Modeling Survival and Testing Biological Hypotheses Using Marked Animals: A Unified Approach with Case Studies

Author(s): Jean-Dominique Lebreton, Kenneth P. Burnham, Jean Clobert and David R. Anderson

Source: *Ecological Monographs*, Vol. 62, No. 1 (Mar., 1992), pp. 67-118
Published by: Wiley
Stable URL: http://www.jstor.org/stable/2937171
Accessed: 22-09-2016 22:16 UTC

Ecological Monographs, 62(1), 1992, pp. 67–118

MODELING SURVIVAL AND TESTING BIOLOGICAL HYPOTHESES USING MARKED ANIMALS: A UNIFIED APPROACH WITH CASE STUDIES[1]

JEAN-DOMINIQUE LEBRETON
CEFE/CNRS, BP 5051, 34033 Montpellier Cedex, France

KENNETH P. BURNHAM
Colorado Cooperative Fish and Wildlife Research Unit, U.S. Fish and Wildlife Service, 201 Wagar Building, Fort Collins, Colorado 80523 USA

JEAN CLOBERT
Laboratoire d'Ecologie, Ecole Normale Supérieure, 46 rue d'Ulm 75231, Paris Cedex 05 France

DAVID R. ANDERSON
Colorado Cooperative Fish and Wildlilfe Research Unit, U.S. Fish and Wildlife Service, 201 Wagar Building, Fort Collins, Colorado 80523 USA

Abstract. The understanding of the dynamics of animal populations and of related ecological and evolutionary issues frequently depends on a direct analysis of life history parameters. For instance, examination of trade-offs between reproduction and survival usually rely on individually marked animals, for which the exact time of death is most often unknown, because marked individuals cannot be followed closely through time. Thus, the quantitative analysis of survival studies and experiments must be based on capture–recapture (or resighting) models which consider, besides the parameters of primary interest, recapture or resighting rates that are nuisance parameters.

Capture–recapture models oriented to estimation of survival rates are the result of a recent change in emphasis from earlier approaches in which population size was the most important parameter, survival rates having been first introduced as nuisance parameters. This emphasis on survival rates in capture–recapture models developed rapidly in the 1980s and used as a basic structure the Cormack-Jolly-Seber survival model applied to an homogeneous group of animals, with various kinds of constraints on the model parameters. These approaches are conditional on first captures; hence they do not attempt to model the initial capture of unmarked animals as functions of population abundance in addition to survival and capture probabilities.

This paper synthesizes, using a common framework, these recent developments together with new ones, with an emphasis on flexibility in modeling, model selection, and the analysis of multiple data sets. The effects on survival and capture rates of time, age, and categorical variables characterizing the individuals (e.g., sex) can be considered, as well as interactions between such effects. This "analysis of variance" philosophy emphasizes the structure of the survival and capture process rather than the technical characteristics of any particular model. The flexible array of models encompassed in this synthesis uses a common notation. As a result of the great level of flexibility and relevance achieved, the focus is changed from fitting a particular model to model building and model selection.

The following procedure is recommended: (1) start from a global model compatible with the biology of the species studied and with the design of the study, and assess its fit; (2) select a more parsimonious model using Akaike's Information Criterion to limit the number of formal tests; (3) test for the most important biological questions by comparing this model with neighboring ones using likelihood ratio tests; and (4) obtain maximum likelihood estimates of model parameters with estimates of precision.

Computer software is critical, as few of the models now available have parameter estimators that are in closed form. A comprehensive table of existing computer software is provided. We used RELEASE for data summary and goodness-of-fit tests and SURGE for iterative model fitting and the computation of likelihood ratio tests.

Five increasingly complex examples are given to illustrate the theory. The first, using two data sets on the European Dipper (*Cinclus cinclus*), tests for sex-specific parameters,

[1] Manuscript received 11 July 1990; revised 25 January 1991; accepted 7 February 1991; final version received 6 March 1991.

explores a model with time-dependent survival rates, and finally uses a priori information to model survival allowing for an environmental variable. The second uses data on two colonies of the Swift (*Apus apus*), and shows how interaction terms can be modeled and assessed and how survival and recapture rates sometimes partly counterbalance each other. The third shows complex variation in survival rates across sexes and age classes in the roe deer (*Capreolus capreolus*), with a test of density dependence in annual survival rates. The fourth is an example of experimental density manipulation using the common lizard (*Lacerta vivipara*). The last example attempts to examine a large and complex data set on the Greater Flamingo (*Phoenicopterus ruber*), where parameters are age specific, survival is a function of an environmental variable, and an age × year interaction term is important. Heterogeneity seems present in this example and cannot be adequately modeled with existing theory.

The discussion presents a summary of the paradigm we recommend and details issues in model selection and design, and foreseeable future developments.

Key words: Akaike's information criterion; capture–recapture; estimation; hypothesis testing; Jolly-Seber; logistic model; mark–recapture; maximum likelihood; model selection; modeling; population dynamics; quasi-likelihood; survival.

Introduction

The study of life history parameters is becoming increasingly important in ecology, particularly in studies of population regulation (Stearns 1980, Barbault 1981, Soulé 1987), community structure and succession (Noble and Slatyer 1980, Crawley and May 1987), and biogeography (Blondel 1986). In this framework, a major goal is to detect and analyze differences in life history traits among groups of individuals through time and space (Clutton-Brock 1988, Lomnicki 1988). Often these differences are supposed to induce changes in fecundity, survival, or both, and thus differences in fitness (Manly 1985, Endler 1986). Estimating fecundity is often possible without major difficulties, whereas estimating survival probabilities raises more problems, particularly under field conditions where time of death is often unknown (Clobert and Lebreton 1990). The probability of survival may vary with individual characteristics such as age, sex, mass, genotype, or phenotype, and also as a function of biotic and abiotic environmental variables. Intra- and interspecific competition and predation can also affect the probability of survival (Loery and Nichols 1985). Testing hypotheses concerning the survival process and estimating survival rates are therefore critical to understanding animal population dynamics.

To estimate survival parameters in the field under natural conditions, one must follow individually marked animals through time. As early as the 1930s, marked animals were used in the study of survival and other parameters (Jackson 1933, 1939). The emphasis in statistical methodology for such capture–recapture data was on the estimation of population size more than on estimation or comparisons of survival rates, whereas ecologists and evolutionists emphasized the structure of survival processes (Stearns 1976, Charlesworth 1980).

Until recently, the literature concentrated on specific models for individual data sets. A major landmark was the Cormack-Jolly-Seber approach to resighting and recapture data (Cormack 1964, Jolly 1965, Seber 1965). The survival model common to the approach of these three authors considered time-dependent survival and recapture rates for a single group of individuals. The explicit formulae for estimating both kinds of rates have been widely used. However, Begon (1983), in a general review, noted that in a large majority of cases the underlying assumptions of capture–recapture were not checked or examined. In many cases survival rates are still frequently estimated in an ad hoc way, e.g., from numbers of known survivors, with the risk of bias and the loss of power it implies (e.g., Dhondt and Eyckerman 1980, Tinbergen et al. 1985, Boyce and Perrins 1987). This inadequacy is partly because the statistical capture–recapture models available until recently were not sufficiently flexible and sometimes not robust to departures from the underlying assumptions (see further comments by Cormack 1979). Moreover, it was seldom feasible to test the fit of the models or to consider alternative models because such statistical tests and alternative models had not been developed.

Advances have been made in the general analysis of capture–recapture data in recent years (see Seber 1986). During the last 10 yr, the primary focus has changed from estimation of population size to estimation of survival (see further comments by Burnham et al. 1987, Clobert and Lebreton 1987). This new focus is advantageous because survival estimators are substantially more robust to the partial failure of assumptions than are estimators of population size (e.g., heterogeneity of individuals to capture or recapture is common and violates a fundamental assumption of capture–recapture theory, Carothers 1973). Simultaneously, a large array of models was developed (Pollock 1975, 1981, Brownie and Robson 1976, 1983, Buckland 1980, 1982, Jolly and Dickson 1980, Pollock and Mann 1983, White 1983, Conroy and Williams 1984, Clobert and Lebreton 1985, Cormack 1985, Crosbie and Manly 1985, Arnason and Schwarz 1986, Burnham et al. 1987, Clobert et al. 1987, Cormack 1989). Current software, such as SURVIV (White 1983) and SURGE (Lebreton and Clobert 1986) allows a number of extensions to

these models such as parameters to be constrained between 0 and 1, or use of transformations on the parameters (e.g., logarithmic or logistic). Parameter values can be set equal, e.g., to model survival as constant over time, with a greater precision because fewer parameters are estimated. Parameters can also be modeled as functions of auxiliary variables, with a regression equation built into the recapture model; thus, survival can be made dependent on rainfall or a measure of winter severity (e.g., North and Morgan 1979); capture rate can depend on measures of effort. Such models are possible (and feasible) to consider, primarily because of advances in computer technology. Maximum likelihood estimation of parameters is easy, given the availability of specialized software. Consequently, the focus for the analysis of capture–recapture data has moved to model selection. Analysts can easily obtain estimates of parameters for a host of models from appropriate computer software. The biology of the population under study drives the selection of an appropriate class of models on which to base data analysis. Statistical theory in terms of goodness-of-fit and likelihood ratio tests allows specific hypotheses to be assessed, leading to further model selection.

In a parallel way, a large variety of marking methods has been developed. Many investigators have initiated well-designed capture–recapture studies, and the number of large data sets from marked populations has increased substantially. In particularly many biologists are involved in collecting data sets concerning several groups of individuals, called hereafter multiple data sets. These data sets may be related to age or sex classes or geographic areas. Alternatively, the data sets may be treatment-control groups as part of a designed experiment. In any case, these data sets may share certain parameters in common, which leads to fairly straightforward new generalizations in modeling (Burnham et al. 1987, Pradel 1988, Pradel et al. 1990). Such modeling will allow greater power in understanding many problems in population biology, using marked animals. From the biological viewpoint, the resulting situation is very different and much more exciting than in the early 1980s.

Our objective is to present a unified view of capture–recapture theory with particular emphasis on multiple data sets, model selection procedures, and practical consequences to population biology. Model development and selection must be based on a priori biological and sampling information and on standard statistical considerations (e.g., goodness-of-fit and likelihood ratio tests). The principle of parsimony leads to a model with as few parameters as possible that provides an adequate explanation of the data (i.e., a model that accounts for the major components of variation in the data). We provide five examples of the application of this theory. The examples chosen include several taxonomic groups that illustrate the types of questions that can be asked in marking studies as well as the computer software available to perform the intensive computations. The examples are restricted to live recaptures (thus, there may be several recaptures for the same individual), including resightings without physical recaptures. The same theoretical framework includes recovery data (of dead animals), to which many of the ideas developed herein would also apply. We do not address the estimation of population size or number of animals entering the population. Our goal in this paper is to synthesize statistical methodology for refined capture–recapture experiments, which are increasingly needed to test theoretical predictions concerning survival. Although this monograph is written primarily for biologists, others (especially in the quantitative and computer sciences) may also find it useful. Readers should be familiar with statistics and have some knowledge of the capture–recapture literature (e.g., Burnham et al. 1987 and Pollock et al. 1990). Ideally, the reader would also have some familiarity with likelihood estimation and inference, analysis of variance and linear model terminology, and logistic regression.

Capture–Recapture Concepts

The ideal situation for a biologist studying processes of mortality would be to follow all individuals from birth to death in the population under study. This situation is nearly achieved for human populations in developed countries. For more specific experimental purposes or biological questions, following a sample of individuals from a given point in time until their death is still the ideal. This situation is sometimes achieved in human epidemiology, although often survivors remain when the experiment or study is terminated (censoring). In survival studies individuals are often grouped for analysis according to categorical variables (e.g., sex, age classes, winter severity), or data analysis is based on individual continuous covariates (e.g., mass, age, body condition), which need to be taken into account in modeling survival. Table 1 provides an example: the simple case of a single batch of n individuals followed over k discrete dates, with time-dependent probabilities of survival $\phi_1, \phi_2, \ldots, \phi_{k-1}$. Because of the simplicity of the underlying binomial structure (Table 1) and parameterization, the development of survival models for this situation has been fairly rapid during the last 20 yr, especially in relation to the epidemiology of cancer (Breslow and Day 1980, Cox and Oakes 1984). The emphasis has been largely on comparisons between groups (see, e.g., Cox and Oakes 1984:Chapter 5; Aitkin et al. 1989:312–315) including possible differences in survival patterns over time (Cox and Oakes 1984:73).

Capture–recapture data

When studying an animal population in the field, it is rarely possible to follow all the individuals of an initial sample over time, even if they are uniquely marked (see, however, Coulson and Wooller 1976).

TABLE 1. A simple example of a time-dependent survival model for a sample followed until a censoring time t_k. Number of survivors first reobserved at time i, m_i, conditional on R_{i-1} (number released at time t_{i-1}) is distributed as BIN(R_{i-1}, ϕ_{i-1}). As a consequence, ϕ_{i-1} is estimated by m_i/R_{i-1}. In contrast to capture–recapture data (Table 2), in this epidemiological model, an individual released in a given year is either observed the following year or never observed again. The m-array is thus diagonal.

	Occasion					
	t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
Numbers released (R_i) and reobserved (m_{ij})	R_1	m_2	0	$\cdots$	0	0
		R_2	m_3	$\cdots$	0	0
			R_3	$\cdots$	m_{k-1}	0
				$\ddots$	$\vdots$	$\vdots$
					R_{k-1}	m_k
Survival	$\leftarrow \phi_1 \rightarrow$	$\leftarrow \phi_2 \rightarrow$	$\leftarrow \cdots \rightarrow$		$\leftarrow \phi_{k-1} \rightarrow$	

Generally, animals will be seen, recaptured, or recorded from time to time, but exact time of death remains unknown. An animal that has not been observed for some time may have survived and escaped recapture by chance or for biological reasons (temporary emigration, e.g., Nichols et al. 1987); its recapture might occur if the study were to continue. This situation is illustrated in the case of two occasions of recapture of a single initial batch of marked animals (=three occasions of capture, see Table 2). The data are a subset of example 1 (European Dipper *Cinclus cinclus*); we use this subset to introduce notation and concepts, although it is not extensive enough for answering biological questions. The data are of the recapture histories of 22 marked individuals (Table 2). Each 1 represents a capture or sighting. The first 1 from the left indicates the time of initial capture, marking, and release, which in this subset example always took place at the first date t_1. Thus "111" indicates an animal that was marked at time t_1 and recaptured both at times t_2 and t_3. The presence of individuals with a recapture history such as "101" is characteristic of capture–recapture data: these individuals were not captured or sighted at time t_2, although they were alive, as proved by their later recapture (at time t_3). Thus, one needs further parameters besides survival to model the recapture process. This presentation of the data is logical because it is a standard statistical array with individuals as rows and variables as columns, the variables being indicators (=0, 1 values) of captures or covariates. Representing the data this way allows any kind of further analysis (e.g., Burnham et al. 1987, Nichols et al. 1987). The data can also be grouped by recapture histories, and presented as recapture trees in which upper branches represent capture and lower branches represent noncapture (Table 3). The number of individuals with recapture history h is noted X_h (here, X_{111}, X_{110}, X_{101}, X_{100}: Table 3).

TABLE 2. Capture histories of 22 European Dippers (*Cinclus cinclus*) followed for 2 yr after initial capture (1 = capture or resighting, 0 = no capture). The first 1 indicates the initial capture and release of an individual after marking. The data are also presented as an m-array (distribution of the dates of first recaptures of individuals released, for the first time or not, in a given year).

a. The 22 capture histories.

1 1 1	1 0 0	1 1 1	1 1 0
1 1 0	1 1 0	1 0 0	1 1 0
1 1 1	1 0 0	1 1 1	1 0 0
1 1 0	1 0 0	1 0 0	1 0 1
1 0 0	1 1 0	1 0 0	1 1 0
1 0 1	1 0 0		

b. The data summarized as an m-array.

Year released i		First recapture, m_{ij}		
		Year 2	Year 3	Never
1	$R_1 = 22$	11	2	9
2		$R_2 = 11$	4	7
3			$R_3 = 6$	6

TABLE 3. Recapture tree for two occasions of capture of a single batch of marked individuals. Upper branches mean capture (1), and lower branches mean no capture (0).* The observed numbers given for occasion 3 deal with European Dippers (*Cinclus cinclus*) followed for 2 yr (see Example 1 later in paper).

Initial number released	Captured, X_{11} not captured, X_{10} on occasion 2	Capture histories on occasion 3	Expected numbers on occasion 3
$R_1 = 22$	X_{11}	$X_{111} = 4$	$R_1\phi_1p_2\beta_3$
		$X_{110} = 7$	$R_1\phi_1p_2(1 - \beta_3)$
	X_{10}	$X_{101} = 2$	$R_1\phi_1(1 - p_2)\beta_3$
		$X_{100} = 9$	$R_1[1 - \phi_1p_2 - \phi_1(1 - p_2)\beta_3]$

	Capture occasion		
	t_1	t_2	t_3
Parameters	survival rates	$\leftarrow \phi_1 \rightarrow$	$\leftarrow \phi_2 \rightarrow$
	capture rates		
	p_1†	p_2	p_3

* This scheme can be modified to take account of losses on capture. Under the assumption of independence and identity of individuals, using survival rates ϕ_1 and ϕ_2 and capture rates p_2 and p_3, one obtains the expected numbers given above. p_1 does not enter these conditional models, hence is not identifiable. ϕ_2 and p_3 appear only as a product $\beta_3 = \phi_2p_3$, and there are only three identifiable parameters (ϕ_1, p_2, and β_3).

† Not identifiable from data conditional on first captures.

Probabilistic framework

The basic probabilistic scheme.—The probabilistic scheme common to all capture–recapture models can be introduced in a simple way with two occasions of recapture. We use survival rates ϕ_1 (from t_1 to t_2) and

ϕ_2 (from t_2 to t_3) and capture rates p_2 (at t_2) and p_3 (at t_3). These parameters are conditional on the animals having survived until the beginning of the period considered (ϕ_i) or until the sampling or recapture occasion considered (p_i). We can thus calculate the probability of the various possible *recapture* histories. For example, given release at t_1, an individual with capture history "101" occurs with probability $\phi_1(1 - p_2)\phi_2 p_3$ because the last 1 in "101" indicates that it survived until t_3 (thus, ϕ_1 and ϕ_2), was not recaptured at t_2 (hence $1 - p_2$), but was recaptured at t_3 (hence p_3). With only two recapture occasions, the parameters ϕ_2 and p_3 appear in the model only as the product, denoted β_3. The capture rate p_1 (at t_1) does not appear and will be as such unidentifiable.

The four recapture histories in the above model are mutually exclusive events. Under the assumption of *i*ndependence of fates and *i*dentity of rates among *i*ndividuals (the **iii** assumption discussed below) the observed numbers are then an observation of a multinomial distribution: $(X_{111}, X_{110}, X_{101}, X_{100})$ is distributed as $\text{MULT}\{R_1;\ \phi_1 p_2 \beta_3,\ \phi_1 p_2(1 - \beta_3),\ \phi_1(1 - p_2)\beta_3,\ 1 - \phi_1 p_2 - \phi_1(1 - p_2)\beta_3\}$.

In general, letting h_j represent the capture history of individual j, the conditional probability $\Pr\{h_j | \text{first release}\}$ depends only on the conditional survival rates (ϕ_i) in the population at risk of capture and the capture rates (p_i) at each occasion. We note $q_i = 1 - p_i$. In the case of $k = 7$ capture occasions, a capture history might be {0110010}, with knowledge that the animal was released alive the last time it was caught. Then the probability of this capture history is

$$\Pr\{0110010 | \text{release at occasion } 2\} = \phi_2 p_3 \phi_3 q_4 \phi_4 q_5 \phi_5 p_6 \chi_6.$$

The probability of not being seen again given release at occasion 6 is symbolized χ_6; this parameter is a function of ϕ_6 and p_7. Some other examples are

$$\Pr\{1001100 | \text{release at occasion } 1\} = \phi_1 q_2 \phi_2 q_3 \phi_3 p_4 \phi_4 p_5 \chi_5;$$

$$\Pr\{0001100 | \text{release at occasion } 4\} = \phi_4 p_5 \chi_5;$$

$$\Pr\{0010001 | \text{release at occasion } 3\} = \phi_3 q_4 \phi_4 q_5 \phi_5 q_6 \phi_6 p_7.$$

Note that by definition $\chi_7 = 1$; thus it does not appear in the last example above. In general, with k occasions of capture, χ_i can be computed recursively (see, e.g., Cormack 1964) working backwards from $i = 6, 5, \ldots, 1$:

$$1 - \chi_i = \phi_i(1 - q_{i+1}\chi_{i+1})$$

with $\chi_7 = 1$. This $\Pr\{h_i | \text{first release}\}$ is always a product of the conditional parameters ϕ, p, and χ because of the assumptions made in capture–recapture and the way ϕ and p are defined. Thus, for every possible capture history h_j, we can write an explicit probability.

Occasionally there are losses on capture (see Jolly 1965) which in terms of modeling these data just means that the animal was not re-released after a capture. Because all losses on capture are known events they do not complicate modeling capture–recapture data (Burnham et al. 1987:28–29). For example, if an animal with capture history "1001100" had actually been sacrificed or deliberately removed on occasion 5, then the probability of this modified capture history is

$$\Pr\{1001100 | \text{release at occasion 1 and removed on occasion 5}\} = \phi_1 q_2 \phi_2 q_3 \phi_3 p_4 \phi_4 p_5.$$

Independence of fates and identity of rates among individuals.—The Jolly-Seber model makes the assumption of independence of fates and identity of rates among individuals (the **iii** assumption) in the population under study. The assumption is commonly not met in field studies. Here we need only deal with independence of fates and identity of rates among marked individuals. A random mixing of marked individuals is neither necessary nor sufficient for the **iii** assumption to hold and is only necessary when population size is to be estimated. The capture design can induce independence of individuals and equality of rates of capture when mixing is not random, e.g., by using random routes across a study area for territorial animals. On the contrary, capture rate may depend on body size, for example, and be heterogeneous even if marked individuals mix randomly; this is frequently the case in fish populations. Another assumption frequently emphasized in the literature is that recaptures should take place at a single point in time (i.e., in a period of negligible length when compared to the between-captures interval) to make survival constant over individuals. This assumption is a consequence of the **iii** assumption, but the effect of violating this assumption can be minimized by using more complex models.

Only a careful examination of the design of recaptures and of the data will make it possible to check, at least partially, the **iii** assumption. This check, as well as further analyses, will often be based on the partitioning of individuals in subcategories; the **iii** assumption will have to hold within each subcategory but not as a whole.

The rates obtained will apply only to the marked fractions of the categories of the study population, and even then one must assume that the act of marking has no effect on survival. Frequently, this restriction will not prevent testing the effect of a factor (such as sex) on survival if any marking effects are additive. The validity of extrapolation of the estimates to the unmarked fractions cannot be tested directly; randomly capturing individuals for marking within each category is the only simple way to provide a basis for this

extrapolation. In particular, one has to keep in mind that most studies are done in areas favorable to the species studied, and that estimates thus apply to the fraction of the population living in those areas.

Estimation of parameters.—The method of maximum likelihood will be used to estimate the parameters in the model. This is a classical method of statistical inference (see, e.g., Edwards 1972, Rao 1972:359, Mood et al. 1974, Burnham et al. 1987:6–22). Because a substantial literature is available on likelihood inference, we give only the essential idea; for more details in an animal marking context see Seber (1982), White et al. (1982), Brownie et al. (1985), and Burnham et al. (1987).

The statistical likelihood of an actual data set is merely the product of the $\Pr(h_j \mid \text{first release})$ over those capture histories actually observed. Because animals with the same recapture history have the same $\Pr(h_j \mid \text{first release})$, the number observed in each recapture history appears as an exponent of the corresponding probability in the likelihood. For instance, in our simple example (Table 3), the likelihood is equal to

$$K(\phi_1 p_2 \beta_3)^{x_{111}}[\phi_1 p_2(1 - \beta_3)]^{x_{110}}[\phi_1(1 - p_2)\beta_3]^{x_{101}} \cdot [1 - \phi_1 p_2 - \phi_1(1 - p_2)\beta_3]^{x_{100}}$$

or, using the intermediate parameter χ_1:

$$K(\phi_1 p_2 \beta_3)^{x_{111}}[\phi_1 p_2(1 - \beta_3)]^{x_{110}} \cdot [\phi_1(1 - p_2)\beta_3]^{x_{101}}(\chi_1)^{x_{100}}.$$

Here, "K" represents a multinomial coefficient which is a known function of the data; K does not depend at all on the parameters. In terms of inference on the parameters in this model, K is irrelevant and does not need to be included in the likelihood; therefore we do not show it any further in our likelihoods.

For notational convenience, let the vector θ represent the identifiable parameters in a given model. Let np be the number of parameters in the model, i.e., the dimension of the vector θ. The corresponding likelihood function for a given data set can be viewed as a function of the parameters and will be noted $L(\theta)$. The maximum likelihood estimate (MLE) of θ, noted $\hat{\theta}$, is the unique value of θ that maximizes $L(\theta)$, or equivalently, the log-likelihood $\ln L(\theta)$. The latter is usually more tractable because it reduces to:

$$\sum_{h_j} (X_{h_j}) \ln\left(\Pr(h_j \mid \text{first release})\right),$$

where the sum is over all observed capture histories h_j. In our simple example in Table 3, the observed log-likelihood is

$$\ln L(\phi_1, p_2, \beta_3) = 4 \ln(\phi_1 p_2 \beta_3) + 7 \ln[\phi_1 p_2(1 - \beta_3)] + 2 \ln[\phi_1(1 - p_2)\beta_3] + 9 \ln(\chi_1).$$

The method of maximum likelihood provides estimators that are asymptotically unbiased, normally distributed, and of minimum variance among such estimators. The first two properties indicate good reliability of the estimates when the number of marked animals is large. The last property (minimum variance) indicates they are fairly optimal in terms of precision. The maximum likelihood method also provides a basis for deriving tests, which will be discussed later (under *Model selection tools*).

Variances and confidence intervals.—Point estimates, $\hat{\theta}$, are only part of the information obtainable from the likelihood function (i.e., from the model). The curvature of the log-likelihood function estimated at the MLEs provides information on the precision of the estimators, i.e., the partial second derivatives of the $\ln L(\theta)$,

$$\frac{\partial^2 \ln L(\theta)}{(\partial\theta_i)(\partial\theta_j)}, \qquad i = 1, \ldots, np, \quad \text{and } j = 1, \ldots, np,$$

computed at $\hat{\theta}$, are used to construct the estimated variance–covariance matrix of $\hat{\theta}$, obtained as the negative of the inverse of the matrix of second-order derivatives. From the variance–covariance matrix we obtain the estimated standard errors, $\widehat{\text{SE}}(\hat{\theta}_i)$, $i = 1, \ldots, np$, of the estimates $\hat{\theta}_i$, i.e., $\widehat{\text{SE}}(\hat{\theta}) = \sqrt{\widehat{\text{var}}(\hat{\theta})}$. An estimated asymptotic 95% confidence interval on θ_i is then $\hat{\theta}_i \pm 1.96\ \widehat{\text{SE}}(\hat{\theta}_i)$. This approach is standard, based on asymptotic normality. Such asymptotic intervals can be improved in many cases by transforming the parameters (e.g., $\log[\phi/(1 - \phi)]$) and then back-transforming the confidence interval endpoints.

Although computationally intensive, another kind of confidence interval, profile likelihood (also called maximum relative likelihood) intervals, can be computed (Kalbfleisch and Sprott 1970, Hudson 1971, Harding et al. 1984, Barndorff-Nielsen 1986, Venzon and Moolgavkar 1988). Profile likelihood intervals have better coverage with small samples than the traditional large-sample approach based entirely on asymptotic normality of the actual $\hat{\theta}$. Morgan and Freeman (1989) provide an example of profile likelihood intervals with band recovery data, and several methods of setting confidence intervals will be illustrated in the examples to follow.

Using $\hat{\theta} \pm 1.96\ \widehat{\text{SE}}(\hat{\theta})$ for a 95% interval on θ (when $\hat{\theta}$ is based on a large sample) does not work well when a parameter estimate is on or near a boundary, such as when we happen to get $\hat{\phi} \geq 1$. Using $\pm 1.96\widehat{\text{SE}}$ also does not work well with some types of models that arise in capture–recapture. This is primarily because the distribution of estimators is very non-normal and the parameter space has boundaries. Even though using $\pm 1.96\widehat{\text{SE}}$ often works very well, this approach is only an approximation to the use of profile likelihood intervals.

We illustrate the idea behind a profile likelihood with a binomial model. Consider the case of sample size $n = 30$ and an observed success count of $y = 0$. Then $\hat{\theta} = y/n = 0.0$ and from the usual estimator of $\text{var}(\theta) = \theta(1 - \theta)/n$ we get $\widehat{\text{SE}}(\hat{\theta}) = 0.0$. The resultant confidence

interval of [0, 0] is ridiculous. The approximate 95% profile likelihood interval in this case is [0, 0.062]. Clearly the lower interval endpoint must be 0. The upper endpoint is obtained from the log-likelihood here: $\ln[L(\theta)] = 30 \ln(1 - \theta)$ as the solution to the equation $-2 \ln(L) = 3.8416$; here 3.8416 is the upper 95% point of the chi-square distribution on 1 df. In general, profile likelihood intervals are based on the relationship between $-2 \ln(L)$ and the chi-square distribution.

Consider the binomial case of $n = 30$ and $y = 1$. The usual 95% interval here is [−0.029, 0.0965]. The profile likelihood interval is the set of all admissible values of θ that satisfy the inequality, in θ, $-2 \ln[L(\hat{\theta})] + 2 \ln[L(\theta)] \leq 3.8416$. Explicitly, we have

$$-2\{\ln(0.03333) + 29 \ln(0.9667)\} + 2\{\ln(\theta) + 29 \ln(1 - \theta)\} \leq 3.8416;$$

the resultant interval is [0.002, 0.139]. Notice that the two intervals have about the same width: 0.126 and 0.137 for standard and profile interval, respectively. However, the standard interval is not admissible as its lower limit is negative.

The above case is somewhat typical; namely profile likelihood intervals often have width about $4\widehat{\text{se}}(\hat{\theta})$, but they are not symmetric about $\hat{\theta}$; rather, profile intervals are shifted and asymmetric in response to the asymmetry of the likelihood. When the likelihood is symmetric and sample size is large, the two approaches give essentially the same result. For the binomial case of $n = 30$ and $y = 15$, the two approaches give

standard confidence interval: [0.321, 0.679]

profile likelihood interval: [0.327, 0.673].

In general, the profile likelihood interval is to be preferred; its only disadvantage is that it is more difficult to compute. The procedure for computing profile likelihood intervals is more complex for multiparameter models, but it is a well-defined procedure (see e.g., Barndorff-Nielsen 1986, Critchley et al. 1988, Venzon and Moolgavar 1988). We expect to see increasing routine use of profile likelihood intervals, and we expect this procedure will soon be programmed into code such as CAPTURE (White 1982), SURVIV (White 1983), and SURGE.

Minimal sufficient statistics.—The information about the np estimable parameters in a given model can be concentrated in r minimal sufficient statistics, which are the only pieces of information needed to obtain the maximum likelihood estimates. In general, r is smaller than d, which is the total number of degrees of freedom (df) associated with the raw data. Therefore, a number of different data sets can lead to the same estimates because they produce the same observed values of the minimal sufficient statistics. In this simple example, $np = r = d = 3$. Minimal sufficient statistics provide a concise summary of the data for the model being considered. They can be used as standard input to computer programs (e.g., Burnham 1989). Thus, identifying the mathematical form of the minimal sufficient statistics for a given model is an important task (see, e.g., Pollock et al. 1985, Burnham 1988).

When there are more elementary parameters than minimal sufficient statistics, at most, r functions of the parameters can be estimated. There are then as many identifiable parameters (i.e., parameters that can be estimated) as minimal sufficient statistics ($np = r$). In such a case the model is said to be of full rank, and the maximum likelihood method often leads to explicit formulae for estimates. This is the case in our Dipper example. The four observed numbers provide only $r = 3$ different pieces of information because they sum to R_1. The $np' = 4$ elementary parameters (ϕ_1, ϕ_2, p_2, p_3) reduced to $np = 3$ (ϕ_1, p_2, β_3) estimable parameters. Maximizing $\ln L(\phi_1, p_2, \beta_3)$ leads to $\hat{\phi}_1 = 0.7500$, $\hat{p}_2 = 0.667$, and $\hat{\beta}_3 = 0.3636$. Because $r = d$ in this particular case, there are no degrees of freedom left to test the fit of the model (see *Model selection tools*): the model is said to be saturated. Full rank models are important in two respects: first, they have been used extensively; explicit estimates can be obtained from a desk calculator. The classic example is the Jolly-Seber model (Jolly 1965, Seber 1965). Second, the explicit formulae can be improved, for instance, to reduce their bias that might be non-negligible for small sample sizes (see example in Seber 1982:204). However, in many cases, there will be fewer parameters to estimate than the number of minimal sufficient statistics ($np < r$). In these cases, numerical iterative fitting is needed to find the maximum likelihood estimates. The numerical approach is the only one possible to obtain MLEs for many reduced models. It has been developed extensively during the last 10 yr (White 1983, Clobert and Lebreton 1985, Brownie et al. 1986) and requires computer software.

A general model

Two occasions of recapture.—The information available, and thus the potential flexibility in modeling, is increased when one considers several batches of marked animals. Traditionally, in the literature, one considered successive batches of animals, similar to the first batch, marked and released at the times of recaptures. In the case of two occasions of recapture, one should consider $R_1 = U_1$ releases at time t_1 and U_2 releases at time t_2 of newly marked animals. (Animals newly marked at the last occasion, time t_3, will not provide any information about survival.) The total number of animals released at time t_2 (R_2) is the sum of the number newly marked (U_2), plus the number previously marked at time t_1 and captured at time t_2 ($X_{111} + X_{110} = X_{11}$), minus any losses on capture at time t_2. This scheme (illustrated with the European Dipper data in Table 4) has been used extensively in cases of physical recapture (rather than resightings) because the field sampling at

TABLE 4. Recapture data represented by a binary tree for an age- and time-dependent model with two occasions of capture and two successive cohorts of marked individuals. The observed numbers given here deal with European Dippers (*Cinclus cinclus*) followed for 2 yr (see Example 1 later in paper). Individuals newly marked at time t_3 do not provide information on survival. The second column of the *m*-array (below) is composed of the $X_{11} = 11$ previously marked individuals and of $U_2 = 49$ newly marked individuals, all released at time t_2.

Initial number released	Captured, X_{11}, X_{01} not captured, X_{10} on occasion 2	Capture histories on occasion 3	Expected numbers on occasion 3
	Occasion		
t_1	t_2	t_3	
$R_1 = 22\ (=U_1)$	X_{11}	$X_{111} = 4$	$R_1\phi_1p_2\beta_3$
		$X_{110} = 7$	$R_1\phi_1p_2(1 - \beta_3)$
	X_{10}	$X_{101} = 2$	$R_1\phi_1(1 - p_2)\beta_3$
		$X_{100} = 9$	$R_1[1 - \phi_1p_2 - \phi_1(1 - p_2)\beta_3]$
	$U_2 = 49\ (=X_{01})$	$X_{011} = 20$	$U_2\beta'_3$
		$X_{010} = 29$	$U_2(1 - \beta'_3)$

Occasion			
t_1	t_2		t_3
$R_1 = 22$		11	2
	$h = \{11\}$	$X_{11} = 11$	4
	$h = \{01\}$	$U_2 = 49$	20
		$R_2 = 60$	24

time t_2 provides marked animals (X_{11}) and unmarked (U_2) animals.

The R_2 animals are a mixture of X_{11} animals belonging to the first batch (marked at time t_1), and of U_2 animals belonging to the second one (marked at time t_2). In some cases, marking will take place immediately after birth, and such batches will be genuine cohorts, in the meaning of demographers. In the general case, we call cohorts the groups made up of U_1, U_2, ... animals marked at successive points in time. Then, age has the meaning of time elapsed since first capture. In some cases, in particular, when trap dependence is of concern, we will call cohorts the groups of R_1, R_2, ... animals released at successive times, whether newly marked or not. Then, age has the meaning of time elapsed since last capture.

In Table 4, we consider a general parameterization allowing time dependence in each cohort independently of the others. The model for these data has $d = 4$ df (3 for the first cohort, 1 for the second) for $np = 4$ parameters to be estimated (ϕ_1, p_2, β_3, β'_3). In addition to the $R_1 = 22$ individuals marked at time t_1, the data include $U_2 = 49$ individuals marked and released at time t_2, among which $X_{011} = 20$ are recaptured at time t_3. The explicit ML estimates are: $\hat{\phi}_1 = 0.7500$, $\hat{p}_2 = 0.6667$, $\hat{\beta}_3 = 0.3636$, and $\hat{\beta}'_3 = 0.4082$ (i.e., 20/49). The parameters can be presented as triangular tables, according to cohort and time, as proposed by Clobert et al. (1987) (Table 5).

General case.—In the case of k occasions of capture (including initial capture), the parameters can also be represented with two indices in triangular tables (Clobert et al. 1987), as shown in Table 6. This parameterization [with $k(k - 1)$ parameters, among which $(k - 1)^2$ are separately identifiable] was introduced in Tables 4 and 5 in the case of two occasions of recapture ($k - 1 = 2$). This general model is obtained by considering time dependence in each cohort independently of the others (Pollock 1975): one refers to full age- (time since first capture) and time-dependent model, or cohort- and time-dependent model. The Robson-Pollock model (Robson 1969) represents such a generality. Explicit estimates for the ϕ_{ij} and p_{ij} are given by Pollock (1975).

More flexibility: classifying and describing models

Introduction.—There is a broad choice of parameterizations when there are several marked cohorts. Also, many models appear as special cases of the full age- and time-dependent model: we will refer to these as constrained models. Constraints involve equality between parameters as well as relationships with external variables (e.g., weather variables for survival, effort variables for capture rates) (Pollock et al. 1984, Clobert and Lebreton 1985, Clobert et al. 1987).

Time-dependent model.—When there are two recapture occasions (Table 4), the constraint $\beta_3 = \beta'_3$ leads to a parameterization that depends only on time, not cohorts. Animals in cohort 1 and cohort 2 released at time t_2 will have the same product $\beta_3 = \phi_2p_3$ for the next time period. There are $d = 4$ df and $np = 3$ param-

TABLE 5. Triangular table of parameters of the age- and time-dependent model for two occasions of recapture with two cohorts of marked individuals. A cohort is the set of all individual animals released at a given time; these releases can come from animals first captured at that time or, in some cases, from all animals recaptured at that occasion, whether for the first time or not.*

Survival parameters (apply to intervals between dates)

Date:	t_1	t_2	t_3
Cohort: 1	$\leftarrow\!\!\!-\phi_1\!\!\!\rightarrow$	$\leftarrow\!\!\!-\phi_2\!\!\!\rightarrow$	
2		$\leftarrow\!\!\!-\phi'_2\!\!\!\rightarrow$	

Capture parameters (apply to occasions)

Date:	t_1	t_2	t_3
Cohort: 1		p_2	p_3
2			p'_3

* ϕ_i = survival probability from time i to $i + 1$, conditional upon an animal being alive at time i; p_i = probability of an animal alive at time i being captured or recaptured at time i.

eters to be estimated. However, β_3 is estimated from the ratio $(4 + 20)/(11 + 49)$, in which contributions from both cohorts are pooled. One degree of freedom was lost by pooling to obtain $r = 3$ minimal sufficient statistics and, as discussed later, can be used for a goodness-of-fit test, which here reduces to a test of the hypothesis $\beta_3 = \beta'_3$. Here the estimates $\hat{\phi}_1 = 0.7273$, $\hat{p}_2 = 0.6875$, $\hat{\beta}_3 = 0.4000$ can be obtained explicitly.

This parameterization extends readily to k occasions of capture at times $t_1, t_2, \ldots, t_k$ (i.e., $k - 1$ occasions of recapture at times $t_2, \ldots, t_k$) for cohorts of animals released at these successive points in time. There are $2k - 2$ parameters $\phi_1, p_2, \phi_2, p_3, \ldots, \phi_i, p_{i+1}, \ldots, \phi_{k-2}, p_{k-1}, \phi_{k-1}, p_k$, which reduce to $2k - 3$ identifiable parameters $\phi_1, p_2, \phi_2, p_3, \ldots, \phi_i, p_{i+1}, \ldots, \phi_{k-2}, p_{k-1}, \beta_k = \phi_{k-1}p_k$. This is the Cormack-Jolly-Seber model (CJS model: see Brownie 1987 and Clobert and Lebreton 1987 for a discussion of the name), for which well-known explicit formulae exist (see Seber 1982 and Burnham 1988 for a full account, and Begon 1983 for a critical review of its use). The triangular tables of parameters for the CJS model are given in Table 7.

The CJS model has been the most general approach to survival estimation by capture–recapture since 1964. However, in many situations, this model is either too general or too restrictive "... the model is accused of being too restrictive by requiring that all individuals, whatever their age or capture history, should have the same probabilities of capture and survival ... by including a separate parameter for each survival and capture probability it is too general" (Cormack 1979:241). The Robson-Pollock model partly answers the first criticism. The Robson-Pollock model is obtained when fitting the CJS model separately for each cohort. Models with further constraints will answer the second criticism; methodology for constructing and fitting such restricted models is dealt with here.

Models with constant rates.—Starting from the CJS model, three straightforward models can be considered: ϕ constant and p time-dependent (model $[\phi, p_t]$), ϕ time-dependent and p constant (model $[\phi_t, p]$), and ϕ and p both constant (model $[\phi, p]$). In this notation, the CJS model is (ϕ_t, p_t). These various models, with constraints of constancy, have been introduced independently by several authors (Jolly and Dickson 1980, Clobert 1981, Cormack 1981, Sandland and Kirkwood 1981, Jolly 1982; see also further developments by Clobert et al. 1985, Crosbie and Manly 1985, Brownie et al. 1986). In model (ϕ_t, p), the hypothesis of constancy of the capture rate cannot be tested for the last year: identifiable parameters can be represented as $\phi_1, \phi_2, \ldots, \phi_{k-2}, \beta_k, p$. Comparing this model to the CJS model tests the hypothesis $p_2 = \cdots = p_{k-1}$ $(=p)$ only (with $df = k - 3$). In turn, ϕ_{k-1} can be estimated by β_k/p, under the hypothesis $p_k = p$, which cannot be tested. Similar remarks hold for model (ϕ, p_t), whereas tests of model (ϕ, p) address, for the last year, the hypothesis $\beta_k = \phi p$ (and not the compound hypothesis $\phi_{k-1} = \phi, p_k = p$).

Equality constraints between parameters of models more general than CJS lead to a variety of models (Table 7B, C, D), which include, for instance, reductions in the number of parameters of the Robson-Pollock age- and time-dependent model. The case of the fully age-dependent model $[\phi_a, p_a]$ is illustrated in Table 7B. A second example is the time-dependent model with two age classes for survival for which Brownie and Robson (1983) obtained explicit estimates (Table 7D). Further models can be deduced from the CJS model by reducing time dependence in survival and capture to a few levels, such as good and poor years (Clobert et al. 1985). All of these models require iterative numerical methods to find the MLEs. They can be compared with full rank models by likelihood ratio

TABLE 6. Parameters of the full age- and time-dependent Robson-Pollock model (Pollock 1975). There are $(k - 1)k$ parameters; however, the components of the $k - 1$ products $\beta_{ik} = \phi_{i,k-1}p_{ik}$ are not separately identifiable; therefore there are only $(k - 1)^2$ separately identifiable parameters.

Survival parameters (apply to intervals between occasions)

	Time					
Cohort	t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
1		ϕ_{11}	ϕ_{12}	$\cdots$	$\phi_{1,k-1}$	
2			ϕ_{22}	$\cdots$	$\phi_{2,k-1}$	
$\vdots$				$\cdots$	$\vdots$	
$k - 1$					$\phi_{k-1,k-1}$	
k						

Recapture parameters (apply to occasions)

	Time					
Cohort	t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
1		p_{12}	p_{13}	$\cdots$	$p_{1,k-1}$	p_{1k}
2			p_{23}	$\cdots$	$p_{2,k-1}$	p_{2k}
$\vdots$				$\cdots$		$\vdots$
$k - 1$						$p_{k-1,k}$
k						

TABLE 7. Parameters of reduced age- and time-dependent capture–recapture models, presented as triangular tables.

A. Time-dependent model (Cormack-Jolly-Seber) ($2k - 3$ identifiable parameters)

Survival parameters

Cohort	Time t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
1	ϕ_1	ϕ_2		$\cdots$	ϕ_{k-1}	
2		ϕ_2		$\cdots$	ϕ_{k-1}	
$\vdots$				$\cdots$	$\vdots$	
$k-1$					ϕ_{k-1}	
k						

Recapture parameters

Cohort	Time t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
1		p_2	p_3	$\cdots$	p_{k-1}	p_k
2			p_3	$\cdots$	p_{k-1}	p_k
$\vdots$				$\cdots$		$\vdots$
$k-1$						p_k
k						

B. Fully age-dependent model ($2k - 3$ identifiable parameters)

Cohort	Time t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
1	ϕ_1	ϕ_2		$\cdots$	ϕ_{k-1}	
2		ϕ_1		$\cdots$	ϕ_{k-2}	
$\vdots$				$\cdots$	$\vdots$	
$k-1$					ϕ_1	
k						

Recapture parameters

Cohort	Time t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
1		p_2	p_3	$\cdots$	p_{k-1}	p_k
2			p_2	$\cdots$	p_{k-2}	p_{k-1}
$\vdots$				$\cdots$		$\vdots$
$k-1$						p_2
k						

C. Cohort-dependent model ($2k - 3$ identifiable parameters)

Survival parameters

Cohort	Time t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
1	ϕ_1	ϕ_1		$\cdots$	ϕ_1	
2		ϕ_2		$\cdots$	ϕ_2	
$\vdots$				$\cdots$	$\vdots$	
$k-1$					ϕ_{k-1}	
k						

Recapture parameters

Cohort	Time t_1	t_2	t_3	$\cdots$	t_{k-1}	t_k
1		p_1	p_1	$\cdots$	p_1	p_1
2			p_2	$\cdots$	p_2	p_2
$\vdots$				$\cdots$		$\vdots$
$k-1$						p_{k-1}
k						

TABLE 7. Continued.

D. Time-dependent model with two age classes for survival (Brownie and Robson 1983) ($3k - 5$ identifiable parameters)

Survival parameters

Cohort	Time t_1	t_2	t_3	t_4	$\cdots$	t_{k-1}	t_k
1	ϕ_{11}	$\phi_{\cdot 2}$	$\phi_{\cdot 3}$		$\cdots$	$\phi_{\cdot,k-1}$	
2		ϕ_{22}	$\phi_{\cdot 3}$		$\cdots$	$\phi_{\cdot,k-1}$	
3			ϕ_{33}		$\cdots$	$\phi_{\cdot,k-1}$	
$\vdots$					$\cdots$	$\vdots$	
$k-1$						$\phi_{k-1,k-1}$	
k							

Recapture parameters

Cohort	Time t_1	t_2	t_3	t_4	$\cdots$	t_{k-1}	t_k
1		p_2	p_3	p_4	$\cdots$	p_{k-1}	p_k
2			p_3	p_4	$\cdots$	p_{k-1}	p_k
3				p_4	$\cdots$	p_{k-1}	p_k
$\vdots$					$\cdots$		$\vdots$
$k-1$							p_k
k							

tests (see *Model selection tools: likelihood ratio tests,* Sandland and Kirkwood 1981, Clobert et al. 1985).

Models with time-dependent external variables.—ϕ_t and p_t can be also constrained to be functions of external variables over time. Relevant variables include environmental variables for survival (Clobert and Lebreton 1985) and measures of effort or efficiency for capture rate (Pollock et al. 1984, Clobert et al. 1987). There are advantages in expressing such constraints in the framework of generalized linear models (McCullagh and Nelder 1983) as ϕ_t or $p_t = f(a_0 + \Sigma\, a_i x_{it})$. The function f links the parameters to a linear formula, and, as such, is usually called a link function. It is usually given by its inverse f^{-1}.

Possible link functions include

$$\text{Identity,}\quad f^{-1}(x) = x$$

$$\text{Logit,}\quad f^{-1}(x) = \text{logit}(x) = \log[x/(1-x)]$$

$$\text{Log,}\quad f^{-1}(x) = \log(x)$$

and

$$\text{Hazard,}\quad f^{-1}(x) = \log[-\log(x)].$$

Each of these functions has its own advantages and disadvantages, and criteria of choice may vary from case to case (see Cox and Oakes 1984:74 concerning epidemiological models). The logit-link function keeps estimates of ϕ_t and p_t within the interval (0, 1) (Buckland 1980). The log-link makes multiplicative formulae additive and can simplify the treatment of data obtained over unequal time intervals. For survival, $\log(1 - \phi)$ approximates $\log[-\log(\phi)]$, which is the discrete time log-hazard rate used in epidemiology (Aitkin et al. 1988: 313).

Dependence of survival and of capture rate on a

variable x is noted $\phi[f(x)]$ and $p[f(x)]$, respectively, or $\phi(x)$ and $p(x)$ when there is no ambiguity about the link function used. In such an approach, the overall precision is increased because time dependence is expressed in a parsimonious way, with a risk of increasing bias. Moreover, the effect of an external variable on survival can be tested by comparing between models, e.g., $[\phi(x), p]$ vs. $[\phi, p]$, whereas an ordinary least-squares analysis of CJS estimates over the variable may be inefficient because of the autocorrelation of estimates (Conroy and Williams 1984, Clobert and Lebreton 1985). Lastly, comparing a model with an external variable to the corresponding time-dependent model (e.g., $[\phi(x), p]$ vs. $[\phi_t, p]$) tests the adequacy of the functional relationship, in a similar way to a test of linearity in regression (cf. Burnham et al. 1984, Clobert and Lebreton 1985).

The general theory of likelihood inference applies irrespective of the type of link function used to reparameterize the model based on survival rates ϕ and capture rates p. In particular, we emphasize the logit link function (see, e.g., Hosmer and Lehmeshow 1989) as a way to facilitate placing capture–recapture theory into a linear model framework. Program SURGE (Lebreton and Clobert 1986, Pradel et al. 1990) computes the MLEs of the parameters on a logit scale and their sampling variances and covariances (using the information matrix and asymptotic assumptions). The back transformation to the survival rate is

$$\hat{\phi} = \{1 + \exp[-\text{logit}(\hat{\phi})]\}^{-1}.$$

The asymptotic standard error of the back transformation is then

$$\widehat{\text{SE}}(\hat{\phi}) = [\hat{\phi}(1 - \hat{\phi})]\widehat{\text{SE}}[\text{logit}(\hat{\phi})].$$

The expression for sampling covariances is similar; for example, for any two parameters $\hat{\phi}_a$ and $\hat{\phi}_b$

$$\widehat{\text{cov}}(\hat{\phi}_a, \hat{\phi}_b) = [\hat{\phi}_a(1 - \hat{\phi}_a)][\hat{\phi}_b(1 - \hat{\phi}_b)] \cdot \widehat{\text{cov}}[\text{logit}(\hat{\phi}_a), \text{logit}(\hat{\phi}_b)].$$

Also, when using the logit-link, confidence intervals should be computed by first getting an interval on logit(ϕ) then back transforming the two interval end points. Such confidence intervals are typically asymmetric about $\hat{\phi}$. Using these intervals is better than using $\hat{\phi} \pm 2\widehat{\text{SE}}(\hat{\phi})$ because logit($\hat{\phi}$) tends generally to be more normally distributed than is $\hat{\phi}$. In addition, such back-transformed interval end points on ϕ are never outside the interval [0, 1].

Dummy variables.—Models with two levels of survival for good and poor years can easily be expressed in this framework by using a dummy categorical variable. The use of dummy variables can be generalized, in the same way that analysis of variance (ANOVA) models can be represented as multiple regression models using dummy categorical variables (see e.g., Kleinbaum and Kupper 1978:Chapter 13). Any subset of parameters can be constrained to be equal using this technique. All the models considered up to now can be represented in this way. For example, the fully age-dependent model in Table 7B can be obtained from the Robson-Pollock model by writing ϕ and p as linear functions of indicator variables of age, whereas the CJS model can be obtained by writing them as linear functions of indicator variables of time (see Appendix). Such dummy variables together with a quantitative covariate are shown for the roe deer example.

Rules for model description.—When constraints based on link functions and linear predictors using external variables (either categorical or not) are used, a description of models based on triangular tables of parameters is incomplete. We extend the notation of models using indices, introduced by Sandland and Kirkwood (1981). In this notation the CJS model is model (ϕ_t, p_t); any other model can be described by noting the link function used and the linear predictor for ϕ and p as proposed by Wilkinson and Rogers (1973). We use the logit as the default link function and, thus, usually omit it in the model notation. In the Robson-Pollock model, time is considered with different parameters in each cohort, i.e., for animals for different (relative) ages: $f^{-1}(\phi) = u_{at}$ and $f^{-1}(p) = v_{at}$, where a and t index, respectively, all possible values of age and time (see Appendix). The Robson-Pollock model is thus noted $(\phi_{a\cdot t}, p_{a\cdot t})$.

Interactions.—In models (ϕ_a, p_a) and (ϕ_t, p_t), respectively, ϕ (or p) vary only with age or time, respectively, and can thus be noted $f^{-1}(\phi)$ [or $f^{-1}(p)$] $= g_a$ and h_t. An obvious intermediate model is based on an additive variation: $f^{-1}(\phi)$ [or $f^{-1}(p)$] $= g_a + h_t$, where f is an appropriate link function. This model is noted (ϕ_{a+t}, p_{a+t}). It differs from the Robson-Pollock model in that variations over time are parallel (in the scale induced by the link function) in the various age classes (and vice versa). The Robson-Pollock model $(\phi_{a\cdot t}, p_{a\cdot t})$, in which ϕ (or p) vary independently over age and time, i.e., as $f^{-1}(\phi)$ [or $f^{-1}(p)$] $= u_{a\cdot t}$, can be rewritten $f^{-1}(\phi)$ (or $f^{-1}(p)$) $= g_a + h_t + m_{a\cdot t}$. It includes interaction terms $m_{a\cdot t}$, which, as in a two-way analysis of variance, can be easily represented by dummy variables (see *Appendix*): comparing models (ϕ_{a+t}, p_{a+t}) and $(\phi_{a\cdot t}, p_{a\cdot t})$ is equivalent to testing $m_{a\cdot t} = 0$, i.e., testing parallelism over time between age classes. A comparison of this approach with two-way analysis of variance is drawn in the Appendix, which provides a summary of model notation. Interaction terms can be considered for any "main effects" in case of need. Interaction terms can be added for realism or dropped for parsimony in modeling survival from capture–recapture data (Pradel 1988, Pradel et al. 1990). A sequence of models progressively deleting interaction terms is given for the roe deer.

Multiple data sets

Introduction.—Besides the scheme of cohorts, Jolly (1965) noted that there are other possible schemes,

even those that rarely have been considered in the literature. The release of cohorts of various types of individuals, as well as the release of individuals at times that do not coincide with recapture (or resighting) occasions, are examples. Each cohort can receive its own parameterization or share some parameters with some other cohorts (e.g., Burnham et al. 1987, Pradel 1988, Burnham 1989). In the first case this leads to estimates that are independent over cohorts. An example here might be a study of small mammals trapped at monthly intervals where a strong seasonal effect is present.

An elementary example.—Let us consider, as an elementary example, data partitioned by sex (denoted s). The classical approach consists of fitting the CJS model separately for each sex. Each parameter is defined separately for the two sexes; thus, we can speak of the $(\phi_{t \cdot s}, p_{t \cdot s})$ model (this model is structure $H_{k-1,\phi}$ of Burnham et al. 1987:113). Fitting the CJS model to pooled data makes sex disappear; we are back to model (ϕ_t, p_t) (model structure H_0 of Burnham et al. 1987: 116). Many intermediate models are possible. If males and females are equally prone to capture, it may be relevant to constrain probabilities of capture to be equal to enhance precision on survival: model $(\phi_{t \cdot s}, p_t)$ or $(\phi_{t \cdot s}, p)$ will be considered. Thus, groups of animals fit naturally in our notation. (See Burnham 1989 for an example using male and female beetles.)

If probability of capture varies between sexes with some kind of parallelism over time, e.g., because variations in effort induce similar variations in capture rates of males and females, otherwise differing in their overall susceptibility to capture, model $\phi_{t \cdot s}, p_{f(t+s)}$ will be useful with an appropriate link function f. If one incorporates sex (s) as a dummy variable (e.g., males = 1, females = 0), then the logistic link function leads to:

$$\text{logit}(p_{t+s}) = a_1 s + b_t,$$

or

$$p_{t+s} = \frac{1}{1 + \exp[-(a_1 s + b_t)]}.$$

The subscript $t + s$ indicates that sex is additive on a logit scale. In this example, the capture probabilities are year-specific (with logit parameters b_t) and the values for males and females differ by a constant value (on a logit scale), a_1 (i.e., parallel over time). This concept of parallelism is illustrated in the Swift example. Similarly, parallel variations in survival of the two sexes can be modeled as $[\phi_{f(t+s)}, p_+]$, p_+ denoting an appropriate model for capture rate. In particular, with the link function $h(x) = \exp[-\exp(x)]$ [or approximately $1 - \exp(x)$ if survival is close to 1] for survival, the log rate of death, or log-hazard rate, is modeled linearly. With p fixed to 1, we obtain model $[\phi_{h(t+s)}, 1]$, which is Cox's proportional hazards model (see, e.g., McCullagh and Nelder 1983:183, Aitkin et al. 1988: 313), which can be extended to capture–recapture data as $(\phi_{h(t+s)}, p_+)$.

Dummy variables and general models.—In the most general case we consider, each individual is characterized by v dummy variables $g_1, g_2, \ldots, g_v$, respectively, with $m_1, m_2, \ldots, m_v$ categories each. There are thus $m = m_1 m_2, \ldots, m_v$ groups. In Wilkinson and Rogers' (1973) notation, we denote $g = g_1 * g_2 * \ldots * g_v$ the categorical variable composed of all the combinations of the g_i. Thus, each category of this new variable characterizes one of the m groups.

The $(\phi_{a \cdot t}, p_{a \cdot t})$ model can be applied separately to these m groups. This collection of $(\phi_{a \cdot t}, p_{a \cdot t})$ models is denoted as model $(\phi_{g1 \cdot g2 \cdot \ldots \cdot gv \cdot a \cdot t}, p_{g1 \cdot g2 \cdot \ldots \cdot gv \cdot a \cdot t})$, or, shortened, as $(\phi_{g \cdot a \cdot t}, p_{g \cdot a \cdot t})$. It has $mk(k-1)$ parameters, of which $m(k-1)^2$ are estimable (intrinsic aliasing of parameters, McCullagh and Nelder 1983:47). In practice, for a particular data set, some of these parameters will not be identifiable (extrinsic aliasing, McCullagh and Nelder 1983:51). The $mk(k-1)$ parameters can be arranged in m pairs of triangular tables of the kind used for a single data set. Linear constraints (with appropriate link functions) between subsets of these parameters induce a general class of models. Among such constraints, those based on dummy categorical variables offer a wide range of possibilities, as noted by Manly (1985). These constraints cover the case of equality constraints between parameters as well as the case of additive effects. Case studies will in general consider a sequence of nested models incorporating progressive constraints. The model to begin with will frequently be already constrained because of identifiability problems.

Capture-history dependence

A simple example of a model with trap dependence (see Cormack 1981, Sandland and Kirkwood 1981) is given in Table 8 for the first cohort of European Dippers, followed for 3 yr. Such a model is more general than the time- and age-dependent models. This model considers different parameters for individuals in the same cohort if they differ in their previous capture history, i.e., if they belong to different subcohorts (see Robson 1969, Burnham et al. 1987). There are seven pieces of information for seven parameters. Unfortunately, X_{1110}/X_{1111} and X_{1010}/X_{1011} both estimate $(1 - \beta_4)/\beta_4$. Thus, the seven parameters in the model cannot be estimated. Many similar models have too many parameters, thus leading to a lack of identifiability. A way to limit over-parameterization is to limit dependence on previous history to a switch in the capture rate according to the previous event. The capture rate is p' instead of p if the previous event was a capture or recapture, with $f^{-1}(p') = f^{-1}(p) + \alpha$, with link function f. Sandland and Kirkwood (1981) considered f^{-1} to be a logarithmic function, and we will use, later, f^{-1} as a linear logistic function (*Example 5, Greater Flamingo*). Standard models, without this Markovian dependence,

are obtained through the constraint $\alpha = 0$. Capture-history dependence will have to be considered only if preliminary tests indicate it seems unavoidable.

A related feature of capture–recapture is the possibility of a handling effect (see Arnason and Mills 1987, Burnham et al. 1987:218–221). Whereas a Markovian switch in capture rates can be modeled, handling effects models suffer from severe parameter identifiability problems. A permanent behavioral switch in capture probability after first capture is different from either a Markovian switch or a handling effect. By a behavioral effect on p we mean the permanent change after first capture, as in the models of Otis et al. (1978). We do not need to consider such a behavioral effect in any of our models here because it is totally nonidentifiable and has no effect. This is because we condition on first release in all our likelihoods, so the only capture probabilities that we ever estimate are really recapture probabilities.

Model Selection

Analysis philosophy—general principles

Introduction.—In all applied subject areas of statistics, the analysis of complex data becomes an exercise in selecting an appropriate model to describe the data (see, e.g., Box and Jenkins 1970:17–19, McCullagh and Nelder 1983:1–6, Pregibon 1984:1995, and Durbin 1987:184–185). Much literature has been written about general and specific aspects of data analysis and model selection; relevant references discussing general philosophy and strategy include Leamer (1978), Gnanadesikan (1983), Pregibon (1984), Linhart and Zucchini (1986), Durbin (1987), Chatfield (1988), and McCullagh and Nelder (1989). Our philosophy of data analysis and strategy for model selection is well illustrated by Chatfield (1988) and is nicely summarized by McCullagh and Nelder (1983:6):

> *Modeling in science remains, partly at least, an art. Some principles exist, however, to guide the modeler. The first is that* all models are wrong*; some, though, are better than others and we can search for the better ones. At the same time we must recognize that eternal truth is not within our grasp. The second principle (which applies also to artists!) is not to fall in love with one model, to the exclusion of alternatives. Data will often point with almost equal emphasis at several possible models and it is important that the analyst accepts this. A third principle involves checking thoroughly the fit of the model to the data, for example by using residuals and other quantities derived from the fit to look for outlying observations, and so on. Such procedures are not yet fully formalized (and perhaps never will be), so that imagination is required of the analyst here as well as in the original choice of models to fit.*

This chapter presents some principles and strategies for model selection for survival analysis of data on marked animals.

TABLE 8. An example of a model with trap dependence for the case of three capture occasions; if an animal is captured on occasion i (for $i = 2$ or 3) then its capture probability on occasion $i + 1$ (for $i + 1 = 3$ or 4) is p_i; if the animal is not caught on occasion i, then its capture probability on occasion $i + 1$ is p_i'. Also, note the identifiable parameters $\beta_4 = \phi_3 p_4$ and $\beta'_4 = \phi_3 p'_4$.

Initial number released	Recapture tree Occasions			
t_1	t_2	t_3	t_4	Expected numbers
R_1	X_{11}	X_{111}	X_{1111}	$\phi_1 p_2 \phi_2 p_3 \beta_4$
			X_{1110}	$\phi_1 p_2 \phi_2 p_3 (1 - \beta_4)$
		X_{110}	X_{1101}	$\phi_1 p_2 \phi_2 (1 - p_3) \beta'_4$
			X_{1100}	$\phi_1 p_2 [1 - \phi_2 p_3 - \phi_2 (1 - p_3) \beta'_4]$
	X_{10}	X_{101}	X_{1011}	$\phi_1 (1 - p_2) \phi_2 p'_3 \beta_4$
			X_{1010}	$\phi_1 (1 - p_2) \phi_2 p'_3 (1 - \beta_4)$
		X_{100}	X_{1001}	$\phi_2 (1 - p_2) \phi_2 (1 - p'_3) \beta'_4$
			X_{1000}	1 − (the 7 terms above)

Principle of parsimony.—Models relating sample data and population parameters should contain enough parameters (such as probabilities of survival and capture) to account for all of the significant variation (patterns) in the data (see, e.g., Box and Jenkins 1970:17–18, Bishop et al. 1975:311–313, Fienberg 1981:56, McCullagh and Nelder 1983:5–6, 68–69, Goodman 1984:34–36, Chatfield 1988:16, and Read and Cressie 1988:125–126). As the number of parameters increases, the bias in the estimators decreases; however, the sampling variance increases. Thus, an important trade-off exists. The principle of parsimony indicates that a model with too many parameters is undesirable; only those parameters that are justified by the data (i.e., are "significant") should be retained in the model actually supported by the data.

The concept of a parsimonious model is illustrated by the following hypothetical capture–recapture study of a resident population over a 6-yr period. The Cormack-Jolly-Seber (CJS) model would have four annual survival probabilities ($\phi_1, \ldots, \phi_4$) and four capture probabilities ($p_2, \ldots, p_5$) plus one confounded parameter, $\phi_5 p_6$ (ϕ_5 and p_6 are not separately estimable with only 6 yr of marking). It is often the case that the variation over years in survival is relatively minor. Thus, a parsimonious approach might be to model annual survival as a constant ϕ, and capture rates as year dependent, $p_2, \ldots, p_6$. This approach decreases the model dimensionality from nine to six. Under this approach it is not valid to infer that true annual survival rates in the population are strictly constant. Rather, given the data, the single "average" value, $\hat{\phi}$, is being

recommended as a better (smaller mean square error) estimate of survival in any given year than would a separate estimate of survival, $\hat{\phi}_i$, for that year from a more general model allowing year-dependent survival. The choice of a single $\hat{\phi}$ vs. $\hat{\phi}_1, \ldots, \hat{\phi}_4$ involves a data-based trade-off on the quantity of data vs. the degree of variation in the true $\phi_1, \ldots, \phi_4$. This trade-off is the essence of parsimony.

Model selection philosophy.—Model selection must be guided primarily by knowledge of the biology of the marked populations under study (cf. Box and Jenkins 1970:18–19, Bishop et al. 1975:168, Chatfield 1988: 13–21). Model selection is a problem in applied data analysis and, as such, is outside classical statistical inference theory to a large degree (see, e.g., Gnanadesikan 1983:1–7, 140–141, Durbin 1987:184–185), but uses classical statistical tools of goodness-of-fit (including inspection of residuals) and between-model tests. The goal in model selection is to identify a biologically meaningful model that explains the significant variability in the data (e.g., age-dependent survival for the first 2 yr of life), but excludes unnecessary parameters (Occam's razor), i.e., these additional parameters, hence, inferences, that cannot be justified on the basis of the actual data (cf. Burnham et al. 1987:54–55).

As the amount of data on a population with complex dynamics increases, the model required for analysis and inference tends likewise to increase in dimensionality (i.e., the number of parameters increases). For example, if young, subadults, and adults of both sexes are marked in large numbers, survival is dependent upon winter severity, and capture rates are dependent upon effort, then a relatively sophisticated model with many parameters will be justified and appropriate. If, still further, a "treatment" is imposed, then additional model structure will be required, again increasing the number of parameters in the model.

It is not always possible to decide upon one model that is "best" for the analysis of a given set of data. Rather, several similar, but different, models may be nearly equally applicable to the data. Thus, inferences must contain an additional component of uncertainty. An honest analysis of a complex set of data, therefore, may have to stop short of giving a single "best" model as the end result of the analysis (cf., Bishop et al. 1975: 168, McCullagh and Nelder 1983:17). In this regard we agree with the advice of Chatfield (1988:17), "In particular the choice between models which fit data approximately equally well should be made on grounds external to the data."

Finally, when a model is selected to fit the data from a given experiment, the investigator must remember that it is just a model. That model is not reality; rather, it provides the best representation of the data at hand for the chosen goal. These data may not be refined enough to demonstrate, at statistically significant levels, minor effects that are nonetheless biologically important. For example, the major effect may be age- and sex-specific differences in survival rates; however, additional time variation in survival rate may also occur, but its effect may be too small to be detected convincingly with the available data. One would then select a model with age and sex effects on survival, but no time variation in survival rates, as the best model to describe the given data. As such, it summarizes the statistically significant information in the data and tells what statistical inferences the data justify, not necessarily what full reality is; "minor" effects that are biologically significant may be missed.

Model selection tools

Likelihood ratio tests.—For selecting a model we will make use of the likelihood ratio test (LRT) between models. Use of this test requires that one of the models be a special case of (i.e., "nested" within) the more general model. More specifically, the LRT is used to test a constraint on the parameter vector θ, such as testing the null hypothesis that $\theta_1 = \theta_2$ and $\theta_3 = \theta_4$ (model 2) vs. $\theta_1, \theta_2, \theta_3, \theta_4$ (model 1) where the parameters are unconstrained. An example would be to test the null hypothesis H_0 of no time effects on the ϕ against the general CJS model as the alternative hypothesis H_a, hence H_0: $\phi_1 = \cdots = \phi_{k-2}$ vs. unconstrained $\phi_1, \ldots, \phi_{k-2}$, for any values of $p_2, \ldots, p_k$. For the sake of simplicity we will only use constraints on θ that are linear.

Let θ of dimension np be the parameters of model 1 and let the constrained model, model 2, correspond to θ_0 with r (estimable) parameters. The two corresponding MLEs are $\hat{\theta}$ and $\hat{\theta}_0$, with corresponding log-likelihoods evaluated at the respective MLEs,

$$\ln L(\hat{\theta}) \quad \text{and} \quad \ln L(\hat{\theta}_0).$$

One always has $\ln L(\hat{\theta}) \geq \ln L(\hat{\theta}_0)$. However, random sampling deviations around expected values of model 2 lead to a fit of model 1, which is better in a limited way than that of model 2. Too large a difference between the two log-likelihoods at their maximums indicates that the more general model (model 1) fits the data significantly better. This intuitive reasoning is formalized in standard maximum likelihood theory. Under the hypothesis that the simpler model parameterization θ_0 fits the data just as well as the more complex model corresponding to θ, then the test statistic

$$-2 \ln L(\hat{\theta}_0) - [-2 \ln L(\hat{\theta})]$$

has asymptotically a central chi-square distribution with $np - r$ degrees of freedom. Thus, the test of H_0: $\theta = \theta_0$ vs. H_a: $\theta \neq \theta_0$ is a chi-square test. In examining the relative fit of nested models, we can use this LRT.

An initial global model that would have enough parameters to fit the data exactly is called a *saturated* model and it has 0 degrees of freedom for goodness-of-fit testing. Let $L_S(\theta)$ be the likelihood function for the saturated model. The value of $-\ln L_S(\hat{\theta})$ would then be minimal for the saturated model; all other

models are special cases of the saturated model and will have larger −log-likelihoods. For any other model the quantity $-2 \ln L(\hat{\theta}) - [-2 \ln L_S(\hat{\theta})]$ is usually defined as the deviance (see, e.g., McCullagh and Nelder 1983:17, Hosmer and Lemeshow 1989:14). The LRT is the difference of two deviances. Program SURGE (Clobert et al. 1987, Pradel et al. 1990) prints a "relative deviance" $-2 \ln L(\hat{\theta})$, without the constant $2 \ln L_s(\hat{\theta})$, for any model fitted. For the sake of brevity we will herein call $-2 \ln L(\hat{\theta})$ the deviance, and we denote it as DEV, associated with any model. Thus, SURGE prints DEV. The user must form the difference $DEV_1 - DEV_2$ (larger minus smaller) to compute the LRT if one of the models is a special case of the other model. To summarize, we have, symbolically

$$\text{deviance} = -2 \ln L(\hat{\theta}) + 2 \ln L_S(\hat{\theta}),$$

$$\text{DEV} = -2 \ln L(\hat{\theta}),$$

and

$$\text{LRT} = \text{DEV}_1 - \text{DEV}_2.$$

Goodness-of-fit tests.—If there was only one model to fit to a data set, the only test to be done would be an overall evaluation of how well that model fits the data. Such a goodness-of-fit test for discrete data, such as capture–recapture data, generally takes the symbolic form of

$$\chi^2 = \sum \frac{(O_i - \hat{E}_i)^2}{\hat{E}_i},$$

where O denotes the observed count data and $\hat{E}$ denotes the estimate of this count based on the fitted model. For large samples, this test statistic will be a central chi-square statistic under the null hypothesis that the postulated model fits the data. In principle, one can use an LRT comparing the specific model to the fully parameter-saturated model as a goodness-of-fit test, i.e., use the classic deviance, $-2 \ln L(\hat{\theta}) - [-2 \ln L_S(\theta)]$. This approach has practical problems with capture–recapture data: small expectations induce a need for too much subjective pooling of cell counts and a difficulty of interpreting the result. Therefore, we recommend the use of what Burnham et al. (1987) called TESTS 3 and 2 for goodness-of-fit to the CJS model. Program RELEASE (Burnham et al. 1987) will compute TESTS 2 and 3 given the relevant input capture histories. The pooled chi-square test statistic from TESTS 2 and 3 yields the same test of goodness-of-fit to the CJS model as the tests first developed by Pollock et al. (1985), except for differences induced by the use of different pooling algorithms. One advantage of basing these tests on contingency tables is that exact randomization P levels can be computed, when the data are sparse, thereby eliminating the need for any pooling (see, e.g., Edgington 1980).

TEST 3 is computed as numerous component chi-square tests, each being independent so the component chi-squares are additive. At any release time (at least after occasion 1), there are potentially individuals released with different capture histories. For example, at release occasion 3 we might have histories $h = \{111\}$, $\{101\}$, $\{011\}$, and $\{001\}$. First recaptures then occur at occasions $4, \ldots, k$ for individuals with these four possible histories at release time three. Hypothetical data might be

History	Released	First recapture occasion 4	5	6	7	Never recaptured
{111}	5	1	0	2	0	2
{101}	18	3	2	1	4	8
{011}	24	5	4	4	3	8
{001}	57	8	10	6	9	24

These "data" constitute the subcohorts of the animals released as cohort 3 (i.e., at occasion 3). The component of TEST 3 computable from the data is denoted 3.S3; it would be the chi-square test of homogeneity computed on the 4 × 5 contingency table, conditional on the row and column sums. In this hypothetical example, this component of TEST 3 has 3 × 4 = 12 df. In principle, this same sort of test is computed for the subcohorts of the release at each occasion. The overall TEST 3 is the sum of the chi-squares.

The practical problem is sparse data. This problem motivates a partition of TEST 3.S3 into components that conveniently allow pooling of data. Component test 3.SR3 is based on whether animals were recaptured or not:

History	Recaptured	Never recaptured
{111}	3	2
{101}	10	8
{011}	16	8
{001}	33	24

Further pooling is done if needed. Often, sparse data reduce TEST 3.SR3 to being based on the 2 × 2 table

	Recaptured	Never recaptured
captured before occasion 3	29	18
not captured before occasion 3	33	24

and the associated chi-square test of homogeneity on 1 df.

TEST 3 component 3.Sm3 would be conditional on recapture, hence computed based on the 4 × 4 table of recaptures here (on 9 df):

1	0	2	0
3	2	1	4
5	4	4	3
8	10	6	9

However, often data are so sparse that no test is possible and no TEST 3.Sm3 component exists. The maximal pooling to a 2 × 2 table attempted by RELEASE will, here, produce the table

	Recaptured at occasion 4	Recaptured after occasion 4
captured before occasion 3	8	20
not captured before occasion 3	8	25

The top row of this 2 × 2 table is a pooling over all animals first captured on occasions 1 and 2, whereas the second row is not pooled (as a row) and is the number of new animals captured and released at occasion 3. Similarly, column 1 is not pooled (as a column), while columns 2–4 are animals captured on occasions 5, 6, or 7, rather than on the occasion immediately after release at occasion 3.

These same pooling strategies are used by RELEASE to compute TEST 3.SRi and 3.Smi component chi-squares for cohort releases $i = 2, \ldots, k - 1$. The summed chi-squares and df produce the TEST 3 goodness-of-fit test. For further details see Burnham et al. (1987).

TEST 2 is based on the m_{ij} array (see, e.g., Table 2.b), given releases R_i, and tests that parameters are the same for release at the different occasions $1, 2, \ldots, k$. This array of data is a pooling of the subcohorts data, and hence is not as sparse as the arrays used in TEST 3. TEST 2 is logically equivalent to fitting the CJS model to the data, then computing the estimated expected values $\hat{E}(m_{ij} | R_i)$ given the fitted parameters and doing the usual observed vs. estimated–expected cells chi-square test of fit. RELEASE computes this TEST 2 as a sequence of interrelated contingency tables without first estimating any parameters. The explicit nature of these contingency tables (see Burnham et al. 1987) allows appropriate pooling of cells to account for sparse data. The sum of the TEST 3 and 2 chi-square statistics provides the basis of a general goodness-of-fit test to our global starting model $(\phi_{a \cdot t}, p_{a \cdot t})$ by applying TESTS 3 and 2 on a cohort-by-cohort basis.

There is a TEST 1 also defined in Burnham et al. (1987), which is based on assuming the CJS model applies for two or more data sets of a study, and TEST 1 tests that those data sets have the same survival and capture rates. For example, if a study has controls and treatments, or if the data are partitioned by sex, then TEST 1 is appropriate. This hypothesis of equal parameters under CJS is also computable by SURGE as a likelihood ratio test.

Let θ_G represent parameters of the global model M_G, the most general model to be considered. Assume a goodness-of-fit test of model M_G is available, based on a statistic χ^2_G following a χ^2 distribution with v_G df under the hypothesis that M_G fits the data. Now let θ_1 be a restriction on θ_G; thus, model M_1 is nested in M_G. Furthermore, let model M_2 correspond to yet a further restriction on model M_1, represented by θ_2. Let the corresponding numbers of parameters be $np_G > np_1 > np_2$. Finally, let the LRT chi-square test statistics of model 2 vs. 1 and 1 vs. G be

$$\chi^2_{2\text{vs.}1} \text{ on } np_1 - np_2 \text{ df}$$
$$\chi^2_{1\text{vs.}G} \text{ on } np_G - np_1 \text{ df.}$$

Then the overall goodness-of-fit test of Model M_1 is given by $\chi^2_G + \chi^2_{1\text{vs.}G}$ on $(v_G + np_G - np_1)$ df. The goodness-of-fit test of Model M_2 is given by

$$\chi^2_G + \chi^2_{1\text{vs.}G} + \chi^2_{2\text{vs.}1},$$

on $(v_G + np_G - np_1 + np_1 - np_2) = (v_G + np_G - np_2)$ df. This generalizes in a straightforward way if there were more levels of nested models. For examples of this methodology, see Burnham et al. (1987:160–168).

We recommend this approach to computing the goodness-of-fit test to special cases of the global model. In essence, the goodness-of-fit test statistic to a model is partitioned into informative components associated with a sequence of more general models in order to better understand lack of fit.

Computing test power.—The power of a statistical test, $P(\theta)$, is the probability of rejecting the null hypothesis when the parameters take the value θ, i.e., when the model departs from the null hypothesis in a specified way, noted H_a. It is important to know the power of these tests in planning capture studies and in interpreting the results of capture data analysis (Burnham et al. 1987). Fortunately, it is relatively easy to compute test power analytically, i.e., without resorting to Monte Carlo simulation. We describe the approach below with reference to chi-square tests. This method applies equally to contingency table tests as to likelihood ratio tests.

The formal test involves specifying a significance level, α, often $\alpha = .05$ (α = Type I error rate = the probability of rejecting the null hypothesis when it is true). Then the computed test statistic, χ^2_v for chi-square on v df, is compared with the α critical level from a central chi-square on v df. The power of this test is the probability of rejecting the null hypothesis, H_0, in favor of H_a. The power equals α under H_0. Under the alternative H_a, the test statistic asymptotically has a noncentral chi-square distribution with a noncentrality parameter, λ. There is software available to compute the probabilities for a noncentral chi-square distribution (e.g., SAS).

A simple way exists to compute the asymptotic value of λ under any specific alternative model (see, e.g., Burnham et al. 1987:214–217). It consists of generating

the expected (count) data under the alternative model (H_a) and then analyzing these expected data exactly as if they were real data (Moore 1984, Drost et al. 1989). (To compute λ, the data analysis software must process the data as real numbers, not as integers). The chi-square "test statistic" of H_0 vs. H_a computed from such expected data is really the approximate value of λ for the parameters used to generate those expected data.

In addition to allowing computation of approximate test power, this approach will also produce (approximate) theoretical biases and standard errors of the estimator under the analysis model chosen when the data arise from the true model (H_0) used to generate the expected data. Thus estimator biases, precision, and test power can all be explored analytically rather than by Monte Carlo simulation. We recommend this approach to explore issues of bias, precision, and test power when planning studies and evaluating analysis of real data.

Model selection strategy

Initial considerations.—It is practical in examining capture–recapture data to first compute detailed goodness-of-fit tests and partition these tests to allow full scrutiny of the assumptions basic to further analysis, estimation, and testing. We recommend computing TESTS 2 and 3 (Burnham et al. 1987:71–77) to assess fit. These tests assess, at each point in time, the equality of future recapture histories of previously marked animals vs. those just marked. Other partitioning can be done. For example, TEST 2 can be further partitioned to test for a Markovian dependence in capture probabilities (Sandland and Kirkwood 1981). This partition allows a test of the hypothesis that the capture (or noncapture) at time t is independent of the capture (or noncapture) at time $t - 1$.

A general strategy that we have found useful is to start with a global model with many parameters and a flexible structure. The global model is a place to start, based on sample size, knowledge of the biology of the population, and statistical considerations. Selection of a global model is not a critical statistical step provided this model is general enough so that it includes the "true" model. For a population of marked males, one might select the CJS model as the global model. Alternatively, if one expects that there might be a marking effect on survival the first year after marking, then the global model might be the model of Brownie and Robson (1983) described in Table 7D. We strongly emphasize that biology plays a central role in choosing the global model.

Assuming a global model has been reasonably selected and found to fit the data, then further model selection can be guided by testing between models (i.e., likelihood ratio tests). Again, the candidate models must represent abstraction of biological reality or interest. These tests are possible if models have smaller dimensionality than the global model selected and are nested within the global model. Sequences of tests will often lead to a parsimonious model that is useful for inference, one that has fewer parameters than the (initial) global model. Between-model tests, with a specific alternative hypothesis, frequently have good power, especially if both models are somewhat parsimonious. In either case, the power can be computed fairly easily (Burnham et al. 1987:214–217). It is desired to know the power of tests as a function of sample size when planning research and when interpreting the results of hypothesis tests (Peterman 1990).

Finally, after an adequate model has been selected, it is prudent to check neighboring models. For example, perhaps an early test result suggested that parameters for adult males and females were similar and, therefore, should be pooled over sex for analysis. Then, likelihood ratio tests were used to select a parsimonious model with, say, only six parameters. At this end point, one might want to recheck to confirm that the hypothesis of a sex effect was still unsupported.

If large numbers of animals have been marked or the capture probabilities are high, it may be possible to select a model and estimate model parameters on about half of the data and use the remaining data for testing the goodness-of-fit of the model (cross validation). This practice is especially useful when functional relationships on model parameters have been specified.

The goodness-of-fit tests have been found to have low power with general models for sample sizes often seen in studies of marked animal populations (see Pollock et al. 1985, Burnham et al. 1987). The likelihood ratio test (LRT) will commonly have low power between nonparsimonious models used at the beginning of the model selection procedure (Goodman 1984:34–36). Thus, one might wish to use P values of .1 instead of .05. This would provide more powerful tests at the expense of more Type I errors. This trade-off seems reasonable here; however, it may lead to difficulties in complex data sets involving age, sex, year, and treatment effect acting, possibly, on both the survival and capture probabilities. Of course, we do not subscribe to the use of an exact a priori P value (e.g., .05) to assess significance strictly.

When model selection is used in capture–recapture via numerous tests of significance involving alternative models there is substantial risk of error: both Type I (reject a true null H_0) and Type II (fail to reject false H_0). The problems of arriving at a "correct" model are the same as encountered in variable selection in regression (see, e.g., Flack and Chang 1987, Rawlings 1988:Chapter 7). Persons seriously doing model selection with complex animal marking data should be aware of these problems and pitfalls in over- and under-fitting models to data. One recommendation from the regression literature that we think applies here is to use an α level of .15 as a basis for rejecting a model (Flack and Chang 1987:85, Rawlings 1988:185). As noted below, however, model selection need not be based only on formal tests of significance

More on model selection.—Model selection has been approached by starting with a simple, low-dimension model and increasing the dimensionality (step-up methods as in least-squares models). This approach is not preferred, as the first, simple model may not fit, which then may produce misleading tests. We thus prefer to begin with a fully parameterized model that fits the data and decrease the dimensionality toward a more parsimonious model that is supported by the data (step-down methods). Frequently, these two methods do not converge to the same model for a given set of data.

In a step-down approach, one can first test and model the capture probabilities while keeping the dimensionality of the survival probabilities high and fixed. Alternatively, one could consider various reduced parameterizations of the survival probabilities, while keeping the dimensionality of the capture probabilities high and fixed. Again, these two approaches may not converge to the same model for a particular data set. We slightly favor modeling capture probabilities first in order to keep as much power as possible for tests on survival parameters, which are the parameters of most biological interest.

The LRTs for these approaches are not strictly hypothesis tests; rather, they compare two alternative models as part of a procedure to select one or several suitable models for the data. Thus, as multiple tests are made for the same data set, the effective α level of the tests is greater than the nominal α level for a single test. This complication makes model selection difficult in complex situations. This dilemma is common to all areas of data analysis (Box and Jenkins 1970:17–19, Searle 1971, Bishop et al. 1975:155–168 and 311–327, Fienberg 1980:56–80, Draper and Smith 1981:299–302, Gnanadesikan 1983:1–7, McCullagh and Nelder 1983:1–5, Goodman 1984:34–36, Linhart and Zucchini 1986:1–24, and Chatfield 1988:5–21).

In theory, a superior solution to model selection is to put the model selection problem into a function optimization framework. In fact, likelihood theory has been extended by Akaike (1973) to put the MLE and model selection in one framework. Akaike's procedure consists of computing the value of the log-likelihood at its maximum for a given model, plus the number of estimable parameters in the model. This quantity, called Akaike's Information Criterion (AIC), is usually defined as

$$\text{AIC} = -2 \ln L + 2 \times (\text{number of parameters}).$$

For each model, one computes the AIC; this calculation is trivial after computing the MLE for the model because computing the MLEs yields the $\ln(L)$. One then selects the model where AIC is smallest. This selection leads to a reasonable model for a particular data set. The likelihood ratio test strategy would lead one to consider $-2 \ln(L)$ + number of parameters as a model selection criterion. However, the effect of the optimization inherent in likelihood maximization increases with the number of parameters. Correcting for this leads one to consider AIC as an overall criterion for model selection.

The AIC has been used in data analysis (i.e., model selection) and found to have good properties (see, e.g., Linhart and Zucchini 1986:198–206, Sakamoto et al. 1986, especially Chapter 4, Bonney 1987, Read and Cressie 1988:124–128). Moreover, AIC is essentially identical to the much used Mallow's C_p statistic for model selection in regression (see, e.g., Draper and Smith 1981:299–302), and Shibata (1989:222–225) shows a clear relationship between AIC and cross-validation. It is known that Mallow's C_p is a satisfactory criterion for preliminary model selection, as is cross-validation (Stone 1974). We will compare the performances of LRT- based step-down model selection and AIC-based model selection in the examples that follow.

Extra-multinomial variation.—The conceptual basis of all models is that the data = structure + residual (stochastic variation). The average residual variation is summarized by a parameter, σ^2, which is often not known. By assuming a multinomial model for the data, we are, among other things, assuming we know σ^2. In effect, if the model structure is correct, then under the multinomial assumption $\sigma^2 = 1$ (where σ^2 is defined as the expected value of the chi-square goodness-of-fit statistic divided by its degrees of freedom).

However, the structural part of the model can hold even if the multinomial variation fails, due to excess variance. The statistical literature has long recognized this problem (see, e.g., Williams 1982, Moore 1987), and one approach to a solution is founded in quasi-likelihood theory (see, e.g., Wedderburn 1974, McCullagh and Nelder 1983, Burnham et al. 1987:243–246. If there is excess variation it will show up in the model goodness-of-fit chi-square statistic (this statistic is a type of residual sum of squares). We recommend use of some type of empirical variances if there is substantial excess variation.

The simplest approach is to estimate a variance inflation factor (Finney 1971, McCullagh and Pregibon 1985, Burnham 1987:243–246), $\hat{c}$. For the selected model, let χ^2 by the chi-square goodness-of-fit statistic to this model, on v df. Then if it is judged that excess variation exists (e.g., if χ^2 is large for its df), compute $\hat{c} = \chi^2/v$ and multiply each theoretical standard error, $\text{SE}(\hat{\theta})$ (for any parameter, θ), by $\hat{c}$. Then treat the empirical estimator, $\widehat{\text{SE}}(\hat{\theta}) = \sqrt{\hat{c}}\,\text{SE}(\hat{\theta})$ as being based on v df, for example, in constructing a confidence interval on θ. This is a common procedure and is recommended for general use, see for example Cox and Snell (1989: 111–115). Finally, it is worth noting here that we expect $\hat{c}$ not to exceed ≈ 3; large values of $\hat{c}$ like 10 almost certainly indicate that our model structure is grossly inadequate and we should continue our search for a better model.

Other modifications to theory are probably needed

if excess variation exists. In priniciple, the LRTs should be modified, as should the AIC criterion. These matters, which need further work, show up in our last example (Greater Flamingo).

When $\hat{c}$ is significantly greater than 1 (by the chi-square test) there is no way to determine from the data if this represents true excess variation and the model structure is correct, or if the multinomial variance assumption is true and $\hat{c}$ represents a failure of the model used to account for all the subtle structure in the data. In fact, we cannot reasonably expect any of our model structures to be exactly correct. A quotation attributed to G. E. P. Box (see Chatfield 1988:15) is relevant here: "All models are wrong, but some are useful." We approach data analysis in this spirit: we want to find a useful model that correctly represents the biologically important structure that really is in the data. We may be unable to ferret out the correct form of the more subtle structure in the data. In this case, we believe it is appropriate to "sweep" this residual structure into the model error component and inflate our standard errors on estimates to reflect properly our degree of uncertainty about the true parameter values.

Discussion

Modeling open-population animal-marking data has now reached the point where we know how to write down models, as long as we condition on first release (hence ignore estimation of abundance). These models can be expressed in various ways; multinomial models are especially convenient with the cells being functions of survival, ϕ, and capture probabilities, p. Statistical theory is well developed for handling a given, prespecified model (as regards MLE of parameters, goodness-of-fit, confidence intervals). Statistical tools are also well developed for model selection (e.g., LRT, AIC, cross-validation). However, what is still needed is to develop and explore model selection strategies.

The properties of complex model selection for capture–recapture data have not been studied extensively. Unsolved problems result when a data set (a capture history matrix) is used both to select a model (e.g., use of AIC) and estimate the uncertainty in the parameter estimators, e.g., $\widehat{\mathrm{SE}}(\hat{\theta})$. That is, the use of the same data set for both structural identification and inference is problematic. Here, the estimated standard errors tend to be too small and the confidence intervals tend to be too narrow, often resulting in poor coverage (see Hurvich and Tsai [1990] for an example of this matter in regression). While a growing literature on this problem exists for least squares analyses (see Dijkstra 1988), little information is available for complex multinomial models. We do not know of a solution to this problem; our preliminary feeling is that the effect on measures of precision is not a major problem in capture–recapture. These are areas in need of more research by statisticians. Nonetheless, our message for biologists is that an analysis theory for capture data is very advanced and much sound science can now be done with animal-marking studies.

Model selection is now a critical aspect of the data analysis, as this process leads to answering the question of what are the important sources of variation in the data (such as control-treatment effects, or time effects on survival rates, or age effects). It may be that the actual parameter estimates are not as important as identifying the significant biological processes in the study population; the processes are more fundamental than actual parameter values, especially average parameter values.

Given this philosophy, the fitting of simplistic models for the purpose of estimating average survival rates is seen to be a great waste of information (unless the data are so sparse that only such a simple model is justified by parsimony). Overly simple models may produce the same average values as biologically realistic models (see, e.g., Spaepen 1988); this is not a good reason to use simple models and thereby lose all the really interesting information in the data.

Capture analysis programs

In Table 9 we summarize important software developed during the past 10 yr. These programs all represent the results of important research publications during the last decade; however, some of these programs are more developed, user friendly, and reliable than others. Most of this software is in the form of stand-alone programs, although in the case of SAS and GLIM the efforts have been to show how these standard, major statistical software packages can be used to analyze data under specialized capture–recapture models.

Also included in Table 9 are both software for the analysis of open-population capture–recapture data (CJS-type models), for recovery data (e.g., band recovery data), and closed population capture–recapture data. We decided to give the information in Table 9 about the programs, even though our judgments often represent personal opinions (but based on our experience). Most analyses of capture–recapture data that a biologist might need can be accomplished using one or more of these programs.

Example 1: European Dipper

Introduction

There are four species of dippers (*Cinclus* spp.) recognized in the world. They inhabit mountain streams, feeding on underwater invertebrates. Their livelihood is closely dependent on streams; their nests of moss are always close to water, sometimes hidden behind waterfalls. Capture–recapture data on the European Dipper (*Cinclus cinclus*) were collected for 7 yr (k = 7) (1981–1987) by G. Marzolin in eastern France (Marzolin 1988). The data consist of marking and recaptures of breeding adults each year during the breeding period from early March to 1 June. Birds were at least 1 yr

TABLE 9. A summary of important software developed during the 1980s for the analysis of animal marking studies. Presence of a particular attribute is denoted by a "+."

	Program name	Key reference	Good documentation	Easy to use	Interactive Input (I), Noninteractive (N), Both (B)	Available without cost	Recommended for general use	Mainframe (M), PC (P), Both (B)	Object code (O), Source (S), Both (B)	Specific software for capture-recapture	Open population models (O), closed pop. (C), Both (B)	Live recapture (L), Recovery data (R), Both (B)	Single data set (S), Multiple data sets (M)	Covariates	User-defined model	Transformations	Constraints: Range (R), Equality (E), Both (B)	Population numbers (N), Survival rates (S), Both (B)	Age-specific parameters	Time-specific parameters	Very general models	Extensive testing available	Treatment/control experiments	Input: CH-matrix (C), m-Array (M), Both (B)	Useful only by statisticians	Simulation capability	Power computation	Bias-adjusted estimates
1	SURGE 4.0	Pradel et al. 1991	+	+	I	+	+	P	B	+	O	B	M	+		+	B	S	+	+	+	+	+	B				
2	RELEASE	Burnham et al. 1987	+	+	B	+	+	P	B	+	O	B	M	+				S		+		+	+	B		+	+	+
3	JOLLY	Pollock et al. 1990	+	+	N	+	+	P	B	+	O	L	S					B		+		+		B				+
4	JOLLYAGE	Pollock et al. 1990	+	+	N	+	+	P	B	+	O	L	S					B	+	+		+		B				+
5	SURVIV	White 1983, 1986			N	+		B	B	+	O	B	M	+	+	+	B	S	+	+	+	+	+	M		+	+	
6	PROPAN3	Arnason & Schwarz 1986	+	+			+	M	O	+	B	L	S	+			B	B	+	+		+		B		+		+
7	RECAPCO	Buckland 1980			N	+		B	S	+	O	B	S			+	R	B		+				C		+		
8	JD	Jolly & Dickson 1980			N	+		M	S	+	O	L	S				E	B		+		+		?				
9	CMR	Crosbie & Manly 1985			N	+		M	S	+	O	L	S				E	B	+	+		+		C				
10	ESTIMATE	Brownie et al. 1985	+	+	B	+	+	B	B	+	O	R	S				E	S		+		+		M				+
11	BROWNIE	Brownie et al. 1985	+	+	B	+	+	B	B	+	O	R	M				E	S	+	+		+		M				+
12	MULT	Conroy et al. 1989	+	+	B	+	+	B	B	+	O	R	S	+			E	S	+	+		+		M				
13	SCHWARZ	Schwarz et al. 1988			N	+		M	S	+	O	R	S				?	S		+		+		M				
14	CAPTURE	White et al. 1982	+	+	N	+	+	B	B	+	C	L	S				E	N		+		+		B		+		
15	CONTRAST	Sauer & Williams 1989	+	+	B	+	+	P	B			B																
16	SAS	Burnham 1989			N			B	O		B	B	M	+	+	+	B	S	+	+	+		+	M	+			
17	GLIM	Cormack 1981, 1985, 1989			B			B	O		B	L	M		+	+	B	B	+	+	+	+	+	M	+			

old when initially banded. A total of 294 birds were marked and a summary of these data, by sex, is shown in Table 10.

Initial model

Known plumage criteria of European Dippers were used to differentiate 1-yr-old breeders from older birds (Glutz Von Blotzheim and Bauer 1985). However, previous work on the influence of age during the breeding period is scarce, and does not point to age specificity in survival rates. On the contrary, severity of winter (e.g., frozen rivers, flooding of nests) is expected to play a crucial role (Glutz Von Blotzheim and Bauer 1985). Furthermore, because the data are limited in number, we choose to start with a time-dependent model, by sex, for both survival and capture rates (i.e., model $\phi_{s\cdot t}$, $p_{s\cdot t}$). A goodness-of-fit test of these data to the Jolly-Seber model, separately by sex, is given below (TESTS 2 and 3, Burnham et al. 1987:64–77):

Sex	χ^2	df	P
Male	6.78	5	.2375
Female	4.98	5	.4183
Total	11.76	10	.3014

The m arrays in Table 10 show that the capture probabilities were fairly high, because approximately half of the birds were recaptured the 1st yr after banding. When the capture probabilities are near 1, then off-diagonal elements in the m arrays (cf. Table 10) are often close to zero. For this reason, the 3.Sm components (Burnham et al. 1987:74–77) of TEST 3 and all of TEST 2 are not informative in this example. Thus, in this case, the information relating to goodness of fit is contained in the 3.SR component of TEST 3 (shown in the table above). The results of these tests suggest that the basic assumptions of the Jolly-Seber model are met (i.e., releases of previously marked animals

TABLE 10. Capture–recapture data for male and female European Dippers (*Cinclus cinclus*) banded 1981–1987 in eastern France (from Marzolin 1988 and G. Marzolin, *unpublished data*).

			m_{ij}						Never recaptured ($R_i - r_i = R_i - \sum_j m_{ij}$)
i	R_i	$j =$	2	3	4	5	6	7	
Males									
1	12		6	1	0	0	0	0	5
2	26			11	0	0	0	0	15
3	37				17	1	0	0	19
4	39					22	0	1	16
5	45						25	0	20
6	48							28	20
Females									
1	10		5	1	0	0	0	0	4
2	34			13	1	0	0	0	20
3	41				17	1	0	0	23
4	41					23	1	1	16
5	43						26	0	17
6	50							24	26

have similar parameters [fates] as newly released animals, part of the **iii** hypothesis). Having small off-diagonal elements is actually an advantage because this means capture rates are near 1 and precision of survival estimates is maximized for the number of animals marked.

Further models

One hypothesis related to whether males and females have different parameters, ϕ_i and p_i. A test of this hypothesis can be computed from the theory presented in Brownie et al. (1985:144–152) or, equivalently, in Burnham et al. (1987:64–77), or using program SURGE. We used program RELEASE (Burnham et al. 1987) to compute TEST 1 and found $\chi^2_{11} = 2.61$ and $P = .9949$. Thus, the data for males and females can be pooled for further analysis and, based on the principle of parsimony, such pooling is advantageous. Because the chi-square value is small, there is little merit in examining models such as (ϕ_{s*t}, p_t) where ϕ but not p is sex specific as well as time specific. The chi-square value for the pooled data is so low that some dependence between males and females is suggested. From the experimental design and field experience, we know that when a nest is occupied, capture of both the male and the female is nearly certain (i.e., the probability of capture approaches 1). In contrast, when a nest is deserted, capture of either mate is unlikely, which might explain the dependence, at least for the capture probabilities in the model. More refined analyses should then be based on only one sex. Nevertheless, we will go on with data for both sexes as a first illustration of capture–recapture modeling for multiple data sets.

For the pooled data, estimates of the capture and annual survival probabilities under the model (ϕ_t, p_t) (CJS model) are shown in Table 11 along with their estimated asymptotic standard errors. Estimated annual survival varied from 0.43 to 0.72. Estimated capture rates were >0.9, except in the 2nd yr (the 1st yr of recapture), when $\hat{p}_2 = 0.70$. (However, $\widehat{\text{SE}}(\hat{p}_2)$ was so large that this estimate is not inconsistent with the other estimates.) In the pursuit of parsimony, models with fewer parameters were investigated, particularly because the sample size, even for the pooled data, is fairly small. Model (ϕ_t, p) described the pooled data as well as the CJS model (LRT, $\chi^2_4 = 2.78$, $P = .5953$). Thus, this model, with constant capture probabilities and year-dependent survival rates (ϕ_t, p) is to be preferred here to the CJS model.

Furthermore, a model with a constant survival rate and a constant capture rate (ϕ, p) is satisfactory (LRT, $\chi^2_5 = 7.11$, $P = .2126$, and goodness of fit $\chi^2_{34} = 33.03$ $P = .5150$) (Table 12). Thus, a two-parameter model yields the estimates $\hat{\phi} = 0.56$ [$\widehat{\text{SE}}(\hat{\phi}) = 0.025$] and $\hat{p} = 0.90$ [$\widehat{\text{SE}}(\hat{p}) = 0.029$]. A considerable increase in precision of the estimators is achieved by using this reduced model, and the tests of fit indicate that this model is statistically acceptable for these data.

Further modeling is possible to attain increased biological reality. It was known before this analysis that a major flood occurred during the 1983 breeding season. Because captures during this breeding season occurred well before and after the flood, survival in the two year 1982–1983 and 1983–1984 was likely to be affected. Thus, it seems reasonable to hypothesize that survival of species living along and feeding in the river in those two years was lower than in nonflood years. Estimates of survival in flood (ϕ_f) and nonflood (ϕ_n) years were computed under the assumption of a constant capture probability p (using SURGE with dummy variable constraints). Thus, we denote the model with three parameters (ϕ_f, ϕ_n, p), where survival rates are different for flood vs. nonflood years, as model (ϕ_{fn}, p). As might be expected from the estimates in Table 11, this three-parameter model fits significantly better than the two-parameter model (LRT $\chi^2_1 = 6.74$, $P = .0094$

TABLE 11. Summary of estimates of annual survival (ϕ) and capture (p) probabilities for the pooled data on male and female European Dippers (*Cinclus cinclus*). The estimates are made under the CJS model (ϕ_t, p_t).

i	Year	$\hat{\phi}$	$\widehat{\text{SE}}(\hat{\phi})$	$\hat{p}$	$\widehat{\text{SE}}(\hat{p})$
1	1981–1982	0.718	0.155	...	...
2	1982–1983	0.435	0.069	0.696	0.166
3	1983–1984	0.478	0.060	0.923	0.073
4	1984–1985	0.626	0.059	0.913	0.058
5	1985–1986	0.599	0.056	0.901	0.054
6	1986–1987	...	...	0.932	0.046

TABLE 12. Summary of information used in model selection for the capture–recapture data on the European Dipper (*Cinclus cinclus*). Data for males and females are pooled. For model notation, see the Appendix.

Model	No. parameters, *np*	DEV*	LRT χ^2	df	*P*	GOF χ^2	df	*P*	AIC†
(ϕ_t, p_t)	11	656.95	2.78	4	.5953	11.76‡	10	.3014	678.95
(ϕ_t, p)	7	659.73	0.37	4	.9849	14.54§	14	.4103	673.73
(ϕ_{fn}, p)	3	660.10	6.74	1	.0094	14.91§	18	.6681	**666.10**
(ϕ, p)	2	666.84‖				21.65§	19	.3020	670.84

* Deviance, defined up to an additive constant.
† Akaike Information Criterion, AIC = DEV + 2*np*.
‡ Result of TEST 2 and TEST 3 chi-squares.
§ Obtained by adding the difference in deviance to the GOF chi-square of model (ϕ_t, p_t); for example, 14.54 = 2.78 + 11.76.
‖ LRT for model (ϕ_t, p) vs. (ϕ, p): $\chi^2_5 = 7.11$, $P = .2126$.

for testing model $[\phi_{fn}, p]$ vs. model $[\phi, p]$). The MLEs are $\hat{\phi}_f = 0.469$ $[\widehat{\text{SE}}(\hat{\phi}_f) = 0.043]$, $\hat{\phi}_n = 0.607$ $[\widehat{\text{SE}}(\hat{\phi}_n) = 0.031]$, and $\hat{p} = 0.900$ $[\widehat{\text{SE}}(\hat{p}) = 0.029]$. This illustrates the value of a priori biological information in model building and inference; without this information we would have been satisfied with the simpler two-parameter model.

Rather than assume asymptotic normality in setting the confidence limits as $\hat{\theta} \pm 1.96\ \widehat{\text{SE}}(\hat{\theta})$, an alternative method is given. With small sample sizes, the distribution of the estimates is skewed to the right; thus, a confidence interval based on the lognormal distribution can be expected to have better properties. The procedure used here is from Burnham et al. (1987:212), and computes an approximate $(1 - \alpha)100\%$ CI using the lower and upper bounds, $\hat{\theta}_l$ and $\hat{\theta}_u$, respectively, as

$$\hat{\theta}_l = \hat{\theta}_u / C \quad \text{and} \quad \hat{\theta}_u = \hat{\theta} \cdot C$$

where

$$C = \exp\left\{z_{\alpha/2}\sqrt{\ln(1 + [\text{CV}(\theta)]^2)}\right\}.$$

The 95% confidence intervals for the three model parameters ϕ_f, ϕ_n, and p are [0.385, 0.553], [0.549, 0.671], and [0.845, 0.959], respectively. These asymmetric confidence intervals have better achieved coverage than the usual procedure based on untransformed estimates.

The difference with the usual symmetric confidence intervals will be negligible for small coefficients of variation $\text{CV}(\theta)$. This is the case here, because of the parsimony of the final model, even for those parameters close to 1. (Symmetric confidence intervals for $\hat{\phi}_f$, $\hat{\phi}_n$, $\hat{p}$, respectively: [0.385, 0.553], [0.548, 0.668], [0.843, 0.957].)

After selecting model (ϕ_{fn}, p) as a basis for inference, it might be appropriate to retest for sex differences. The power of this final test is increased because fewer parameters (three) are involved. We tested for sex specificity in the three parameters, but found none (LRT, $\chi^2_3 = 0.66$, $P = .8826$).

We considered Akaike's Information Criterion (AIC) for the models involved. This criterion also supports the use of the (ϕ_{fn}, p) with constant survival rates for flood and nonflood years and constant capture rate (i.e., model $[\phi_{fn}, p]$ with AIC = 666.10). The other models had larger AIC values ($[\phi_t, p_t]$: 678.95, $[\phi_t, p]$: 673.73, and $[\phi, p]$: 670.84). When there are only a few models to consider (four here), and especially if they are nested models, then the use of LRTs can be very effective in selecting the appropriate model. Hence we used both approaches here to demonstrate that they agree here, as they often do. However, in complex situations with many models from which to select, AIC will be much faster and easier to use and avoids some of the problems inherent in making multiple tests.

Discussion

The European Dipper example illustrates several general points and three basic results. First, the questions of goodness-of-fit to the Cormack-Jolly-Seber model (ϕ_t, p_t) were addressed by TEST 2 and TEST 3 for each sex. Second, data for males and females could be pooled for further analysis. Third, no variation in survival or capture parameters could be detected using the usual nonspecific likelihood ratio tests. However, the most important point was that, using a priori biological information to formulate an hypothesis, survival in flood years was shown to be significantly lower than in more normal years using a corresponding specific likelihood ratio test. This last finding illustrates the value of specific alternative hypotheses if the a priori basis exists. Program CONTRAST (Hines and Sauer 1989) might also be useful in this context.

The differences in survival could include emigration, particularly in flood years. However, the study area covered $\approx$2200 km^2 (Marzolin 1988), and any such emigration, if it occurred, would be extensive and permanent. We say this because the flood years were early

TABLE 13. A summary of Swift (*Apus apus*) capture data (G. Gory, *unpublished data*). Birds are classified by columns according to time of next recapture, and by rows according to year of release. Newly marked birds and already banded birds released in the same year are thus pooled in each row. Note the uneven distribution of the dates of first recaptures for the birds released in 1981 in the Good colony (this cohort was discarded for survival analysis).

Released		Year next recaptured								Never
Year	Number	1982	1983	1984	1985	1986	1987	1988	1989	recaptured
		Recaptures, m_{ij}								
Poor colony										
1981	11	1	2	0	0	0	0	0	0	8
1982	7		4	0	0	0	0	0	0	3
1983	14			6	0	1	1	0	0	6
1984	6				3	0	0	0	0	3
1985	7					1	1	1	0	4
1986	4						2	0	0	2
1987	5							3	0	2
1988	11								2	9
Good colony										
1981	12	3	0	4	0	0	0	0	0	5
1982	13		10	2	0	0	0	0	0	1
1983	24			13	1	1	0	1	0	8
1984	25				13	4	0	0	0	8
1985	22					15	2	0	0	5
1986	27						20	0	0	7
1987	29							19	1	9
1988	43								15	28

in the study period and therefore there was ample opportunity for recapture of returning emigrants after the floods. This analysis indicates that annual survival of adult dippers in eastern France averages ≈0.60 in normal years, but can be substantially less in flood years (see Glutz Von Blotzheim and Bauer 1985). Galbraith and Tyler (1982), working on color-banded birds, estimated survival of adults at 0.54, but raised this estimate to ≈0.55–0.60 after making a correction for emigration.

EXAMPLE 2: SWIFT

Introduction

The Swift (*Apus apus*) is an abundant and noisy colonial breeder in Asian and European cities. It nests from March to July, nearly exclusively in buildings (see Cramp 1985). The Swifts breeding in the Museum of Natural History of Nîmes, in southern France, have been studied since 1981 (G. Gory, *unpublished data*). This study of Swifts attempts to understand the processes of colony and nest site selection, both for fundamental reasons (regulation of a population in an heterogeneous environment) and for applied reasons (defining the characteristics of artificial nest sites to be built in modern buildings).

Breeding birds were captured by hand at night from inside the building, using prebuilt accesses to the holes used by Swifts for nests. Nonbreeding birds were not captured because they do not remain at night on the nest. Homogeneity of capture rates of birds is thus expected, in particular with respect to age. Some irregularities in checking sites occurred in the first few years of the study. Birds could not be sexed reliably in the field. Captures of breeding birds were concentrated at the chick stage to ensure maximal capture rate and minimize possible nest desertion.

In this example we use data from two neighboring colonies. One colony (hereafter designated G, for Good) faced west and was in a cloister closed to the public. It was a priori a good site, protected against the mistral, a cold northerly wind frequent in southern France in spring. The other colony (hereafter designated P, for Poor) faced south and was of poorer quality because it faced a busy boulevard and was subject to swirling winds. Swifts are extremely faithful to the colonies where they breed (various observations summarized by Cramp 1985), and only one example of a bird temporarily changing colonies was observed in our data. For the sake of simplicity, that bird was assigned to its original colony. The differences in quality between colonies could induce differences in the survival rate (through the quality of birds selecting each site as well as through direct effects) and in the capture rate (through differences in the rate of presence at nests). Moreover, differences between colonies in accessibility to nests for recaptures can also influence capture rates.

We used birds banded from 1981 to 1988 and recaptures from 1982 to 1989 in this example ($k = 9$). The data consisted of 127 individuals which belonged to one of two groups (88 and 39 for G and P, respectively) over 9 yr. The capture rates were fairly high, as most of the birds were recaptured in the 1st yr after banding, as seen from a summary of the data (Table 13).

Our basic objective is to test for differences between

colonies in survival and capture rates and estimate the parameters. In model notation, colony will be noted c, with two categories, G and P.

Basic tests and models

We begin with model $(\phi_{c\cdot t}, p_{c\cdot t})$, i.e., the Cormack-Jolly-Seber model considered separately for each colony. TEST 3 (Program RELEASE; Burnham et al. 1987) did not indicate heterogeneity of fates within groups $\chi^2_{20} = 21.34$, $P = .3774$, with sparse data. The results of TEST 2 (Burnham et al. 1987), although showing heterogeneity, could not be interpreted, as none of the nine calculable components had enough data. Attempts to fit age-dependent models failed to reduce the heterogeneity, which was thus within cohorts much more than between cohorts. This indicated that the **iii** hypothesis was not fully met. The first marked cohort was deleted because of small sample size. Restricting the analysis to capture years 1982–1988 (with recaptures in 1983–1989, i.e., $k = 8$) results in accepting the CJS model for both colonies (TEST 3: $\chi^2_{16} = 17.58$, $P = .3491$; TEST 2: $\chi^2_8 = 11.78$, $P = .1613$, this value resulting from several low expected numbers).

The deviance for model $(\phi_{c\cdot t}, p_{c\cdot t})$ (now with $k = 8$ yr), using SURGE, was 339.705, and all 26 parameters ($np = 2[2k - 3] = 26$) (ϕ, p, and β) were estimable. An overall colony effect did not seem to be present, whether quantified by a contingency table χ^2 (TEST 1, Burnham et al. 1987, for group effect; RELEASE: $\chi^2_{13} = 16.30$, $P = .2333$), or alternatively by a likelihood ratio test between models $(\phi_{c\cdot t}, p_{c\cdot t})$ and (ϕ_t, p_t), with respective number of parameters 26 and 13 (SURGE, $\chi^2_{13} = 15.24$, $P = .2926$). These two tests use the same data and are asymptotically equivalent.

However, the power of these overall tests to detect differences in survival is limited with such sparse data. The power of the test to distinguish between survival rates of 0.6 and 0.8 for colonies P and G, respectively, and with $p = 0.7$ (constant over the years and across the colonies), calculated using option EXPECT in RELEASE (Burnham et al. 1987:214–217), is only 0.0997 for $\alpha = .05$. It took 5 s of CPU time to run RELEASE (option EXPECT) to compute this power value. By comparison, using Monte Carlo simulation, 1000 repetitions, RELEASE took 398 s of CPU time to estimate the power as 0.0760 (SE = 0.0084). Either approach to computing power is reliable, but it is more practical to compute power analytically when numerous scenarios are to be examined, because simulation evaluation could then take days, compared with minutes for analytical evaluation.

The following analysis partitions the likelihood ratio test statistic for a colony effect into components, starting first with tests on the capture rate, concentrating on the comparison of survival rates after having reached as much parsimony as possible in capture rates. The sequence of models considered is summarized in Table 14.

Modeling capture rate.—The first attempt towards parsimony for capture rate consisted of dropping the time × colony interaction, i.e., to hypothesize parallelism over time of capture rates between colonies, on a logit scale. Anderson et al. (1980:164–165) provide an illustration of parallelism on a logit scale. This parallelism hypothesis is compatible with the design, variations in effort over time as a result of practical constraints having affected both colonies in a comparable way. This parallelism hypothesis seemed readily acceptable (model $[\phi_{c\cdot t}, p_{c\cdot t}]$ vs. model $[\phi_{c\cdot t}, p_{c+t}]$, $\chi^2_5 = 4.44$, $P = .4873$; Table 14). The effect of parallelism in the logit scale on capture rates in their original scale (i.e., not transformed to logits) is illustrated in Fig. 1, using estimates from model (ϕ_c, p_{c+t}).

Removing either time or colony effect or both seemed, on a first analysis, to be possible (model $[\phi_{c\cdot t}, p_c]$ vs. model $[\phi_{c\cdot t}, p_{c+t}]$, $\chi^2_5 = 7.817$, $P = .1666$; model $[\phi_{c\cdot t}, p_t]$ vs. model $[\phi_{c\cdot t}, p_{c+t}]$: $\chi^2_1 = 2.619$, $P = .1056$; and model $[\phi_{c\cdot t}, p]$ vs. model $[\phi_{c\cdot t}, p_{c+t}]$: $\chi^2_6 = 9.766$, $P = .1349$). The P levels from these tests were close to the $\alpha = .05$ level; however, we recommend a less stringent α level (e.g., .1 or .15) in judging the significance of factors known a priori to be of likely biological significance. Colony effect, when looked for after having accepted model $(\phi_{c\cdot t}, p_c)$, was also close to significance (model $[\phi_{c\cdot t}, p]$ vs. model $[\phi_{c\cdot t}, p_c]$: $\chi^2_1 = 1.949$, $P = .1627$). The estimates of p_g (0.775) and p_p (0.584) obtained from model $(\phi_{c\cdot t}, p_c)$ differed as expected (*G*ood better than *P*oor). The one-sided test computed from the above χ^2 suggests the difference may be real ($P = .1627$ [two-sided test] or $= .0813$ [one-sided test]). The asymptotically equivalent test based on a direct one-sided comparison of estimates was closer to the 5% significance level ($[\text{logit}\,\hat{p}_g - \text{logit}\,\hat{p}_p]/\widehat{\text{SE}} = 1.485$, $P = .0688$). Given this limited statistical evidence, the known low power of these tests in this study, and the a priori information on the design, it seemed preferable not to bias further tests on survival by using too simple a model for the capture rates. Hence, we will keep model $(\phi_{c\cdot t}, p_{c+t})$, in accordance with the general strategy outlined in *Model selection.* Further tests will be done after modeling survival in a more parsimonious way.

Modeling survival rate.—The term for time × colony interaction could be dropped for survival (model $[\phi_{c+t}, p_{c+t}]$ vs. model $[\phi_{c\cdot t}, p_{c+t}]$, $\chi^2_5 = 1.780$, $P = .8787$). Then the time effect was not significant (model $[\phi_{c+t}, p_{c+t}]$ vs. model $[\phi_c, p_{c+t}]$, $\chi^2_6 = 2.273$, $P = .8930$). The comparison of average survival rates between the two colonies was done by comparing the two parsimonious models: (ϕ_c, p_{c+t}) and (ϕ, p_{c+t}). This comparison suggests that average survival may be better in the Good colony ($\chi^2_1 = 2.580$, $P = .0541$ for a one-sided test). This test is asymptotically equivalent to a direct com-

TABLE 14. Results of capture–recapture models and of comparisons between models for the Swift (*Apus apus*) (data from G. Gory, *unpublished data*). For each model we give the number of estimable parameters (*np*), the deviance (DEV), and the Akaike Information Criterion (AIC). For model notation, see the Appendix.

Model	*np*	DEV	AIC	Comparison
I. Basic models				
(1) $(\phi_{c\cdot t}, p_{c\cdot t})$ (CJS model by colony)	26	339.705	391.705	(fits the data)
				Overall diff. between colonies (2) vs. (1): $\chi^2_{13} = 15.239$, $P = .2927$
(2) (ϕ_t, p_t) (CJS model for pooled data)	13	354.944	380.944	
II. Modeling capture rate				
				Parallelism over time on p (3) vs. (1): $\chi^2_5 = 4.445$, $P = .4873$
(3) $(\phi_{c\cdot t}, p_{c+t})$	21	344.150	386.150	
				Time effect in capture rates (5) vs. (3): $\chi^2_5 = 7.817$, $P = .1666$
				Colony effect in capture rates (4) vs. (3): $\chi^2_1 = 2.619$, $P = .1056$
(4) $(\phi_{c\cdot t}, p_t)$	20	346.769	386.769	
(5) $(\phi_{c\cdot t}, p_c)$	16	351.967	383.967	
				Time and colony effect on p (6) vs. (3): $\chi^2_6 = 9.766$, $P = .1349$
				Colony effect (one-sided) on p (6) vs. (5): $\chi^2_1 = 1.949$, $P = .0813$
(6) $(\phi_{c\cdot t}, p)$	15	353.916	383.916	
III. Modeling survival rate				
				Parallelism in survival rates (7) vs. (3): $\chi^2_5 = 1.780$, $P = .8787$
(7) (ϕ_{c+t}, p_{c+t})	16	345.930	377.930	
				Time effect in survival rates (9) vs. (7): $\chi^2_6 = 2.273$, $P = .8930$
				Colony effect in survival rates (8) vs. (7): $\chi^2_1 = 2.877$, $P = .0899$
(8) (ϕ_t, p_{c+t})	15	348.807	378.807	
(9) (ϕ_c, p_{c+t})	10	348.203	368.203	
				Colony effect in survival rates (10) vs. (9): $\chi^2_1 = 2.580$, $P = .1082$
(10) (ϕ, p_{c+t})	9	350.783	368.783	
IV. Further models				
(11) (ϕ_c, p_t)	9	350.870	368.870	
				Colony effect in capture rates (11) vs. (10): $\chi^2_1 = 2.667$, $P = .0501$ (one-sided test)
				Time effect in capture rate (12) vs. (9): $\chi^2_6 = 17.591$, $P = .0073$
(12) (ϕ_c, p_c)	4	365.794	373.794	
				Overall colony effect in rates (13) vs. (9): $\chi^2_2 = 8.285$, $P = .0159$
(13) (ϕ, p_t)	8	356.488	372.488	

parison of the estimated logit ϕ_g and logit ϕ_p obtained from model (ϕ_c, p_{c+t}) ($P = .0584$, Table 15). As expected, the survival in colony P is lower than in colony G (0.621 and 0.763, respectively).

Final analyses.—Once a parsimonious model for survival has been determined, an additional check for differences in probability of capture leads to clearer results. Time variation in the capture rate was highly significant $[\phi_c, p_c]$ vs. model $[\phi_c, p_{c+t}]$, $\chi^2_6 = 17.591$, $P = .0073$; Table 14), and differences in capture rate between colonies were also significant in our opinion (model $[\phi_c, p_c]$ vs. model $[\phi_c, p_{c+t}]$, $\chi^2_1 = 2.667$, $P = .0512$, one-sided test).

The overall comparison between model (ϕ, p_t) and (ϕ_c, p_{c+t}) was significant in a two-sided test $\chi^2_2 = 8.285$, $P = .0159$; Table 14). Moreover, differences between estimates were all in the expected direction (G better than P; Table 15). This comparison looks for differences between colonies in a much more parsimonious and powerful way than TEST 1 or the equivalent LRT comparison between (ϕ_t, p_t) and $(\phi_{c\cdot t}, p_{c\cdot t})$, neither of which was significant ($P = .2333$ and .2926, respectively).

The difficulty in assessing separately the role of survival and capture rate is made clear by the estimated negative sampling correlation (estimated value: −0.42)

TABLE 15. Estimates of survival and capture rates for the Swift (*Apus apus*) (data from G. Gory, *unpublished data*), by colony. For model notation, see the Appendix.

A. Results under model (ϕ_c, p_{c+t}).

	Survival rates			
	Good colony		Poor colony	
	$\hat{\phi}_g$ = 0.763	$\widehat{SE}$ = 0.039	$\hat{\phi}_p$ = 0.621	$\widehat{SE}$ = 0.083
	Capture rates			
	$\hat{p}$	$\widehat{SE}$	$\hat{p}$	$\widehat{SE}$
1983	$\hat{p}_2$ = 0.931	0.067	0.838	0.148
1984	$\hat{p}_3$ = 0.778	0.096	0.574	0.164
1985	$\hat{p}_4$ = 0.574	0.116	0.340	0.152
1986	$\hat{p}_5$ = 0.722	0.098	0.499	0.180
1987	$\hat{p}_6$ = 0.872	0.080	0.723	0.181
1988	$\hat{p}_7$ = 0.888	0.093	0.753	0.193
1989	$\hat{p}_8$ = 0.499	0.102	0.276	0.125

B. Results under ad hoc methods, "return rates."

Return rate	Good colony	Poor colony
Recaptured (next year)*† / all releases	96/195 = 0.492	22/65 = 0.338
Recaptured (any year)‡ / all releases	112/195 = 0.574	28/65 = 0.430
Recaptured (next year)*§ / newly released	45/87 = 0.517	10/39 = 0.256
Recaptured (any year)‡ / newly released	55/87 = 0.632	14/39 = 0.359
Capture–recapture results (from model in part A)	$\hat{\phi}$ = 0.771 $\hat{p}$ = 0.752‖ $\hat{\phi}\hat{p}$ = 0.580	$\hat{\phi}$ = 0.633 $\hat{p}$ = 0.572‖ $\hat{\phi}\hat{p}$ = 0.362

* Estimates of ϕp neglecting variations over time in p.
† χ^2_1 = 4.655, P = .0310.
‡ Does not estimate ϕp because the recaptured individuals have been exposed to several recapture occasions. The return rates ϕp cannot be compared between colonies because the distribution of exposure times is likely to differ between groups.
§ χ^2_1 = 7.448, P = .0063.
‖ Unweighted mean of time-dependent estimates.

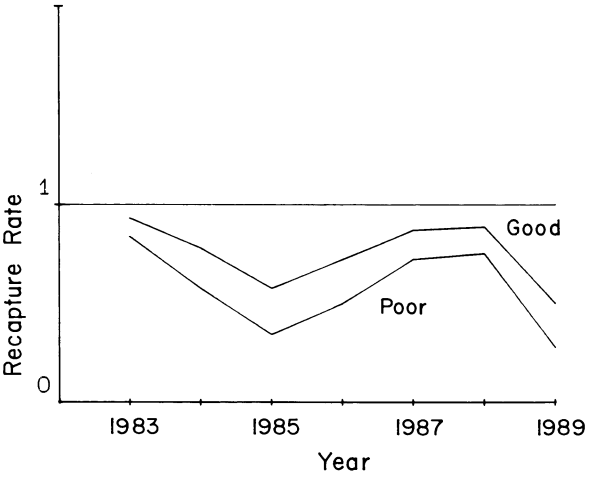

FIG. 1. Estimates of capture rate of Swifts (*Apus apus*) from 1983 to 1989 in the Good and Poor colonies, according to model (ϕ_c, p_{c+t}). The structure of this model forces the estimated capture rates to be parallel over time when expressed on a logit scale, which is on the interval $(-\infty, +\infty)$.

of colony effects on survival and on capture rate in model (ϕ_c, p_{c+t}).

Using Akaike's Information Criterion as a shortcut for model selection, we selected model (ϕ_c, p_{c+t}), which has the lowest AIC value among all the models fitted (Table 16). With this approach, we began testing with model (ϕ_c, p_{c+t}), and only tested this model against its close neighbors. This limits the number of formal tests to four (the last four tests in Table 14).

For the survival rate in the good colony, the asymptotic confidence interval (95% level), back-transformed from the logits, is [0.678, 0.830]. Surprisingly enough, the profile likelihood interval is not markedly different: [0.682, 0.835]. The asymptotic confidence interval ob-

TABLE 16. Results of various capture–recapture models for the Swift (*Apus apus*) data; the deviance (DEV), the number of estimable parameter (*np*), Akaike's Information Criterion (AIC = DEV + 2*np*) are given for each model. AIC, which emphasizes parsimony, is minimal for model (ϕ_c, p_{c+t}), from which further tests can be conducted. For model notation, see the Appendix.

Survival model		Capture model				
		p_{c*t}	p_{c+t}	p_t	p_c	p
ϕ_{c*t}	DEV =	339.705	344.150	346.769	351.967	353.916
	np =	26	21	20	16	15
	AIC =	391.705	386.150	386.769	383.967	373.613
ϕ_{c+t}	DEV =	340.591	345.930	348.494	353.710	355.613
	np =	21	16	15	10	9
	AIC =	382.591	377.930	378.494	373.710	373.613
ϕ_t	DEV =	343.962	348.807	354.944	356.928	362.266
	np =	20	15	13	9	8
	AIC =	383.962	378.807	380.944	374.928	378.266
ϕ_c	DEV =	342.297	348.203	350.870	365.794	367.405
	np =	16	10	9	4	3
	AIC =	374.297	**368.203**	368.870	373.794	373.405
ϕ	DEV =	345.011	350.783	356.488	367.938	372.853
	np =	15	9	8	3	2
	AIC =	375.011	368.783	372.488	373.938	376.853

tained without a logit transform is also similar: [0.686, 0.839]. The close similarity of these intervals may be thought to result from the relatively good precision on survival in the good colony (estimated asymptotic standard error = 0.039). However, the same similarity shows up for the poor colony, despite a lower precision (standard error = 0.083 for the estimated survival rate), as we obtain [0.459, 0.782] for the confidence interval without the logit transform, and [0.457, 0.776] for the profile likelihood interval. This supports our belief in the validity of the general use of profile likelihood intervals, a belief expressed by several authors (Harding et al. 1984, Aitkin and Stasinopoulos 1989, Morgan and Freeman 1989).

Discussion

The use of ad hoc "return rates" is common in bird population studies. Although not based on a formal model, return rates actually estimate only a product of survival and capture rates, ϕp (see Table 15B). An advantage of capture–recapture modeling is that separate inferences about survival and capture rates can be made. The advantage in this example is somewhat limited because the study period is short compared to the average life-span of Swifts. Here, although there is evidence of an overall difference in survival and probability of capture between colonies (Table 15), the present set of data does not allow one to draw definitive conclusions about the relative roles of survival and capture probability in the difference between colonies. Series of zeros at the end of capture histories can be interpreted in various models either as a result of low survival or of low probability of capture. A low p can result from temporary random absences (not detected by tests of temporary emigration if absences are independent over the years). Similarly, a low ϕ can result from permanent emigration. Temporary vs. permanent emigration cannot be distinguished in studies done for only a few years (say 2, 3, or 4 yr), but these factors might be distinguished in long-term studies (say two or three times as many years as the average life-span). Similarly, in a long-term study survival rates can all be estimated separately from capture rates.

Capture–recapture modeling allows a better treatment of sources of variation and better use of available information. A third advantage of capture–recapture modeling is the ability to test the fit of the model. The basic goodness-of-fit tests revealed limited heterogeneity at the beginning of the study. Thus, in such studies we recommend dropping the 1st yr of study, as design becomes more stable after this 1st yr. Another recommendation is to achieve a high capture rate when neither the number of marked animals nor the number of years is high.

Although differences in survival and capture rates between colonies are difficult to demonstrate in this example, capture–recapture modeling makes better use of the information than ad hoc statistics, such as ratios of the number captured to the number of marked birds. These ratios estimate only products like ϕp and should thus be discarded as a basis for inference on survival mechanisms (see, e.g., Nichols and Pollock 1983).

In this example, using the AIC reduces the number of formal tests to four, and thus helps considerably in selecting a model from which the relevant biological questions can be addressed more powerfully.

The annual survival rate obtained in the Good colony (0.763, SE = 0.039, confidence interval [0.678, 0.830], from model $[\phi_c, p_{c+t}]$) is close to the value 0.84 (Perrins 1971) obtained in a successful colony in Oxford, and compatible with other published estimates (0.81 in Sweden and 0.83 in Switzerland; see summaries by Glutz Von Blotzheim and Bauer 1980, Cramp 1985; 0.78 in Britain, Perrins 1971). The apparent survival in the Poor colony (0.621, SE = 0.083) could explain why the survival estimate 0.78 obtained on a larger scale (Perrins 1971), by mixing good and poor colonies, was lower than the value obtained for a successful colony. However, despite the rarity of short-scale exchanges between colonies (1 out of 127 birds), and the general faithfulness of Swifts to their breeding site, this apparent survival in the Poor colony may partly reflect permanent emigration.

Example 3: Roe Deer

Introduction

The roe deer (*Capreolus capreolus*) is a small, strongly territorial cervid. Males are only slightly larger than females. In cervids, survival rates are usually age- and sex-specific (see Clutton-Brock et al. 1982 for an example). In particular, in red deer (*Cervus elaphus*), patterns of survival over age differ between the sexes as a result of sexual selection (Clutton-Brock et al. 1982). However, only species showing an important sexual dimorphism in size or mass have been studied (Gaillard 1988). Thus, one goal of the study was to look for sex-specific survival in relation to age (Gaillard 1988).

The study area is a 260-ha enclosure in the Chizé forest, located 200 km north of Bordeaux, France, close to the Atlantic coast. The enclosure is managed by the "Office National de la Chasse," mainly to produce animals to restock other forests. (See a description of the area and of the roe deer research program in Boiseaubert et al. 1979, Gaillard 1988.)

Each year, recapture samples are obtained on 10–15 d evenly distributed throughout winter. Some of the roe deer captured are released with individual marks (collar plus ear tags to limit the effect of mark losses). Others (including some recaptures) are removed for restocking. These removals decrease sample size, particularly in the oldest age classes. The sampling design was modified in 1985–1986 by an increase in the area sampled. We considered only capture–recapture data of individuals caught in their 1st yr of life (Table 17), as their age can be reliably determined from their size.

TABLE 17. A summary of roe deer (*Capreolus capreolus*) recapture data (Gaillard 1988 and J. M. Gaillard, *unpublished data*). Individuals are classified by columns according to time of next recapture, and by row according to year of release. Newly marked animals (i.e., young) and already marked ones are summarized in different sub-tables. The first diagonal for the newly marked animals has fewer individuals, in proportion, because of a lower survival rate of young animals.

Animal category	Released		Year of next recapture									Never recaptured
	Year, i	R_i	1978–1979	1979–1980	1980–1981	1981–1982	1982–1983	1983–1984	1984–1985	1985–1986	1986–1987	1987–1988	
Males, newly marked	1977–1978	18	7	3	1	0	0	1	1	0	0	0	5
	1978–1979	20		7	3	3	0	0	0	0	0	0	7
	1979–1980	12			3	3	0	1	0	0	0	0	5
	1980–1981	11				4	2	2	0	0	0	0	3
	1981–1982	11					5	2	3	0	0	0	1
	1982–1983	10						3	1	1	1	0	4
	1983–1984	4							2	0	0	1	1
	1984–1985	2								1	0	0	1
	1985–1986	15									4	4	7
	1986–1987	9										5	4
Males, already marked	1977–1978	...	...	...	...	...	...	...	...	...	...	...	...
	1978–1979	5		3	0	1	0	0	0	1	0	0	0
	1979–1980	9			6	3	0	0	0	0	0	0	0
	1980–1981	12				7	1	2	0	0	0	0	2
	1981–1982	17					6	4	1	0	0	0	6
	1982–1983	14						6	2	2	0	0	4
	1983–1984	21							9	3	2	0	7
	1984–1985	19								12	1	0	6
	1985–1986	21									12	2	7
	1986–1987	21										13	8
Females, newly marked	1977–1978	11	3	2	1	2	1	0	0	0	0	0	2
	1978–1979	17		7	4	1	2	0	0	0	0	0	3
	1979–1980	12			6	0	0	0	1	2	1	0	2
	1980–1981	9				3	4	0	0	0	0	0	2
	1981–1982	15					6	3	2	0	0	0	4
	1982–1983	8						2	1	2	0	0	3
	1983–1984	14							3	4	3	1	3
	1984–1985	10								5	0	1	4
	1985–1986	17									4	4	9
	1986–1987	7										2	5
Females, already marked	1977–1978	...	...	...	...	...	...	...	...	...	...	...	...
	1978–1979	3		2	1	0	0	0	0	0	0	0	0
	1979–1980	10			5	1	1	1	1	1	0	0	0
	1980–1981	16				5	4	3	1	0	0	0	3
	1981–1982	12					6	4	1	0	0	0	1
	1982–1983	23						11	4	2	1	0	5
	1983–1984	23							11	8	2	1	1
	1984–1985	25								20	2	2	1
	1985–1986	43									21	4	18
	1986–1987	33										23	10

Age was not thought to influence capture rate and was not considered in this respect.

To account for the expected lower survival of young and to be able to investigate in a convenient and parsimonious way a possible effect of senescence on survival (Gaillard 1988), we used two age structures. The first considered three age classes: a_1, a_2, and a_3, spanning 1, 4, and 5 yr, respectively (Table 18). The corresponding categorical variable is denoted by *ac*. The second age structure (denoted by *ag*) lumps a_2 and a_3, thus only considering an age effect for yearlings. Comparisons of models incorporating *ac* and *ag* will thus test for senescence effects.

TABLE 18. Definition of roe deer (*Capreolus capreolus*) age classes. The effect of age in the different models is denoted, respectively, as *ac* (three age classes) and *ag* (two age classes). For model notation, see the Appendix.

Age (yr)	No. years spanned	Three age classes (*ac*)	Two age classes (*ag*)
0.5–1.5	1	a_1	a_1
1.5–5.5	4	a_2	$\{a_2\}$
5.5 to —	5	a_3	

Basic test and models

Because we expected possible differences in age-specific survival rates of males and females, a convenient model to begin with is model ($\phi_{a \cdot t \cdot s}$, $p_{a \cdot t \cdot s}$). In theory, a goodness-of-fit test of this model can be performed by testing the fit of the CJS model, using RELEASE, separately for each age- and sex-determined cohort. Unfortunately, the data are too sparse to make such

TABLE 19. Modeling roe deer (*Capreolus capreolus*) capture rates (*np* = number of parameters, DEV = deviance, AIC = Akaike's Information Criterion). In all cases, survival is modeled as $\phi_{ac\cdot t\cdot s}$, noted $\phi_{+\cdot}$. "Period" is a categorical variable that splits the years of study in two parts, before and after a change in design thought a priori to have affected the recapture pressure. For model notation, see the Appendix.

Model	*np*	DEV	AIC	Tests between models
(1) $(\phi_{+\cdot}, p_{t\cdot s})$	66	1348.853	1480.853	(fits the data) Time variation within period (2) vs. (1): $\chi^2_{14} = 10.176$, $P = .7487$
(2) $(\phi_{+\cdot}, p_{period\cdot s})$	52	1359.029	1463.029	Period effect (3) vs. (2): $\chi^2_2 = 21.893$, $P = .0000$
(3) $(\phi_{+\cdot}, p_s)$	50	1380.922	1480.922	Sex effect (4) vs. (2): $\chi^2_2 = 0.891$, $P = .6405$
(4) $(\phi_{+\cdot}, p_{period})$	50	1359.920	1459.920	

an approach meaningful. This means we cannot fully assess the validity of the **iii** hypothesis for these data.

The CJS model, considered separately for each sex, model $(\phi_{t\cdot s}, p_{t\cdot s})$, fits the data poorly (TEST 2 + TEST 3, components with enough data: $\chi^2_{50} = 66.666$, $P = .0575$). However, a large part of this χ^2 statistic can be explained by an age effect on survival; the LRT statistic, using SURGE, between model $(\phi_{t\cdot s}, p_{t\cdot s})$ and model $(\phi_{ac\cdot t\cdot s}, p_{t\cdot s})$ equals 38.724, with df = 28. An approximate goodness-of-fit test of model $(\phi_{ac\cdot t\cdot s}, p_{t\cdot s})$ is thus given by $\chi^2_{22} = 65.666 - 38.724 = 27.942$. This model appears to fit the data satisfactorily ($P = .1776$). This goodness-of-fit test is a very rough approximation because it subtracts an LRT χ^2 statistic from a Pearson's contingency table χ^2 statistic (TEST 2 + TEST 3, subject to pooling). Clearly, more research is needed to provide flexible goodness-of-fit tests for complex models.

Similarly, the model $(\phi_{ag\cdot t\cdot s}, p_{t\cdot s})$ seems to fit the data ($\chi^2_{32} = 35.039$; $P = .3259$). The two age classes (*ag*) for survival thus seem enough. We will nevertheless start from model $(\phi_{ac\cdot t\cdot s}, p_{t\cdot s})$ to be able to address questions on senescence later on, with more parsimonious models providing more power. More refined age effects on survival and capture rates are not needed (model $[\phi_{ac\cdot t\cdot s}, p_{t\cdot s}]$ vs. model $[\phi_{a\cdot t\cdot s}, p_{a\cdot t\cdot s}]$, $\chi^2_{60} = 50.144$, $P = 8139$).

Modeling capture rates

The time variation in capture rate was significant (model $[\phi_{ac\cdot t\cdot s}, p_s]$ vs. model $[\phi_{ac\cdot t\cdot s}, p_{t\cdot s}]$, $\chi^2_{16} = 32.07$, $P = .0098$, Table 19). An increase in the area sampled occurred in winter 1985–1986; thus, we tested the overall differences between capture rates before and after this change. Mean capture rates differed significantly between periods model $[\phi_{ac\cdot t\cdot s}, p_s]$ vs. model $[\phi_{ac\cdot t\cdot s}, p_{period\cdot s}]$, ($\chi^2_2 = 21.89$, $P = .0000$, Table 19), and variation of capture rates within each time period did not seem to be significant ($\chi^2_{16} = 10.18$, $P = .7489$). Finally, capture rates did not differ significantly between sexes ($\chi^2_2 = 0.89$, $P = .6405$). For the two periods, respectively, $\hat{p}_1 = 0.477$ (95% CI is 0.429–0.525) and $\hat{p}_2 = 0.695$ (0.617–0.763). This small chi-square statistic (0.89 on 2 df) simultaneously tests for sex effect and the period × sex interaction; because the test statistic is <1, the separate tests would not be significant, hence need not be done.

Modeling survival rates

Proceeding from the model $(\phi_{ac\cdot t\cdot s}, p_{period})$, we tried to reduce variation in survival rate, eliminating one term at a time (Table 20). The triple interaction $ac\cdot t\cdot s$ was not statistically significant (model 1 vs. model 2, $\chi^2_{10} = 7.77$, $P = .6513$, Table 20). The two-way interactions $t\cdot s$ and $ac\cdot t$ were not significant either in the presence or in the absence of other interaction terms. The $ac\cdot s$ interaction in which we were mostly interested did not seem to be significant, as judged from comparisons between models 2 and 5 (Table 20) or between models 3 and 6 (Table 20), or even from the more parsimonious test (without the other interaction terms) of model 7 vs. model 9 (Table 20). All three factors: age, time, and sex, influence survival rate (ϕ_{t+s}, ϕ_{ac+s}, ϕ_{ac+t} against ϕ_{ac+t+s}: respectively, $\chi^2_2 = 13.66$, $P = .0011$; $\chi^2_9 = 15.58$, $P = .0763$; and $\chi^2_1 = 5.35$, $P = .0208$), independently from each other on a logit scale.

Senescence

We assumed, in the beginning, that roe deer life stages could be divided into three age classes (a_1, a_2, and a_3, Table 18). We also found that age has a significant influence on survival rate. Still allowing survival rates to be sex-specific, a model allowing an age stratification into two age classes (noted *ag*, Table 18), combining the mature stage and the senescence period ($a_2 = a_3$) seemed sufficient to reflect age dependence in survival. We can test again for senescence with more parsimony. Model $(\phi_{ag+t+s}, p_{period})$ and model $(\phi_{ac+t+s}, p_{period})$ did not differ greatly ($\chi^2_1 = 1.36$, $P = .1216$, one-sided test) (ϕ_{a_3} is believed a priori to be smaller than ϕ_{a_2}), implying that senescence is at most weak in this population, contrary to other roe deer populations (Gaillard 1988).

TABLE 20. Modeling roe deer (*Capreolus capreolus*) survival rates (*np* = number of parameters, DEV = deviance, and AIC = Akaike's Information Criterion). In each model for ϕ, the capture rate is modeled as (p_{period}). For model notation, see the Appendix.

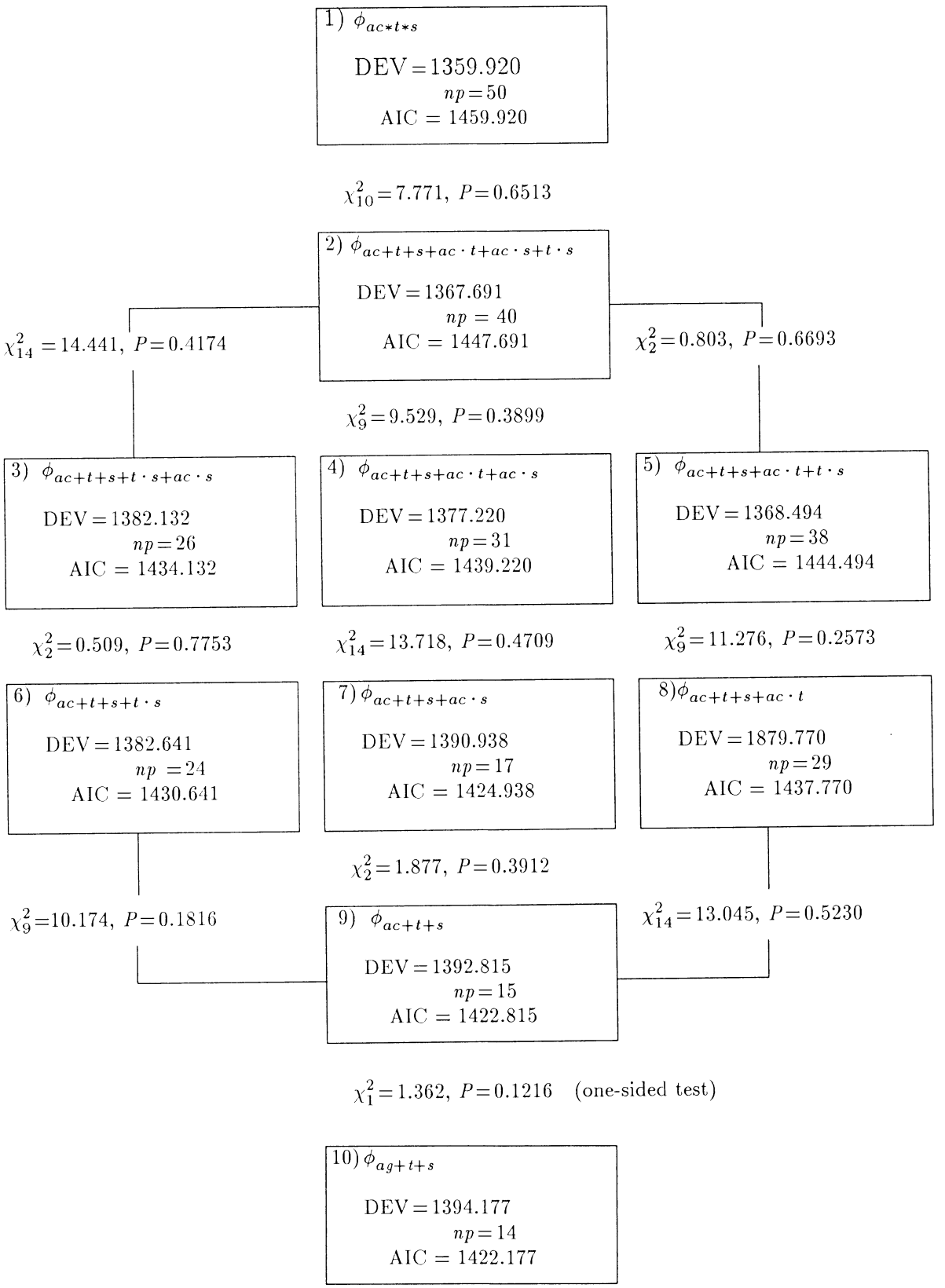

TABLE 21. Variables used to model roe deer (*Capreolus capreolus*) survival. "*D*" is an index of density (number of different roe deer caught per year, including young and adults, adults being the majority). Because of a change in the sampling area then, *D* cannot be used after 1984–1985. The values for *D* for the last two years are thus arbitrary. Time dependence in these last two years is accounted for by the categorical variables "*ta*" with three categories. Because the second and third categories of *ta* take a particular value for 1985/1986 and 1986/1987, respectively, the values of *D* for these years do not influence the predicted ϕ in models incorporating $D + ta$.

Year:	1977/1978	1978/1979	1979/1980	1980/1981	1981/1982	1982/1983	1983/1984	1984/1985	1985/1986	1986/1987
D:	270	270	288	244	304	338	321	285	0*	0*
ta: 1	1	1	1	1	1	1	1	1	0	0
2	0	0	0	0	0	0	0	0	1	0
3	0	0	0	0	0	0	0	0	0	1

* Arbitrary values.

Density dependence

Gaillard (1988) suggested that survival rates in roe deer populations are density dependent, mainly because the mass of the young is negatively related to density, which probably affects their life-span. We use the total number of individuals caught at the beginning of each time interval as a rough index to density. This variable is denoted as *D* (number caught in winter 1977–1978 for survival rate between winters 1977–1978 and 1978–1979, Table 21). Unfortunately, the area sampled was enlarged in 1985–1986 and, as a result, the number of unmarked animals caught increased. We had to restrict the analysis of the relationship between density and survival rate to the period from 1977–1978 to 1985–1986 (Table 21), during which the capture rate seemed to be constant.

To test for a possible density effect on survival, we had thus to relate ϕ to density for the first 8 yr and to leave ϕ unrelated to density, *D*, for the last 2 yr. This was done, using SURGE, by constraining logit(ϕ) to be a linear compound of density *D* and a categorical variable *ta* with three categories, as described in Table 21. The corresponding model incorporated also additive age and sex effects and was thus denoted $[\phi_{ag+s+D+ta}, p_{period}]$. Reducing time dependence to this density effect was acceptable (Table 22; $\chi^2_6 = 2.235$, $P = .8969$). Density dependence could then be tested on a single degree of freedom by testing model $[\phi_{ag+s+ta}, p_{period}]$ vs. model $[\phi_{ag+s+D+ta}, p_{period}]$. Density dependent measured in this way was close to statistical significance (Table 22; $\chi^2_1 = 2.339$, $P = .0631$ as a one-sided test).

The remaining time dependence, concerning the last 2 yr, was extremely strong (model $[\phi_{ag+s}, p_{period}]$ vs. model $[\phi_{ag+s+ta}, p_{period}]$, $\chi^2_2 = 13.248$, $P = .0013$). It resulted from an unexpected low survival in 1986–1987 (year 9, Table 23). Compared to this variation, the time dependence in the first 8 yr, even if slightly related to density, was at most weak.

TABLE 22. Modeling roe deer (*Capreolus capreolus*) survival rates as a function of time effects and density (*np* = number of parameters, DEV = deviance, AIC = Akaike's Information Criterion). In all cases, capture rate is modeled as p_{period} (Table 19). *D* is density (see Table 21), *ag* is a categorical variable separating yearlings from older animals (see Table 18), and *ta* is a categorical variable for dates 9 and 10 (Table 21). For model notation, see the Appendix.

Model	*np*	DEV	AIC	Tests between models
Variation over time				
(1) (ϕ_{ag+t+s})	14	1394.177	1422.177	Reduction of time effect to density effect in first 8 yr plus time dependence for last 2 yr (4) vs. (1): $\chi^2_6 = 2.235$, $P = .8969$
(2) (ϕ_{ag+s})	5	1411.999	1421.999	Time effect in last 2 yr (3) vs. (2): $\chi^2_2 = 13.248$, $P = .0013$
(3) $(\phi_{ag+ta+s})$	7	1398.751	1412.751	Density dependence (one-sided test) (4) vs. (3): $\chi^2_1 = 2.339$, $P = .0631$
Density dependence				
(4) $(\phi_{ag+D+ta+s})$	8	1396.412	1412.412	Density dependence differing with age (5) vs. (4): $\chi^2_1 = 0.223$, $P = .6368$
Interactions of density dependence				
(5) $(\phi_{ag\cdot D+ta+s})$	9	1396.189	1414.189	Density dependence differing between sexes (6) vs. (4): $\chi^2_1 = 2.370$, $P = .1237$
(6) $(\phi_{ag+ta+D\cdot s})$	9	1394.042	1412.042	

TABLE 23. The final model (model 3 of Table 22) for the roe deer (*Capreolus capreolus*) data: model ($\phi_{ag+ta+s}$, p_{period}), with seven parameters. The first period was taken apart from the two final dates to test for the effect of density, of which an estimate was available during this first period.

Estimates of additive components of logit(ϕ) (mean ± SE) (five parameters)	
Male, yearling, first period	1.1433 ± 0.2215
Adult–yearling	1.0269 ± 0.2833
Female–male	0.5060 ± 0.2196
t_9–first period	−1.427 ± 0.3121
t_{10}–first period	−0.0401 ± 0.8344
Estimates of survival rates (with asymptotic 95% confidence interval) (estimated from the five primary parameters above)	
Male, yearling, first period	0.758 [0.671, 0.829]
Male, adult, first period	0.897 [0.851, 0.931]
Male, yearling, t_8–t_9	0.430 [0.281, 0.592]
Male, adult, t_8–t_9	0.678 [0.533, 0.795]
Male, yearling, t_9–t_{10}	0.751 [0.367, 0.940]
Male, adult, t_9–t_{10}	0.894 [0.633, 0.976]
Female, yearling, first period	0.839 [0.762, 0.894]
Female, adult, first period	0.936 [0.904, 0.957]
Female, yearling, t_8–t_9	0.555 [0.389, 0.710]
Female, adult, t_8–t_9	0.833 [0.473, 0.965]
Female, yearling, t_9–t_{10}	0.777 [0.658, 0.863]
Female, adult, t_9–t_{10}	0.933 [0.731, 0.986]
Estimates of capture rates (with 95% asymptotic confidence interval) (two parameters)	
First period	0.471 [0.423, 0.521]
Second period	0.664 [0.578, 0.741]

Assuming nevertheless in a provisional way an effect of density, no difference in slope was found either with age (Table 22, model 5) or with sex (Table 22, model 6). Although the AIC criterion was the smallest for model 6, the difference from model 3 or model 4 was less than unity (Table 22), indicating that these three models had almost the same goodness of fit (Sakamoto et al. 1986). In that case, the model with the fewest parameters (model 3, Table 23) is chosen (principle of parsimony).

Discussion

In this example, the CJS model (by sex) fitted the data poorly, although it was not formally rejected. The fit was better when age dependence was considered. This was confirmed in turn by a test for age dependence in a more parsimonious model. The final model ($\phi_{ag+ta+s}$, p_{period}) was rather simple, with only seven parameters to describe the 10-yr data set. Although this model fitted the data well, there was no clear biological interpretation of the time dependence, which is almost entirely due to a low survival rate in 1986–1987.

There were three particular biological questions addressed regarding survival rates: possible interaction between age and sex, possible effect of senescence, and possible density dependence. The first two questions are answered negatively, with reasonable certainty, because of the parsimonious model selected, which excludes these effects. However, there is limited evidence for density dependence ($P = .0631$, one-sided test). Although Akaike's Information Criterion is smaller for two models with a density effect, it differs by less than unity from the value in the more parsimonious final model (model 3, Table 22).

We conclude that the data from this study support a model in which females survive, on average, better than males: mature individuals (a_2 and a_3) have parallel increased survival rate (on logits) for both sexes (in the first period, 0.758 and 0.897 for young and mature males and 0.839 and 0.936 for young and mature females; Table 23). Differences between sex and age are similar to those found in a much more dimorphic species like red deer (0.7 and 0.85 for males, 0.75 and 0.90 for females, unpublished results on data from the island of Rhum, T. H. Clutton-Brock, *personal communication*). Sexual dimorphism is obviously not the only explanation for a differential survival rate between sexes (Gaillard 1988).

When test power in a general model is feared to be small, it is worth testing again some previous hypothesis based on the final model when the model allows such a test (Goodman 1984). In this example, parallelism in the pattern of temporal fluctuations in survival rate across ages (the main question) was tested twice, first by comparing models ϕ_{ac+t+s} and $\phi_{ac+t+s+ac\cdot s}$, and secondly (in a more restricted, but potentially more powerful way) by comparing slopes of the survival rate vs. density relationships between young and mature animals. A similar answer of no significant interactions for both tests provided more confidence in the additivity of age, sex, and time effects.

The detection of interaction terms is critical in ecology. Particularly in the case of experiments that involve a deliberate change at some point in the course of the study, results of this experiment are expected to be found in a different reaction of some category of individuals through time, i.e., in the interaction of time and some categorical variable. The next example provides such a case.

EXAMPLE 4: THE COMMON LIZARD

Introduction

The common lizard (*Lacerta vivipara*) is a small viviparous lizard (50–60 mm adult body length) widely distributed across Europe. This species is particularly abundant in humid habitats such as peat bogs and heathland. Demographic traits vary considerably across populations (Bauwens et al. 1987, Pilorge et al. 1987), particularly density. A key question is thus to know if density plays a role in shaping the demographic strategy of each population.

In the course of a population study of *Lacerta vivipara* in southern France (see Pilorge 1988), a density manipulation was performed (Massot 1987, Pilorge 1988) using two neighboring fields. Population size in each field was estimated at each capture "occasion" by use of the closed-model methods in program CAP-

TABLE 24. A summary of capture–recapture data of the lizard (*Lacerta vivipara*). The *m* arrays (number captured according to time of capture and time of next capture) are given by cohorts (i.e., by time of first capture) for each field. The number released and the numbers recaptured classified according to time of next recapture are given in each *m* array. Among the 46 individuals recaptured at time 2 in field D, only 18 were released (density manipulation).

Field D: density decreased

Time	Cohort 1						Cohort 2					Cohort 3				Cohort 4			Cohort 5	
1	72	46	4	4	0	0														
2	18		5	3	0	1	7	2	1	0	1									
3	9			3	2	1	2		0	1	0	27	12	1	2					
4	10				3	3	1			0	0	12		3	3	18	4	3		
5	5					4	1				1	4			4	4		3	54	26

Field I: density increased

Time	Cohort 1						Cohort 2					Cohort 3				Cohort 4			Cohort 5	
1	58	24	4	1	1	1														
2	24		5	0	1	1	32	6	5	0	2									
3	9			2	1	1	6		2	0	0	23	8	1	3					
4	3				0	0	7			4	0	8		2	1	32	9	4		
5	3					2	4				4	3			0	9		6	47	15

TURE (Otis et al. 1978, White et al. 1982). Thus, a capture occasion was the result of using Pollock's robust design (Pollock 1982). In the first, field density was multiplied by two (field I, for Increased); in the second, field density was divided by two (field D, for Decreased). The categorical variable *f*, for fields, thus has two categories, I and D. We restricted our attention here to individuals that were adult males (at least 2 yr old) when first captured.

The lizards were individually marked by toe clipping. The data (Table 24) were collected during $k = 6$ sessions of capture: June 1986, August 1986, June 1987, August 1987, June 1988, and August 1988. The individuals newly marked in 1987, whether in June or August, and those marked in June 1988 were mostly 2 yr olds because the number of older individuals missed in earlier sessions decreased strongly with time. This result led us to consider the year of first capture (categorical variable y with three categories) as possibly affecting the survival and capture rates. We also considered the cohort, in the meaning of month of first capture (categorical variable c with five categories), from which y was derived (year 1 = cohorts 1 and 2; year 2 = cohorts 3 and 4; year 3 = cohort 5).

The experimental manipulation of density (transfer of individuals from field D to field I) occurred immediately after the second session of capture (August 1986). The transferred individuals were not taken into account in our analysis because their behavior was most likely affected by the transfer.

The comparison of appropriate models with both field and time dependence in survival rates addresses two main biological questions:

1) Do survival rates differ between sites after August 1986? One expects survival to be lower in site I.

2) Do survival rates differ between sites only immediately after the manipulation (August 1986–June 1987) or for a longer time?

First, the populations were not geographically closed. Permanent emigration could induce an underestimation of the survival rates. However, random samples out of the experimental areas led to the capture of only 20 marked animals out of 250, most being close to the boundaries (M. Massot et al., *unpublished data*). Furthermore, the majority of the individuals captured out of the experimental areas, regardless of sex and age, were also recaptured on the experimental sites during the same time period, indicating that emigration was not permanent.

Second, some lizards lose toes accidentally (this is not unlike tag loss, which can cause problems, see e.g., Seber 1982). Thus, to prevent confusion, at least two toes were clipped. However, in some cases, it was impossible to identify some individuals with lost toes. Although this problem leads to an underestimation of the survival rate, such cases are rare (<2%).

Third, some lizards were less catchable than others. Cohort, status, and location could be important factors of heterogeneity in capture rates. Although we could account for cohort effects indirectly, it was almost impossible to account for location effect. Fortunately, the common lizard has a relatively large home range in regard to the area explored (≈90 m^2, Ortega-Rubio et al. 1989). Any individual was likely to be caught in different types of micro-habitats during one session of capture. As a consequence, the risk of heterogeneity in capture rates according to locations was limited.

Basic models and tests

The goodness-of-fit tests of the CJS model over the two fields separately (model $[\phi_{f\cdot t}, p_{f\cdot t}]$) revealed a somewhat poor fit (components of TEST 2 + TEST 3 with enough data: $\chi^2_{18} = 25.748$, $P = .1057$). These results tend to corroborate the **iii** hypothesis. The components of TEST 3 showed a strong heterogeneity for the last date of capture between newly and already marked

TABLE 25. Differences in numbers of male lizards (*Lacerta vivipara*) recaptured at occasion *nr* 5 (August 1988) among those captured at occasion 4 (June 1988). The lizards first captured in June 1988 (cohort 5) were recaptured in lower proportion in August 1988 than those marked earlier. This fifth cohort effect is significant ($\chi^2_2 = 11.826$, $P = .0027$).

	Lizards captured in June 1988			
	For the first time (cohort 5)		Already marked (cohorts 1–4)	
	Recaptured in August 1988?			
	Yes	No	Yes	No
Field D: (density decreased)	26	28	12	2
Field I: (density increased)	15	32	12	7

individuals in both fields ($\chi^2_1 = 6.364$, $P = .0116$, and $\chi^2_1 = 5.464$, $P = .0194$). Among the individuals captured at date 5, only 27 of the 66 (41%) new individuals were recaptured at date 6 compared to 24 of the 33 (73%) previously marked individuals captured at date 5 (Table 25). Although cohort effects were a priori suspected, such a strong effect restricted to the last cohort released (cohort 5, composed of individuals first captured in June 1988) in a consistent way in both fields was surprising. This effect was highly significant (Table 25), whether measured by a χ^2 (adding the two components of TEST 3 above: $\chi^2_2 = 11.826$, $P = .0027$), or by the asymptotically equivalent LRT ($\chi^2_2 = 12.484$, $P = .0019$). Although the null hypothesis addressed by this component of TEST 3 was not of particular interest a priori, it seemed difficult not to account for this difference between the last cohort and the previous ones.

The reason for this lack of fit was a difference in $\beta_6 = \phi_5 p_6$, the compound survival–capture parameter for the last period, between cohorts 1 and 4 on the one hand, and cohort 5 on the other. The difference cannot be attributed to a difference in survival rates only or in capture rates only; rather, it can be accounted for by a last cohort effect undifferentiated among survival rates, capture rates, or both.

Denoting l the categorical variable separating the first four cohorts from the last cohort made model ($\phi_{f\cdot l\cdot t}$, $p_{f\cdot l\cdot t}$), equivalent to model ($\phi_{f\cdot l\cdot t}$, $p_{f\cdot t}$), with 20 identifiable parameters, the starting point. The likelihood ratio test between model ($\phi_{f\cdot l\cdot t}$, $p_{f\cdot l\cdot t}$) and the CJS model ($\phi_{f\cdot t}$, $p_{f\cdot t}$) was precisely the LRT test comparing cohort 5 to cohorts 1–4 (Table 25) and was thus highly significant ($\chi^2_2 = 12.484$, $P = .0019$). A goodness-of-fit test of this model was obtained by removing the corresponding two components from TEST 2 + TEST 3. This test indicated that the model fit the data well ($\chi^2_{10} = 7.011$, $P = .7244$). When comparing models allowing different β_6 parameters for each field in the last cohort, the last cohort will not contribute to the difference in deviance between models. As a consequence, restricting our attention to such models is strictly equivalent to removing the last cohort from the analyses.

There did not seem to be any reason to consider more detailed models with cohort c or year y effect (model [$\phi_{f\cdot l\cdot t}$, $p_{f\cdot l\cdot t}$] vs. model [$\phi_{f\cdot y\cdot t}$, $p_{f\cdot y\cdot t}$]: $\chi^2_{20} = 5.160$, $P = .8820$; model [$\phi_{f\cdot y\cdot t}$, $p_{f\cdot y\cdot t}$] vs. model [$\phi_{f\cdot c\cdot t}$, $p_{f\cdot c\cdot t}$]: $\chi^2_{20} = 19.939$, $P = .4618$).

There was probably still some heterogeneity in the data, as the distribution of first recaptures at dates 5 and 6 was uneven between individuals captured only at dates 1, 2, and 3 and those captured at date 4 (corresponding component of TEST 2: $\chi^2_1 = 5.398$, $P = .0202$). Trap dependence might be needed in further

TABLE 26. Modeling capture rates for the lizard (*Lacerta vivipara*) data: f = field (in relation with density manipulation), l = last cohort effect, t = time, and e = effort (see Table 27). In all cases, the survival model is $\phi_{f\cdot l\cdot t}$, noted $\phi_{\cdot}$. For each model, we give the number of estimable parameters (*np*), the deviance (DEV), and Akaike's Information Criterion (AIC). For model notation, see the Appendix.

Model	*np*	DEV	AIC	Comparison
(1) ($\phi_{\cdot}$, $p_{f\cdot t}$)	20	972.117	1012.117	
				time × field interaction (2) vs. (1): $\chi^2_3 = 1.802$, $P = .6145$
(2) ($\phi_{\cdot}$, p_{f+t})	17	973.919	1007.919	
				time effect (3) vs. (2): $\chi^2_3 = 19.457$, $P = .0002$
(3) ($\phi_{\cdot}$, p_f)	14	993.376	1021.376	
				effort effect (4) vs. (3): $\chi^2_1 = 18.328$, $P = .0000$ shape of relation with effort (4) vs. (2): $\chi^2_2 = 1.129$, $P = .5686$
(4) ($\phi_{\cdot}$, p_{f+e})	15	975.048	**1005.048**	
				effort × field interaction (5) vs. (4): $\chi^2_1 = 1.054$, $P = .3046$
(5) ($\phi_{\cdot}$, $p_{f\cdot e}$)	16	973.994	1005.994	
				field effect (6) vs. (4): $\chi^2_1 = 3.292$, $P = .0696$
(6) ($\phi_{\cdot}$, p_e)	14	978.340	1006.340	

TABLE 27. Effort of recapture (in days of work) for the lizard (*Lacerta vivipara*) data. This variable, denoted *e*, is used in Table 26 to model the capture rate. A = August, J = June.

	Date				
	A 1986	J 1987	A 1987	J 1988	A 1988
Field D (density decreased)	13	5	7	5	17
Field I (density increased)	12	7	9	6	20

models to account for this heterogeneity. However, this heterogeneity was limited mostly to the last recapture session and we did not consider it in what follows.

Modeling capture rates

Our starting point was model $(\phi_{f\cdot/\cdot t}, p_{f\cdot t})$ or, equivalently, model $(\phi_{f\cdot/\cdot t}, p_{f\cdot/\cdot t})$. While the time × field interaction was negligible (model [2] vs. model [1], Table 26, $\chi^2_2 = 1.802$, $P = .6145$), the time effect was strongly significant (model [3] vs. model [2], Table 26, $\chi^2_3 = 19.457$, $P = .0002$). Because the number of days of work at each session (Table 27) was a good measure of the capture effort, we attempted to model the strong time dependence in capture rate as a function of this variable, denoted *e*, to achieve more parsimony. The effort variable accounted for 94% of the 19.457 difference in deviance attributable to time dependence (model $[\phi_{f\cdot/\cdot t}, p_f]$ vs. model $[\phi_{f\cdot/\cdot t}, p_{f+e}]$, $\chi^2_1 = 18.328$, $P = .0000$, Table 26). The remaining 6% indicated that the functional form logit(p) = $a + be$ was acceptable within each field (model $[\phi_{f\cdot/\cdot t}, p_{f+e}]$ vs. model $[\phi_{f\cdot/\cdot t}, p_{f+t}]$, $\chi^2_2 = 1.129$, $P = .5686$, Table 26). The slopes b were similar between fields (model $[\phi_{f\cdot/\cdot t}, p_{f+e}]$ vs. model $[\phi_{f\cdot/\cdot t}, p_{f\cdot e}]$, $\chi^2_1 = 1.054$, $P = .3046$, Table 26), in accordance with the absence of field × time interaction in capture rates. However, the difference between the intercepts a in the two fields was nearly significant (model $[\phi_{f\cdot/\cdot t}, p_e]$ vs. model $[\phi_{f\cdot/\cdot t}, p_{f+e}]$, $\chi^2_1 = 3.292$, $P = .0696$, Table 26). Such a difference in intercepts expressed a possible difference in efficiency of capture between the two fields, for a given effort. Following the same philosophy as in the previous examples, we considered model (p_{f+e}) as the most appropriate for further tests on survival rates. It has the lowest Akaike's criterion among the models considered in this first part of the analysis (Table 26).

Under model $(\phi_{f\cdot/\cdot t})$ for survival, the model (p_{f+e}), with three parameters, adequately accounted for variations in capture rates over time. This is a definite reduction compared to the more general model where eight parameters were separately identifiable from survival rates (model [4] vs. model [1], Table 26: $\chi^2_5 = 2.931$, $P = .7106$).

Modeling survival rate

At first sight, the survival rates estimated from the current model $(\phi_{f\cdot/\cdot t}, p_{f+e})$ did not seem to differ strongly between fields (Fig. 2). The survival rate in field D was higher than field I during the first period, before the manipulation of density (0.864 vs. 0.733). The log odds ratio, logit ϕ_D − logit ϕ_I, measures the difference in survival between the two fields. It decreased from 0.84 for the first period to 0.52 immediately after the manipulation. The change (0.32, $\widehat{SE} = 0.994$) did not differ significantly from zero ($\chi^2_1 = 0.201$, $P = .7494$). Moreover, it was not in the direction expected on a priori grounds, as the manipulation of density was expected to increase the survival rate in D and to decrease it in I. In the following periods, D tended to keep its original advantage in survival. There was thus no noticeable effect of the experimental changes in density on survival.

Further tests

Several models led to the same conclusion. For instance, the field × time interaction was not significant (model $[\phi_{f\cdot t\cdot/\cdot}, p_{f+e}]$ vs. model $[\phi_{(f+t)\cdot/\cdot}, p_{f+e}]$: $\chi^2_5 = 3.613$, $P = .6063$, Table 28). The slight overall difference in survival, approximately constant over time (Fig. 2), was nearly significant (model $[\phi_{(f+t)\cdot/\cdot}, p_{f+e}]$ vs. model $[\phi_{/\cdot t}, p_{f+e}]$: $\chi^2_1 = 2.919$, $P = .0875$, Table 28).

However, the effect on capture rate of the field × effort interaction, negligible in previous models, was nearly significant (Table 29, model [5] vs. model [6]: $\chi^2_1 = 2.892$, $P = .0890$). In the presence of the field × effort interaction, the field effect in survival vanished (Table 29: model [8] vs. model [5], $\chi^2_1 = 1.084$, $P = .2978$). As in the Swift example, we were thus faced with an effect that could not be attributed clearly to either survival or to capture rates. Because the possible difference in survival did not change through time, this

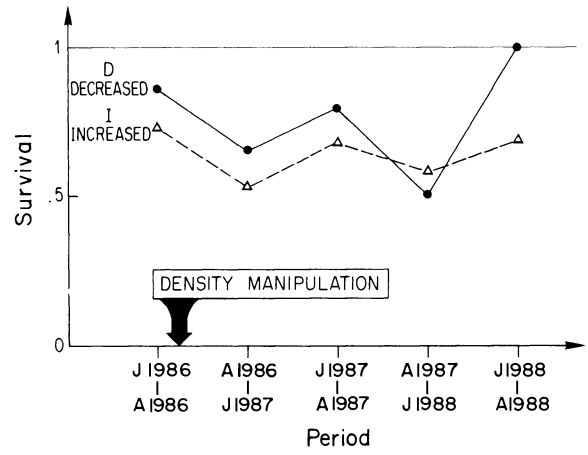

FIG. 2. Survival rates of adult male common lizards, estimated from model $(\phi_{f\cdot/\cdot t}, p_{f+e})$ (see *Discussion*). J = June, A = August. The survival rates concern individuals first captured before June 1988. The validity of the last estimates (J1988–A1988) depends on the validity of the model (p_{f+e}) for this date, which cannot be tested. D = field where lizard density was decreased and I = field where lizard density was increased.

TABLE 28. Modeling survival rate for the lizard (*Lacerta vivipara*) data: f = field (in relation with density manipulation), l = last cohort effect, t = time, and e = effort (see Table 27). In all cases, the capture model is p_{f+e}. For each model, we give the number of estimable parameters (*np*), the deviance (DEV), and Akaike's Information Criterion (AIC). For model notation, see the Appendix.

Model	*np*	DEV	AIC	Comparison
(1) $(\phi_{f\cdot t\cdot l}, p_{f+e})$	15	975.048	1005.048	
				field × time interaction
				(2) vs. (1): $\chi^2_5 = 3.313$, $P = .6063$
(2) $(\phi_{[f+t]\cdot l}, p_{f+e})$	10	978.661	**998.661**	
				field effect
				(3) vs. (2): $\chi^2_1 = 2.919$, $P = .0875$
(3) $(\phi_{t\cdot l}, p_{f+e})$	9	981.580	999.580	

difficulty did not affect our main conclusion on the absence of effect of the density manipulation.

A logical hypothesis at this point was to consider whether or not survival was constant over years. In such a case survival could still differ between seasons, especially because the two seasons were of unequal length and were not biologically equivalent. We thus considered a categorical variable s, with two categories: June–August vs. August–June. Time dependence was reduced to this seasonal effect (model $[\phi_{s\cdot l}, p_{f\cdot e}]$ vs. model $[\phi_{t\cdot l}, p_{f\cdot e}]$: $\chi^2_3 = 2.660$, $P = .4471$, Table 29). The cohort effect was still significant (model $[\phi_s, p_{f\cdot e}]$ vs. model $[\phi_{s\cdot l}, p_{f\cdot e}]$, $\chi^2_1 = 13.538$, $P = .0002$, Table 29). Table 29 summarizes the main models used, in particular model $(\phi_{s\cdot l}, p_{f\cdot e})$, which has the lowest AIC among all models fitted. Adding the goodness-of-fit χ^2 of model $(\phi_{f\cdot t\cdot l}, p_{t\cdot l})$ ($\chi^2_{10} = 7.011$) and its difference in deviance with model $(\phi_{s\cdot l}, p_{f\cdot e})$ ($\chi^2_{13} = 7.396$), shows that the latter fits the data ($\chi^2_{23} = 14.407$, $P = .9146$).

Interest in testing equality of survival over the two seasons must first account for their differing lengths (i.e., June–August = 2 mo vs. August–June = 10 mo). Here it is appropriate to compute estimates of the daily log-survival rate, δ, where $\hat{\delta} = -\ln(\hat{\phi})/(\text{number of days})$. The estimated $\text{var}(\hat{\delta}_1)$ and $\text{var}(\hat{\delta}_2)$ for the two seasons, and $\text{cov}(\hat{\delta}_1, \hat{\delta}_2)$ can be computed from the delta method (see, e.g., Seber 1982). A test of the null hypothesis H_0: $\delta_1 = \delta_2$ can be made based on the test statistic

$$z = \frac{\hat{\delta}_1 - \hat{\delta}_2}{\sqrt{\widehat{\text{var}}(\hat{\delta}_1) + \widehat{\text{var}}(\hat{\delta}_2) - 2\,\widehat{\text{cov}}(\hat{\delta}_1, \hat{\delta}_2)}}$$

where under the null hypothesis, z has asymptotically a standard normal distribution (i.e., $z \approx N[0, 1]$). Computed in this manner, $\hat{\delta}$ is a ML estimator because $\hat{\phi}$ is a one-to-one transformation of the MLE $\hat{\phi}$.

Using the test described above, the daily log-survival rates did not differ significantly between seasons (June–August = −0.00288; August–June = −0.00201; $z = -0.592$, $P = .5541$) although, as expected, survival was lower during the reproduction season (June–August), when males experience more stress, than during the rest of the year. The annual survival rate 0.455 (product of the seasonal rates), has the approximate 95% CI [0.355, 0.611] (Table 30).

TABLE 29. Results of various capture–recapture models for the lizard (*Lacerta vivipara*) data. For each model, the deviance (DEV), number of estimable parameters (*np*), and the value of Akaike's Information Criterion (AIC = DEV + 2*np*), which is minimal for model $(\phi_{s\cdot l}, p_{f\cdot e})$, are given. For model notation, see the Appendix.

Survival model	Capture model		
	$p_{f\cdot t}$	$p_{f\cdot e}$	p_{f+e}
$\phi_{f\cdot t\cdot l}$	DEV = 972.117	973.994	975.048
	np = 20	16	15
	AIC = 1112.117	1005.994	1005.048
$\phi_{[f+t]\cdot l}$	DEV = 973.033	975.769	978.661
	np = 18	11	10
	AIC = 1009.033	995.769	998.661
$\phi_{t\cdot l}$	DEV = 974.457	976.853	981.580
	np = 16	10	9
	AIC = 1006.457	996.769	999.580
$\phi_{s\cdot l}$	DEV = 975.798	979.513	984.469
	np = 13	7	6
	AIC = 1001.798	**993.513**	996.469
ϕ_s	DEV = 987.361	993.051	996.370
	np = 12	6	5
	AIC = 1011.361	1005.051	1006.370

Discussion

In this example, a close examination of the components of the goodness-of-fit tests showed a strong difference in survival or in capture rate between the first four cohorts and the last cohort. Although differences between cohorts were plausible, there was no particular reason to expect that these differences were restricted to the last cohort. After one additional session of capture, it will become possible to separate this difference in distinct survival and capture components.

As in the previous example on roe deer, capture rate was eventually modeled in a most parsimonious way (four parameters), using a regression built into the models on a measure of effort of capture. This way to achieve more parsimony and thus more power should be considered in any capture–recapture experiment.

The final model had the lowest AIC among all models fitted. However, it was tailored to the data, to some

TABLE 30. Results of the final model ($\phi_{s\cdot l}$, $p_{f\cdot e}$) (seven parameters) for the lizard (*Lacerta vivipara*) data (s = season, l = last cohort, f = field, e = effort; see Table 27).

		Estimate	Standard error
Survival rates (cohorts 1–4)	June–August:	0.8364	0.0520
	August–June:	0.5445	0.0491
	Annual survival:	0.4554	0.0778
Survival by capture rate (cohort 5)	June–August 1988:	0.5332	0.0741
Capture rates	Field D (density decreased):	$\text{logit}(p) = -1.041 + 0.0788 \cdot e$	
	Field I (density increased);	$\text{logit}(p) = -1.245 + 0.1910 \cdot e$	

extent: first, because the last cohort effect was not expected a priori; second, because a different equation for the effect of capture effort was used in each field. There was no noticeable effect of density manipulation on survival rates anywhere in the modeling process. The strategy adopted limited the risk of attributing possible differences in capture rates to differences in survival. This risk would have been present with ad hoc calculations on return rates.

From the biological point of view, the insensitivity of survival rates to density manipulation might result from compensatory mechanisms between reproduction and survival, because the density manipulation strongly decreased reproductive rates in field I (T. Pilorge et al., *unpublished data*). Another feature of this example is that seasonal sessions of recapture provided estimates of seasonal survival rates that could be compared both between and within years. A similar approach was used in a study of small-mammal populations by Granjon (1987). Season-specific survival is an important issue in studies of small-mammal populations.

In such experiments, randomized replicates are difficult for reasons of cost. Density manipulation is expected to be detectable mainly as a site × time interaction term in survival. In terms of design, it thus seems advisable to begin manipulation only after a few capture sessions, to be able to assess differences between fields before manipulation with enough precision.

EXAMPLE 5: GREATER FLAMINGO

Introduction

The Greater Flamingo (*Phoenicopterus ruber*) breeds colonially in lagoons and lakes of the tropical and Mediterranean zone. It has been mentioned as observed in the Camargue, the Rhone delta in southern France, as early as 1551 (Johnson 1983) and has bred there since at least 1914 (Cramp and Simmons 1977). Its reproductive biology has been studied in the Camargue since 1977 (Johnson 1983). In particular, resightings of flamingos ringed as chicks were used to investigate migratory movements (Johnson 1983:109) and dispersal to other breeding sites (A. R. Johnson et al., *unpublished manuscript*). The few breeding sites around the Mediterranean seem to be used in an intermittent way, according to their suitability over the years (Cramp and Simmons 1977). The irregular breeding success is counterbalanced by a high survival, attested by maximum longevity records (27 yr in the wild, >50 yr in captivity, Johnson 1983), with few quantitative results available. Johnson (1983:168) estimated an annual survival rate varying from 0.83 to 0.92, according to age, from recoveries of birds ringed as chicks. However, such estimates can be strongly biased, particularly for long-lived species (Anderson et al. 1985). Johnson et al. (1991) used all resightings of chicks marked between 1977 and 1985 in the Camargue for an exploratory analysis of survival rates. They emphasized the effect of wintering strategies on survival rates.

The data used here are from the study period 1977 to 1987. A mass ringing of flamingo chicks was done one day each summer on the main colony in the Camargue just before the earliest fledging. At that time the young cannot be sexed. The recapture data are visual resightings done with the aid of spotting telescopes that allow the leg band code to be read. We restricted our attention to resightings done in the spring and summer on the colony. Birds resighted as breeding adults can often be sexed. The study period includes an unusually severe winter in 1984–1985 during which coastal lagoons remained frozen for 2 wk and a massive mortality of flamingos was observed. One objective of this example is to see if survival rate estimates from these resighting data show an effect of this severe winter; another objective is to look for differences in survival between sexes. Our approach is based first on an age- and time-dependent analysis of resightings in the Camargue colony from 1978 to 1987 ($k - 1 = 10$) of 6622 flamingos ringed as chicks in this same colony from 1977 to 1986. This exploratory analysis is similar to that of Johnson et al. (1991), in that it looks for a general pattern of survival, using models that obviously do not fit the data. To concentrate on adult survival, we then considered the first resighting in the colony as the initial marking and restricted our attention to birds ≥5 yr old. By this age, a high proportion of the birds were confidently sexed. The data set consists then of further resightings in the colony, over a 5-yr time period ($k - 1 = 5$), of 284 males and 331 females.

TABLE 31. Age- and time-specific estimated annual survival rates ($\hat{\phi}$) from the 6622 Greater Flamingos (*Phoenicopterus ruber*) banded as young of the year from 1977 through 1986; rows are for annual cohorts, column headings correspond to age, diagonal elements correspond to rates at different ages (cohorts) within the same calendar year; these estimates were computed by RELEASE as unconstrained estimators. The 1986 cohort provides no separately estimable survival rate. The estimates for the severe winter 1984–1985 are boldface.

Year	Cohort	Age of bird (yr) 1	2	3	4	5	6	7	8	9
		Age- and year-specific $\hat{\phi}$								
1977	1	...*	...	...	1.081	0.965	0.962	1.034	**0.802**	0.969
1978	2	0.797	0.500	1.000	1.063	1.013	1.041	**0.733**	0.942	
1979	3	0.630	0.989	0.761	1.051	0.988	**0.766**	0.981		
1980	4	0.577	0.571	1.300	1.056	**0.649**	0.977			
1981	5	0.518	0.990	1.063	**0.690**	0.928				
1982	6	0.477	1.529	**0.540**	0.994					
1983	7	0.900	**0.309**	1.146						
1984	8	**0.141**	1.280							
1985	9	0.226								

* Insufficient data to compute any survival estimate.

Exploratory analysis of all the data

The full flamingo data set consists of 6622 banded birds from years 1977 through 1986. It is clear that there must be age effects at the younger ages and that there could be sex effects, except sex is only determined for resighted adult birds and often only after several resightings. In the full data set, sex is unknown for 4529 birds that were either never resighted, or not resighted frequently enough (or not sighted as breeding adults) to determine their gender.

Assuming there is no sex effect, the basic CJS model can be applied separately to each cohort of newly banded flamingos as a way to fit model ($\phi_{a\cdot t}$, $p_{a\cdot t}$). We separately apply the CJS model formulae for $\hat{\phi}_i$ and $\hat{p}_i$ to each release, and subsequent resightings, as if those were the only data we had. Each such cohort potentially produces survival estimates for 1-yr-old birds in year i, then 2-yr-old birds in year $i + 1$, then 3 yr olds in year $i + 2$, and so forth. Age and time effects are confounded within a single cohort, but because we have several different years of cohorts, we can substantially separate age from time effects. These survival estimates are given in Table 31, as computed by program RELEASE. The individual estimates were not constrained to be <1, because we intended to smooth these estimates at a second step in the analysis.

In Table 31, each row shows the $\hat{\phi}_i$ that came from that year's release of birds. For example, from the cohort released in 1979, $\hat{\phi}_1 = 0.630$, $\hat{\phi}_2 = 0.989$, $\hat{\phi}_3$, and so forth. These represent estimated survival rates at ages 1, 2, 3, and so forth. For each $\hat{\phi}$ in Table 31 there is a corresponding estimated standard error produced by RELEASE. The survival estimates in Table 31 may now be averaged (inversely weighted by their estimated sampling variances) in various ways to focus on different aspects of the data. Column averages represent (as well as is possible here) age-specific survival rates, ignoring year effects. Such age-specific rates still are partially confounded with year effects. The left-to-right upward diagonal of rates 0.141, 0.309, . . . , 0.802 are the rates that applied to 1st-, 2nd-, etc., year birds all in the severe winter of 1984–1985. A priori we are interested in comparing age-specific survival rates in this severe winter to those in the normal winters. Table 32 shows average (weighted) age-specific rates for the normal winter vs. the age-specific survival rates for the severe winter and estimated standard errors of these rates. The same data are shown, with a different emphasis, as a graph in Fig. 3. There is compelling evidence here of a winter effect and that this effect was very different by age. This analysis is in the spirit of an exploratory, or initial data analysis (Chatfield 1988), which could be formalized by the use of program SURGE.

As in all the other examples except roe deer, in which the studied population was enclosed, survival rates are apparent survival rates where the effects of permanent emigration and mortality are confounded. The values obtained are thus difficult to interpret in absolute terms, especially for the younger age classes, which are more prone to permanent emigration and more heterogeneous in their resighting rates. The model ($\phi_{a\cdot t}$, $p_{a\cdot t}$) is strongly rejected by the goodness-of-fit tests (Table 33),

TABLE 32. Average age-specific annual survival rates (and SEs estimated by RELEASE) of Greater Flamingos (*Phoenicopterus ruber*) in the severe (1984–1985) and in normal winters (from Table 31; see also Fig. 3).

Bird age	Winter type: Normal Mean	Normal SE	Severe Mean	Severe SE
1	0.558	0.017	0.141	0.070
2	0.876	0.103	0.309	0.158
3	1.091	0.065	0.541	0.117
4	1.044	0.023	0.690	0.078
5	0.964	0.079	0.650	0.060
6	0.981	0.020	0.776	0.040
7	0.996	0.021	0.733	0.045
8	0.942	0.034	0.802	0.042
9	0.969	0.030	...	

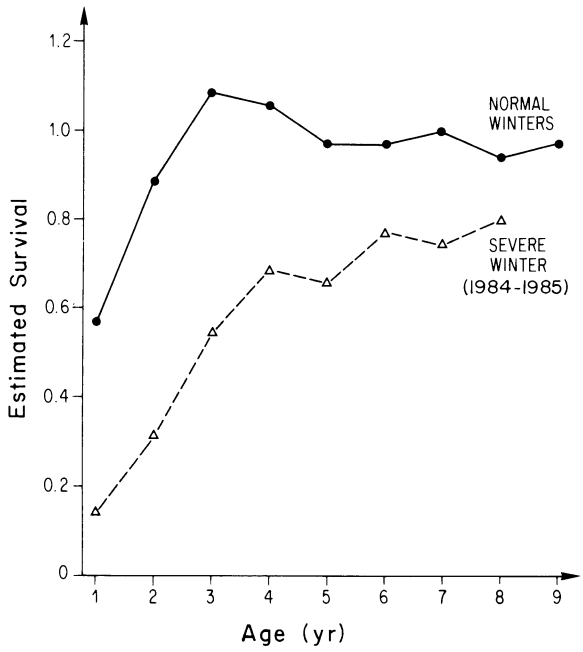

FIG. 3. Average age-specific survival rates of Greater Flamingos (*Phoenicopterus ruber*) in severe and normal winters (see Tables 31 and 32).

which causes us to consider these analyses as only exploratory. Within each cohort of flamingos marked as chicks, there are major heterogeneities. The lack of fit, if measured by the P value of the goodness-of-fit tests, depends strongly on the amount of data (i.e., the R_i). A more absolute method is to measure the amount of lack of fit by the ratio χ^2/df, which, on the average, equals 1 under the null hypothesis that the model fits. Here, this ratio equals 1.982 and 6.913 for TESTS 3 and 2, respectively. The high chi-square test statistic for TEST 2, even relative to the value for TEST 3, reveals a structural failure in the model, mostly because individuals marked as chicks return gradually to the colony: their first date of reobservation is spread over several years. On the contrary, individuals already reobserved at least once tend to be reseen in majority in the next year because of increasing faithfulness to the colony site and a high resighting pressure. The slow increase in the probability of reobservation of individuals not yet seen probably results mostly from an increase in the probability of presence rather than in the probability of reobservation of birds present in the colony. Such data can be used to estimate age-specific probabilities of presence on a breeding site or age-specific probabilities of breeding, with the help of more specific models that have been developed by Lebreton et al. (1990) and J. Clobert et al. (*unpublished manuscript*).

Because heterogeneity of capture rates decreases strongly after the first reobservation, a necessary step for further analysis of survival consisted of considering the first observation on the colony as the initial release of marked individuals. To remove as much heterogeneity as possible, we considered birds that were at least 5 yr old when first reobserved. Because of all these sources of heterogeneity, which we do not know how to model, we use here only data on individuals known to have returned to the study area as adults.

Analysis of adults only: basic tests and models

A summary of the resighting data for birds older than 5 yr is given in Table 34. The starting point for the analysis was model ($\phi_{t\cdot,s}$, $p_{t\cdot,s}$), i.e., the CJS model considered separately for each sex. Both TESTS 3 and 2 (Table 35) indicated that this model did not fit the data.

However, the lack of fit was less pronounced than in the overall analysis (TEST 2/df = 2.664, TEST 3/df = 2.597). Many reasons could be responsible for the remaining heterogeneity, which indicated, nevertheless, a strong departure from the **iii** hypothesis. Age, breeding status, and wintering habits (Johnson et al. 1990) may affect survival rates. Capture rates may vary with age and with breeding status, and also as a consequence of site tenacity within the colony. Differences in visibility between various parts of the colony and an increasing faithfulness with age could induce trap dependence, although the recaptures are only resight-

TABLE 33. Results (chi-square values), by cohorts, of the goodness-of-fit tests of model ($\phi_{a\cdot,t}$, $p_{a\cdot,t}$) for the Greater Flamingo (*Phoenicopterus ruber*) resighting data (overall analysis of flamingos marked as chicks). Nearly all components are significant, indicating a severe lack of fit to the model.

Cohort	χ^2, TEST 2	df	χ^2, TEST 3	df	Pooled χ^2 (TEST 2 + TEST 3)	df
1977	132.3278	15	13.7257	5	146.0535	20
1978	128.3303	13	11.8500	5	140.1803	18
1979	89.2250	12	13.3876	6	102.6126	18
1980	50.8853	9	1.3653	3	52.2506	12
1981	42.9379	8	5.0283	3	47.9662	11
1982	13.6031	6	1.5515	2	15.1546	8
1983	2.4621	2	0.3896	1	2.8517	3
1984	0.4399	1		0	0.4399	1
1985		0		0		0
1986		0		0		0
Total	460.2124	66	47.2980	25	507.5094	91

TABLE 34. A summary of Greater Flamingo (*Phoenicopterus ruber*) resighting data (A. R. Johnson et al., *unpublished data*) for birds older than age 5 yr seen at least once on the Camargue colony. Birds are classified by columns according to time of next resighting, and by rows according to year of previous resighting (i.e., the reduced *m* array of Burnham et al. 1987).

Females

Year	Released Number	Year next recaptured 1983	1984	1985	1986	1987	Never resighted
1982	32	27	4	0	0	0	1
1983	80		57	7	3	4	9
1984	112			58	18	4	32
1985	132				109	13	10
1986	211					169	42

Males

Year	Released Number	Year next recaptured 1983	1984	1985	1986	1987	Never resighted
1982	27	20	3	2	1	0	1
1983	85		56	9	4	6	10
1984	118			61	28	4	25
1985	152				120	17	115
1986	253					203	50

ings. Much more could be explored with regard to modeling age effects; we did not choose to tackle that problem here.

The kind of trap dependence expected is positive ("trap happiness"), in the sense of an increased probability of resightings of birds already observed. We looked for trap dependence using an overall difference in capture rate immediately after each reobservation, following Sandland and Kirkwood (1981), utilizing, however, a logit rather than log scale. Thus, we considered for each sex: logit $p = u_t + v_d$, with $d = 1$ the year after each capture, $d = 0$ otherwise (R. J. Pradel and J. D. Lebreton, *unpublished manuscript*). In this way we fitted, using SURGE, model $(\phi_{s\cdot t}, p_{s\cdot(t+d)})$. The comparison with model $(\phi_{s\cdot t}, p_{s\cdot t})$ seemed to indicate a strong trap dependence (difference in deviance: 11.260, df = 2). As in the roe deer example, a rough measure of fit of model $(\phi_{s\cdot t}, p_{s\cdot(t+d)})$ was obtained by subtracting the LRT $\chi^2_2 = 11.260$ from the goodness-of-fit χ^2, leading to $\chi^2_{20} = 46.277$ ($P = .0007$). Given the list of potential sources of heterogeneity mentioned above, it was not surprising to obtain a model that did not fit the data. Because neither model $(\phi_{s\cdot t}, p_{s\cdot t})$ nor model $(\phi_{s\cdot t}, p_{s\cdot(t+d)})$ fit the data, the difference in deviance $\chi^2_1 = 11.260$ could not be supposed to follow a χ^2 distribution.

The lack of fit revealed departure from the underlying assumptions leading to the multinomial and binomial models. These departures can be of two types: failure of the model *structure* to hold and (or) excess variation compared to that assumed by binomial theory (cf. Burnham et al. 1987:243–246, Moore 1987). Extra-binomial variation can be caused by heterogeneity of true capture and survival rates among individuals and (or) by failure of the statistical independence of the fates of individuals. In models for categorical data, these departures can be summarized in a single parameter *c*, a variance inflation factor, that indicates the amount of extra-binomial variation present in the data. In our case, *c* could, in theory, be estimated from the residual deviance of model $(\phi_{s\cdot t}, p_{s\cdot t})$ divided by its degrees of freedom, i.e., by the number of independent possible capture histories. Such an estimator is not robust because of the large number of capture histories, of which many have observed values equal to zero. The pooled chi-square test statistics (TEST 2 + TEST 3), with corresponding pooled rdf, (df2 + df3) = rdf, the residual degrees of freedom, provides a more robust estimator:

$$\hat{c} = \frac{\text{TEST 2} + \text{TEST 3}}{\text{rdf}}$$

(Burnham et al. 1987:246).

It seemed preferable to start from model $(\phi_{s\cdot t}, p_{s\cdot(t+d)})$, which seemed to account for a major structural feature and, thus, to use the rough estimates $\hat{c} = 46.277/20 = 2.314$, with rdf = 20. Such a value is not unusually high for models based on count data exhibiting excess binomial variation (see, e.g., Eberhardt 1978, Burnham et al. 1980:55, Williams 1982, Burnham et al. 1987:246–252, Moore 1987).

The likelihood ratio tests will tend to be inflated by extra binomial variation. The degree to which the LRT will be affected has not yet been explored. Standard considerations on quasi-likelihood theory would lead one to use of $\hat{c}$ as a reference for variation. The chi-square statistic given by (TEST 2 + TEST 3) is independent of the LRT between reduced models, which leads to the use of *F* tests. Instead of comparing the LRT statistic (on "df" degrees of freedom) to the distribution of χ^2_{df}, one would use as a test statistic the LRT/df statistic divided by $\hat{c}$ treated as an *F* test on df and rdf degrees of freedom:

$$F_{df,rdf} = \frac{\text{LRT/df}}{\hat{c}}.$$

TABLE 35. Results (chi-square values), by sex, of the goodness-of-fit tests of model $(\phi_{s\cdot t}, p_{s\cdot t})$ for the Greater Flamingo (*Phoenicopterus ruber*) resighting data, birds older than age 5 yr seen at least once on the Camargue colony (A. R. Johnson et al., *unpublished data*).

	χ^2	df	*P*
Males			
TEST 2	9.4066	3	.0243
TEST 3	14.3893	8	.0722
Females			
TEST 2	6.5762	3	.0867
TEST 3	27.1648	8	.0007
Total	57.5369	22	.0001

TABLE 36. Modeling capture rates of Greater Flamingos (*Phoenicopterus ruber*) older than 5 yr (A. R. Johnson et al., *unpublished data*). For each model, we give the deviance (DEV), number of estimable parameters (*np*), and Akaike's Information Criterion (AIC modified; see *Analysis of adults only: basic tests and models*). Models are compared by treating differences in deviance, divided by *np* and by the estimated overdispersion factor $\hat{c} = 2.314$ (df = 20), as F statistics. s = sex, t = time, and d = "trap dependence" (increased probability of reobservation of birds sighted the year before). For model notation, see the Appendix.

Model	*np*	DEV	Modified AIC	Comparison
(1) $(\phi_{s\cdot t}, p_{(d+t)\cdot s})$	20	1771.889	805.726	
				(2) vs. (1) inter. $(d + t)\cdot s$ $F_{4,20} = 0.364$, $P = .8313$
(2) $(\phi_{s\cdot t}, p_{d+t+s})$	16	1772.732	798.090	
				(3) vs. (2) effect of sex $F_{1,20} = 0.876$, $P = .3605$
(3) $(\phi_{s\cdot t}, p_{d+t})$	15	1774.758	**796.966**	
				(4) vs. (3) effect of time $F_{3,20} = 3.560$, $P = .0327$
(4) $(\phi_{s\cdot t}, p_{d})$	12	1797.911	800.971	
				(5) vs. (3) trap-dependence $F_{1,20} = 5.013$, $P = .0367$
(5) $(\phi_{s\cdot t}, p_{t})$	14	1786.359	799.979	

Similarly, the AIC should be modified as

$$\frac{\text{DEV}}{\hat{c}} + 2np.$$

We caution that these ideas are exploratory and not yet confirmed by fundamental statistical theory, but the ideas are consistent with quasi-likelihood theory.

Thus, the analysis continued from model $[\phi_{s\cdot t}, p_{s\cdot(t+d)}]$, with 20 estimable parameters, and relied on an estimated overdispersion factor $\hat{c} = 2.314$ (with rdf = 20 degrees of freedom) to test for the significance of differences in deviance between models using approximate F statistics. For this initial model, the deviance, DEV (still defined up to an additive constant as in all other examples) was 1771.889 (Table 36).

Analysis of adults only: modeling capture rates

The interaction between sex and (time + trap dependence), made up of the two interactions sex × time and sex × trap dependence, was not significant (model [1] vs. [2], $F_{4,20} = 0.364$, $P = .8313$, Table 36). The main effect of sex then could also be removed from the model (model [2] vs. [3], $F_{1,20} = 0.876$, $P = .3605$, Table 36). Neither time effect nor trap dependence could be removed from model $(\phi_{s\cdot t}, p_{t+d})$ (Table 36: model [3] vs. models [4] and [5], respectively. $F_{3,20} = 3.560$, $P = .0327$ and $F_{1,20} = 5.013$, $P = .0367$). Model $(\phi_{s\cdot t}, p_{t+d})$ had the lowest modified AIC among these five preliminary models.

Analysis of adults only: modeling survival rates

In a similar way, neither the sex × time interaction nor sex (as a main effect) influenced noticeably survival rates (Table 37: model [1] vs. [2] and [2] vs. [3]: $F_{4,20} = 0.142$, $P = .9635$ and $F_{1,20} = 0.575$, $P = .4652$, respectively). Then, time dependence in survival appeared as strong (model $[\phi_t, p_{t+d}]$ vs. model $[\phi, p_{t+d}]$, $F_{3,20} = 3.894$, $P = .0242$). This time dependence could be reduced in a straightforward way to the effect of the severe 1984–1985 winter (model $[\phi_t, p_{t+d}]$ vs. model $[\phi_w, p_{t+d}]$, $F_{2,20} = 0.210$, $P = .8123$, Table 37). The winter effect itself was strongly significant (model $[\phi_w, p_{t+d}]$ vs. model $[\phi, p_{t+d}]$, $F_{1,20} = 11.261$, $P = .0031$, Table 37). Model (ϕ_w, p_{t+d}) had the lowest modified AIC among all models considered.

Analysis of adults only: final model and tests

Having reached model (ϕ_w, p_{t+d}) with eight parameters (with the lowest modified AIC up to this point), it was time to retest for sex effect on survival and for trap dependence. Once again, there was a nonsignificant effect of sex on survival (model $[\phi_w, p_{t+d}]$, $F_{1,20} = 0.542$, $P = .4779$, Table 37). As expected, the difference was in favor of males: 0.980 against 0.972 in normal winters, 0.810 against 0.755 during the severe 1984–1985 winter. These differences resulted from the back transformation of a constant difference (0.323) on a logit scale because the sex × winter interaction was not included in the model. A one-sided test, considering that males would tend to survive better because of their larger size, was still nonsignificant.

Trap dependence was still significant (model $[\phi_w, p_{t+d}]$ vs. model $[\phi_w, p_t]$, $F_{1,20} = 5.230$, $P = .0332$, Table 37). The model (ϕ_w, p_{t+d}) was thus our final model. The estimates of survival and capture rates are given in Table 38. The results of model (ϕ_w, p_{t+d}) lead to two main conclusions. First, the strong time dependence in the capture rate resulted largely from a lower rate of presence of flamingos on the colony the year after the severe winter (Table 38), although an increasing efficiency of resighting over the years has also played a role. Second, the severe winter induced a reduction in survival of ≈ 0.19. Such a reduction in survival can be viewed as an ecological catastrophe for a long-lived species such as these flamingos.

Another consequence of extra-binomial variation that

TABLE 37. Modeling survival rates of Greater Flamingos (*Phoenicopterus ruber*) older than 5 yr (A. R. Johnson et al., *unpublished data*). For each model we give the deviance (DEV), number of estimable parameters (*np*), and Akaike's Information Criterion (AIC). Models are compared by treating differences in deviance, divided by *np* and by the estimated overdispersion factor $\hat{c} = 2.314$ (df = 20), as *F* statistics (see *Analysis of adults only: basic tests and models*). *s* = sex, *t* = time, and *d* = trap-dependence. For model notation, see the Appendix.

Model	*np*	DEV	Modified AIC	Comparison
(1) (ϕ_{s*t}, p_{d+t})	15	1774.758	796.966	
				(2) vs. (1) inter. $s \cdot t$ $F_{4,20} = 0.142$, $P = .9644$
(2) (ϕ_{s+t}, p_{d+t})	11	1776.074	789.534	
				(3) vs. (2) effect of sex $F_{1,20} = 0.575$, $P = .4571$
(3) (ϕ_t, p_{d+t})	10	1777.404	788.109	
				(4) vs. (3) effect of time $F_{3,20} = 3.894$, $P = .0242$
(4) (ϕ, p_{d+t})	7	1804.436	793.791	
				(5) vs. (3) reducing time effect to winter effect $F_{2,20} = 0.210$, $P = .8123$
				(5) vs. (4) effect of winter $F_{1,20} = 11.261$, $P = .0031$
(5) (ϕ_w, p_{d+t})	8	1778.377	**784.530**	
				(6) vs. (5) effect of sex $F_{1,20} = 0.542$, $P = .4702$
(6) (ϕ_{w+s}, p_{d+t})	9	1777.123	785.988	
				(7) vs. (5) trap-dependence $F_{1,20} = 5.230$, $P = .0332$
(7) (ϕ_w, p_t)	7	1790.479	787.759	

should be accounted for is that confidence intervals become wider. Basically, standard errors that are usually derived from theory, such as $\widehat{\text{SE}}[\text{logit}(\hat{\phi})]$, should be inflated (multiplied) by $\sqrt{\hat{c}}$, and then these standard errors should be treated as being estimated on the rdf degrees of freedom of $\hat{c}$. Confidence intervals then are based on the *t* distribution rather than the standard normal distribution. For instance, in this flamingo example, the confidence intervals based on the logit(ϕ) are $\text{logit}(\hat{\phi}) \pm (2.086)\ \widehat{\text{SE}}[\text{logit}(\hat{\phi})]$, where 2.086 is the *t* statistic on 20 df needed for a 95% CI, and $\widetilde{\text{SE}}(\text{logit}(\hat{\phi})) = \sqrt{\hat{c}}\ \{\widehat{\text{SE}}[\text{logit}(\hat{\phi})]\}$. Here, $\hat{c} = 2.314$, hence $\sqrt{\hat{c}} = 1.521$, and $\widehat{\text{SE}}[\text{logit}(\hat{\phi})]$ is the estimated theoretical standard error of $\text{logit}(\hat{\phi})$ based on the **iii** hypothesis under the fitted model. SURGE prints $\widehat{\text{SE}}[\text{logit}(\hat{\phi})]$.

Discussion

The flamingo data represent a difficult example in that many real world difficulties are present. The fact that chicks cannot be sexed at the time of ringing imposes severe limitations in the modeling and analysis. Preliminary analysis seems to make it clear that there are significant age effects in both ϕ and p and that substantial trap dependence exists in the capture probabilities. Heterogeneity in capture rates, especially for younger age classes, is substantial, and methods for modeling such heterogeneity have not yet been developed. Interesting covariates are available (e.g., winter conditions); such covariates should be incorporated in the analysis. The example illustrates some analysis directions and philosophy and helps the reader evaluate the evidence and make decisions concerning further analyses. More analytical work could clearly be done for these data. Hence, our example is preliminary; no final results are to be implied.

TABLE 38. Estimates of survival and capture rates, standard errors (SE), and approximate 95% confidence intervals (CI) for the Greater Flamingo (*Phoenicopterus ruber*) from the model (ϕ_w, p_{d+t}), applied to birds older than 5 yr resighted at least once on the colony (A. R. Johnson et al., *unpublished data*). The asymmetric confidence intervals are back-transformed from symmetric confidence intervals on logits. Both SE and CI account for an estimated overdispersion factor, $\hat{c} = 2.314$.

Survival rates			
Winter	$\hat{\phi}$	$\widetilde{\text{SE}}$	CI
Normal	0.973	0.027	[0.813, 0.997]
Severe (1984–1985)	0.785	0.046	[0.674, 0.866]
Capture rates			
Year	$\hat{p}$	$\widetilde{\text{SE}}$	
For birds seen in the previous season:			
1983	0.810	0.080	
1984	0.714	0.057	
1985	0.657	0.050	
1986	0.826	0.035	
1987	0.822	0.040	
For birds not seen in the previous season:			
1984	0.523	0.111	
1985	0.457	0.097	
1986	0.676	0.090	
1987	0.671	0.105	

While the analysis presented is both preliminary and only partially satisfactory, it also illustrates the need for a sophisticated approach, including intensive testing of assumptions and model fit. The use of a simple model will be unlikely to provide valid inferences. Worse yet would be the use of various ad hoc methods, usually assuming constant capture probability over age, sex, time, etc. (see Begon 1983). Invalid inference is the most likely outcome from the use of models that clearly do not fit the data. The approach advocated here includes biologically driven model building, model testing and selection, evaluation of model fit, and parameter estimation based on a "best" model. Then adjustment of precision for overdispersion may be required. These measures, while imperfect, help protect the integrity of the inferences made. The real world problems of analyzing capture–recapture data such as the flamingo data must be faced, rather than ignored by doing simplistic, ad hoc analyses.

The use of an overdispersion factor is clearly justified if the remaining lack of fit results from more or less random (i.e., nonsignificant) sources of heterogeneity or random biases. Here age effects may cause some nonrandom (i.e., significant) bias. Even with nonrandom bias, we believe it is proper to inflate the standard errors based on theory partially to account for this bias. It is possible that the lack of fit here could be reduced by use of models with a more complex structure for the particular data set under study, because the lack of fit indicates to us that there are subsets of animals exhibiting different behavior with respect to recapture. If we could have identified these subsets and defined particular parameters for them, mostly in terms of recapture rates, an improved model structure would result. However, when such structural features do not arise from a priori hypotheses, post hoc attempts at improving the model structure are likely to overfit the data. This overfitting will then risk producing very precise but badly biased parameter estimates whose resultant small standard errors will be misleading.

When the reported model does not fit we must be mostly concerned about bias of estimators. The concept of an ideal replicate of the data is relevant. Would true replication of this study lead to the same amount of lack of fit as in the original data? If so, would this lack of fit concern the same subsets of individuals? Differences in estimates arising, sample after sample, from the same structural failure, would induce a bias, whereas those arising from such random sources of heterogeneity do not bias estimators. This is why the confidence intervals are made larger. Which features should be kept in the model structure (e.g., trap dependence) and which can be considered as overdispersion is somewhat subjective, with no sharp borderline. Through the reduction of lack of fit to an overdispersion parameter, we provide a way to put the emphasis on external validity rather than on internal validity. The price to pay for lack of fit, in terms of either bias or precision, indicates that not much confidence can be placed in capture–recapture estimates based on a model that fits quite poorly. Thus, more attention is needed here, with more formal developments, in the case of capture–recapture models.

The estimated annual survival rate of flamingos in normal years is higher (0.973) than the values found in the literature (e.g., $\phi = 0.931$, Johnson et al. 1991), although our estimate has a large confidence interval (0.813, 0.997) which reflects a strong remaining heterogeneity. This increase in estimated survival results mostly from the fact that trap dependence is accounted for. When trap dependence is unaccounted for (model $[\phi_w, p_t]$), the estimated survival rates are lower (0.954 and 0.775 vs. 0.973 and 0.785, respectively, for normal and severe winters). Underestimation is indeed expected when this kind of positive trap dependence is neglected (see Sandland and Kirkwood 1981:538, Table 2). Contrary to Johnson et al. (1991), we were unable to detect a significant difference in survival between sexes, once again because we took account of the lack of fit induced by heterogeneity.

Several structural features could help improve the fit of the models, in relation to flamingo biology. First, age may affect survival, and this effect should be considered in further modeling. Second, resighting rates are likely to vary with the site used by each bird within the colony. Because of the tenacity to the breeding site, each bird could be attributed, at least approximately, to its site of first resighting. Some heterogeneity in capture rate could then be removed by considering such a site effect, with a limited cost in terms of number of parameters if the year × site interaction may be neglected. Such structural improvements could clearly be detrimental to external validity. Splitting the data to keep part of them for cross-validating more sophisticated models would then be recommended. A final run of the selected model with the total data would then provide final estimates of parameters and their standard errors. One might also expect in the future parametric models accounting for between-individual variation in capture rate, already available for closed population recapture models (Otis et al. 1978).

In terms of design, studies of this kind in the future may well consider possibilities of sexing chicks. The physical capture and marking of adults should be considered in parallel with marking of chicks. This would improve the ability to model not only survival rates but also recruitment rates (Lebreton et al. 1990). Last but not least, one cannot expect precise results for long-lived species unless the study period is long enough compared with the average life-span (see also the study of albatrosses by Weimerskirch et al. 1987).

Discussion

Survival analyses from capture–recapture data seem, at first view, to be a specific statistical methodology. The data are a particular case of categorical data; the model structure results from chains of specific events,

such as surviving or being captured at some date conditional on having survived. However, the recent development of a large array of models and of flexible software, which we synthesize here, has shown that the subject fits into a classical situation, common to many statistical problems, with a clear emphasis on model selection. We will discuss the model building, model selection, design issues, and suggested new work.

Models

Models are not reality, but at most, relevant tools to explore and, hopefully, satisfactorily answer specific questions about reality. "Even the most general mathematical model is a plaything relative to the complexities of an animal population" (Cormack 1968). A major problem with the Cormack-Jolly-Seber (CJS) approach for a single data set was its lack of relevance for the kinds of questions biologists want to address with the data, in particular, the comparisons of survival rates among groups. This lack of relevance probably explains why such specific questions have been repeatedly addressed by ad hoc methods, with a clear preference of biologists for relevance as opposed to statistical coherence (Lebreton 1989).

The CJS model structure remains at the core of all survival capture–recapture models. The synthesis presented here applies the CJS structure to groups or subcategories of animals (as an extreme case, cohorts), as anticipated by Jolly (1965). This application of a separate CJS model to all relevant subcategories of animals constitutes an appropriate global model to begin with. Many reduced models, based on constraints among the parameters of the global model, have been produced during the decade of the 1980s. This process of both generalization and specialization of the CJS model meets most of the limitations discussed by Cormack (1979:240–241), except for dealing with individual level heterogeneity in survival and capture parameters. Moreover, models for several groups (Burnham et al. 1987, Pradel 1988, Pradel et al. 1990) go one step further in bridging the gap between biological relevance and statistical soundness. These models are shown to fit into a common framework in which survival rate ϕ and capture rate p are modeled in generalized linear model philosophy (McCullagh and Nelder 1983, 1989).

The model-building techniques presented here and extensively used in the five examples encompass:

goodness-of-fit tests of the CJS model and nested models with fewer parameters than CJS;

flexible age and/or time-dependent variation in survival and/or capture rates;

flexible handling, using linear constraints on transformed parameters, of covariates and qualitative factors susceptible of affecting survival and capture rates, such as weather variables as main effects or as interactions;

flexible parameterization of multiple groups with some parameters in common across groups;

the possibility of accounting, in a robust way, for trap dependence and capture history effects;

the possibility of accounting for some lack of fit by estimating an overdispersion parameter (i.e., quasilikelihood).

This progress in capture–recapture modeling leads us to expand and advocate a simple notation proposed by Sandland and Kirkwood (1981); in particular, this notation includes models for several groups. We believe that this general model notation will help to refocus the attention from the technical aspects of the models onto their structure and meaning in terms of survival and capture rates.

It may be noted that several of the models considered here with capture probability $p \leq 1$ simplify to the classical models used in human epidemiology when p is forced to be 1. In particular, Cox's proportional hazards model (see, e.g., Cox and Oakes 1984) allows a straightforward capture–recapture generalization. This link with models used in human biology will probably be of benefit to both fields (epidemiology and ecology). In the same way, capture–recapture studies can also be linked to studies of survival in animal populations based on radio-tracking data and nest survival studies for which the use of models with $p = 1$ has been proposed (Pollock et al. 1989, White and Garrott 1990).

There are some important shortcomings to the present state of the art:

problems of identifiability may exist in complex models, making it difficult to determine their number of identifiable parameters;

standard, reliable goodness-of-fit tests of complex models, with sources of variation more complex than groups and time dependence, are not available;

general heterogeneity, in particular in capture rates, is not explicitly handled;

individually defined covariates cannot be modeled if they vary over time or if the individual changes group membership. This would require models where covariate values must be assigned to each individual, even if not captured.

In this general framework we advocate formal consideration of an array of models for the analysis of a particular set of data, including multiple data sets. The initial model should incorporate enough parameters and structure to serve as a global model. Biological knowledge should be the primary consideration in model building. For example, many avian populations have been studied and frequently indicate sex specificity in population parameters; thus this source of variation should often be in the global model. Similarly, capture rates frequently vary by age and time and these sources of variation should be considered in building the global model. Other models to be considered would then be special cases of this global model. Relatively minor effects, say the sex × year interaction, can be included in the array of models to be considered, but one must realize that the statistical significance of such

effects often cannot be detected with small sample sizes. Generally, biological knowledge plays a central role in model building.

Model selection

As seen in the five examples, the availability of a comprehensive modeling framework and of comprehensive software moves the emphasis from the fit of a particular model to procedures for model selection. The difficulties of model selection are common to many areas of applied statistics: multiple regression, categorical data analysis, analysis of variance, time-series analysis (e.g., Linhart and Zucchini 1986, Sakamoto et al. 1986, Dijkstra 1988). Most of the recommendations discussed have already been advocated in a more or less similar way in these other subject areas.

A priori biological information on the species and practical information on the design should be used to limit the number of models explored. For example, if we know that different numbers of person-hours have been spent in collecting data in different years, we need to incorporate a time-dependent capture rate. Moreover, in that case, the measure of effort should be recorded as a covariate and the capture rate should be modeled as a function of effort (e.g., Pollock et al. 1984, Clobert et al. 1987). If this variation in capture effort is expected to affect in a similar way the capture probability of, say, males and females, we may specifically drop the time $\times$ sex interaction in capture rate.

Model selection, given a biologically sound global model, is based on the principle of parsimony. This principle recognizes the importance of a trade-off between the bias resulting from the use of an overly simple model vs. the lack of prcision resulting from the use of an overly general model. The parsimony principle implies that one should risk some bias to gain precision. One should thus look for a reasonable compromise, knowing that a true model cannot be determined from a finite amount of data. Most difficulties of model selection arise from the large number of possible models to be handled in that search for a compromise. The first risk is that multiple tests are made between successive models, in such a way that the overall risk of incorrectly rejecting at least one of the many null hypotheses is much increased over the nominal α level. This risk of Type I error is enhanced by the lower power of many tests because low power motivates us to use a larger α level (such as .1 rather than .05) for individual tests. The second risk is that the selected model will be tailored to the data (i.e., over-fitting) in such a way that the external validity of the model, i.e., the possibility of correctly extrapolating the conclusions from a particular data set, is considerably decreased (see Cook and Campbell 1979 regarding internal vs. external validity).

Historically, the selection among models of the type presented here has been based on extensive comparisons of LRTs and goodness-of-fit tests by persons expert in such methodology. This is a difficult process, essentially becoming a de facto implementation of an expert system applied to a maze of models. Modern model selection has recast the entire problem as a classical optimization problem. Within a likelihood-based paradigm, Akaike (1973) has shown that AIC is the theoretically appropriate objective function to use in approaching model selection as an optimization problem: the best model is the one with the smallest AIC. We compared the performance of AIC on our five examples (and other data sets) vs. our own expert selection of the model. We found that AIC performed exceedingly well in routinely selecting the same, or almost the same, model as we did. Therefore we reached the conclusion that AIC is a reliable, objective basis for screening the set of possible models to select a best model and its near neighbors for further study. The final judgment on the choice of a model(s) can be left to the investigator within the constraint that the model should be very close to AIC-optimal (no more than about one unit larger than optimal AIC).

We recommend use of Akaike's Information Criterion here as a way to assist in selecting a basic model from the global model. Then specific biological questions can be addressed by using only a few formal tests between this AIC-selected model and neighboring ones, thus limiting the increase in the overall risk of rejection of at least one null hypothesis otherwise caused by multiple tests. The results obtained in our examples with this strategy, and the theoretical background of Akaike's Information Criterion, lead us to advocate its use for model selection in capture–recapture studies. We note, however, that even this strategy of model selection cannot totally avoid the trap of overfitting the model to the data. Finally, we note that there are reasonable competitors to AIC (Schwarz 1978). The important point, however, is that model selection should now be viewed as an optimization problem.

Design

Capture–recapture studies have most often been thought of in terms of surveys rather than experiments. The possiblity of considering groups within the same modeling framework in addition to the classical sources of variation such as time, and the possibility of separating main effects from interactions, opens the way to both more refined and more specific designs. The full consequences of such a change have yet to be drawn. Nevertheless, a few basic guidelines can already be given.

The design of a study using capture–recapture sampling should start with a clear identification of objectives (Hayne 1978). A list of the factors and variables likely to influence survival and capture rates will be useful when designing a study. Such a list will include factors of direct interest (i.e., related to biological questions, most often about survival rates; see, however, the Swift example), and factors which are not avoidable

in the practical context of the study, and which will be represented by nuisance parameters. Particular attention should be given to the possibilities of parsimonious modeling of such nuisance factors using, e.g., a measure of effort to model capture rate. The number of animals released, R_i, should be as large as practical; there is little risk that too many animals will be released! Sampling effort should be as high as possible, to ensure that the capture probabilities, p_i, are high. Large R_i and p_i guard against inadequate model selection, and help improve the power of tests and the precision of the estimates (Rexstad et al. 1991). Calculations of power can be made beforehand to check if the questions under examination can be reasonably addressed with the planned design.

Sampling effort (e.g., the number of people, number of hours or days, number of traps used) and other covariates (particularly at the individual animal level: age, sex, mass, condition, etc.) should be carefully recorded. It is critical that the investigator use a capture history matrix as a summarization of the data. As portable microcomputers are becoming increasingly available, data entry, checking, editing, and storage will benefit from this technology. An applied statistician knowledgeable in capture–recapture should often be consulted for both the design and analysis of capture–recapture or capture–resighting studies, especially if the population issues are complex. The theory for the analysis of such studies is now fairly sophisticated, with an extensive and widely scattered literature. Seeking expertise in these areas is advisable.

The number of capture occasions is an important design consideration. Three occasions are the absolute minimum needed to estimate one survival rate, namely ϕ_1. No meaningful inferences about population dynamics can be made based on one estimated survival rate. Even after 4 or 5 occasions, one is just starting to have enough data to test assumptions and do a model selection as well as to begin to understand any dynamics. Hence, we recommend the investigator think in terms of getting 10, or better yet, 20 survival rates. This means 11 or 21 capture occasions. The value of the data increases with each year the study continues, so when occasions are yearly, continue capture–recapture studies as long as possible: these are long-term, not short-term studies.

Long-term studies do invite problems of mark loss. It is imperative that there be no mark (band, tag, etc.) loss. It is important that the act of marking does not affect ϕ because such an effect cannot be tested for. Also, it is not possible to test for the validity of the inference from the marked to the unmarked population; hence the validity of this inference must be established by the design and conduct of the study.

Using the framework proposed here, the effect of population manipulations can be assessed at the functional level of survival rates in a strictly experimental way (e.g., the lizard example). Such experiments have been properly analyzed up to now only when the capture rate is equal to 1, e.g., studies of nest survival after brood manipulation. When such a deliberate experiment, contrasting a control group vs. a treatment group or groups, is planned, then proper replication is particularly important (Burnham et al. 1987:240–278).

Conducting capture–recapture studies at randomly selected sites is nearly always not feasible because several such sites could have few if any animals available for study. However, it may be quite possible for large studies to have several study sites chosen nonrandomly. Analysis of capture–recapture data from multiple sites must deal with additional variance components issues (i.e., between-site variation in true parameters vs. within-site sampling variances of parameter estimators), additional interactions (i.e., site × year), and hence more problems of parsimonious modeling (e.g., some parameters need to be common across sites). These problems are offset by statistical inferences being directed at a larger geographic area. SURGE will be valuable in the analysis of multi-site data.

When only one site is used there is a danger of committing pseudoreplication (Hurlbert 1984). For example, if the site were partitioned into four subsites, these would not constitute four true replicates. Partitioning the set of capture histories randomly would also be pseuroreplication as concerns inference to other sites. However, such data partitioning could be useful for getting empirical sampling variances at the one site. In the typical capture–recapture study at one site, statistical theory only justifies inferences to that site. In extending our inferences about survival rates to the species as a whole, we are assuming no, or minor, site effects on survival.

Monte Carlo simulation can be effective in planning a capture–recapture study or experiment (Burnham et al. 1987:319–324). Wilson et al. (1989) presented software to compute approximate sample sizes needed to achieve desired levels of precision. Model bias and precision of estimators can be assessed in a preliminary manner, particularly if expected values are input into the software (this is feasible using RELEASE and SURGE). Rexstad et al. (1990) provide an extensive example of the use of simulation in study design.

Future opportunities

New developments will appear, in at least four directions: first, new methods of marking are appearing; second, the model-building and computations will be made easy for the users of estimators and test statistics; third, some technical (statistical) shortcomings of the present procedures will be overcome; and fourth, more sophisticated models will appear as deeper biological questions will be asked through more sophisticated experiments.

New methods of marking animals help to refine designs in capture–recapture studies and experiments (e.g., passive integrated transponder (PIT) tags, genetic

markers, etc.). New taxonomic groups become amenable to study, capture rates are increased, the system under study is less and less modified by less harmful marking techniques. In general, sample sizes will be increased. The resulting increase in power will accelerate the need for more sophisticated modeling, in particular in what concerns heterogeneity.

The second point is largely related to the availability of computer software. Software must become increasingly easy to use, not only for handling statistical models, but also for data management. For both design and analysis issues, such software will need to include calculations of analytical test powers and easy-to-use Monte Carlo procedures. New software using computationally intensive methods such as resampling approaches and profile likelihood intervals will become available as the speed of microcomputers increases still further: such approaches are needed to give more realistic confidence intervals, robust against the shrinkage effects on standard errors induced by model selection. With large data sets, cross-validation should be used in the model selection process. This type of procedure can be most useful in protecting against using models with too many parameters for a given data set, to warrant the external validity of conclusions drawn from a particular data set.

Regarding improvements in statistical procedures, the situation will probably become very parallel to that of epidemiological modeling. In this field, after an explosive development symbolized by Cox's proportional hazards model and by Mantel's log rank test (see, e.g., Cox and Oakes 1984), a number of shortcomings have been solved by further research refining such central tools. In capture–recapture modeling, one may expect progress on such directions as: more general goodness-of-fit tests, finite sample properties of statistical procedures such as likelihood ratio tests, relative merits of the various kinds of confidence intervals, influence of model selection on these confidence intervals (shrinkage), information criteria refining over AIC (e.g., Read and Cressie 1988:127 or Schwarz 1988), generalized smoothing procedures applied to survival estimates.

About the fourth point, more sophisticated models will broaden biological relevance. One of the most promising directions will be to allow capture–recapture models to include recruitment and dispersal. Such models would be able to address questions relative to trade-offs between dispersal and survival, and to the regulation of spatially organized populations. More generally, multi-site models, and models considering individual covariates susceptible of changing through time, provide a challenging direction of development. The estimation of population size and its modeling as a function of various external variables can also be developed in a framework similar to that used here. However, such generalizations will only be possible under restrictive assumptions (Burnham 1988). Lastly, one can expect, in a broader perspective, models progressively mixing population dynamics based on life cycle features (e.g., the Leslie matrix model) and statistical models for the estimation of rates of the type presented here (Anderson et al. 1990). Such a framework would lead to models for the growth rate (i.e., fitness) of classes of differing individuals.

Conclusion

Capture–recapture sampling is a widely used methodology applicable to a large number of population problems and taxonomic groups. The recent shift in emphasis from population size estimation to survival rate estimation that took place in capture–recapture modeling has been paralleled by a similar shift for population biologists. Survival rates are considered key parameters in studies of the evolution of demographic strategies, as well as in conservation biology or in game management. The refined questions arising in this context can no longer be addressed by ad hoc calculations, which can be grossly misleading. Statistical theory and comprehensive software now exist for most types of capture–recapture studies and experiments, and provide powerful ways to address efficiently many questions in animal population dynamics, ranging from evolutionary to conservation biology.

Acknowledgments

This work was done under the U.S.–France Cooperative Science Program. This research was supported by grant number INT-8714519 from the National Science Foundation, Division of International Programs, and by grant number 8992/0036 from the Centre National de la Recherche Scientifique, Direction des Relations et de la Coopération Internationale.

R. Pradel was most helpful in preparing software for the analysis of multiple data sets (program SURGE). Drs. R. G. Anthony, R. M. Cormack, C. M. Perrins, and P. M. North provided comments and suggestions on an earlier draft of the manuscript. In particular, the extensive review comments of Drs. C. Brownie and M. J. Conroy provided on the ms. copy submitted for publication significantly helped improve this monograph. We appreciate the assistance of these people.

The five examples would not have been possible without the data sets that others provided us to use to illustrate the theory. We are grateful to the following individuals and organizations for allowing us to use their data sets: G. Marzolin (European Dipper); G. Gory (Swift); B. Boisaubert, J. M. Gaillard, and the Office National de la Chasse (roe deer); T. Pilorge (common lizard); and A. R. Johnson and the Tour du Valat Biological Station (Greater Flamingo).

D. R. Smith aided in the manuscript preparation. Special appreciation is extended to B. A. Knopf for editorial help and manuscript preparation.

Literature Cited

Aitkin M., D. Anderson, B. Francis, and J. Hinde. 1988. Statistical modelling in GLIM. Clarendon, Oxford, England.

Aitkin, M., and M. Stasinopoulos. 1989. Likelihood analysis of a binomial sample size problem. Pages 399–411 *in* L. J. Gleser, M. D. Perlman, S. J. Press, and A. R. Sampson, editors. Contributions to probability and statistics: essays in honor of Ingram Olkin. Springer-Verlag, New York, New York, USA.

Akaike, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 *in* B. N. Petran and F. Csáki, editors. International Symposium on Information Theory. Second edition. Akadémiai Kiadi, Budapest, Hungary.

Anderson, D. R., J. Bart, T. C. Edwards, Jr., C. B. Kepler, and E. C. Meslow. 1990. 1990 status review: northern spotted owl (*Strix occidentalis caurina*). United States Unpublished report. United States Fish and Wildlife Service, Washington, D.C., USA.

Anderson, D. R., K. P. Burnham, and G. C. White. 1985. Problems in estimating age-specific survival rates from recovery data of birds ringed as young. Journal of Animal Ecology **54**:89–98.

Anderson, S., A. Auquier, W. W. Hauck, D. Oakes, W. Vandaele, and H. I. Weisberg. 1980. Statistical methods for comparative studies. John Wiley & Sons, New York, New York, USA.

Arnason, A. N., and K. H. Mills. 1987. Detection of handling mortality and its effect on Jolly-Seber estimates for mark–recapture experiments. Canadian Journal of Fisheries and Aquatic Sciences **44**(Supplement 1):64–73.

Arnason, A. N., and C. J. Schwarz. 1986. POPAN-3: extended analysis and testing features for POPAN-2. Charles Babbage Research Centre, St. Pierre, Manitoba, Canada.

Barbault, R. 1981. Biologie des populations et des peuplements. Masson, Paris, France.

Barndorff-Nielsen, O. E. 1986. Inference on the full or partial parameters based on the standardized log likelihood ratio. Biometrika **73**:307–322.

Bauwens, D., B. Heulin, and T. Pilorge. 1987. Variations spatio-temporelles des caractéristiques démographiques dans et entre populations du lézard *Lacerta vivipara*. Pages 531–536 *in* Colloque National du CNRS Biologie des Populations, Université Claude-Bernard, Villeurbanne, France.

Begon, M. 1983. Abuses of mathematical techniques in ecology: application of Jolly's capture–recapture method. Oikos **40**:155–158.

Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland. 1975. Discrete multivariate analysis: theory and practice. MIT Press, Cambridge, Massachusetts, USA.

Blondel, J. 1986. Biogéographie évolutive. Masson, Paris, France.

Boiseaubert, B., J. Vassant, and D. Delorme. 1979. Contribution a l'étude de la mise au point d'une méthode de recensement applicable a l'espèce chevreuil (*Capreolus capreolus*) vivant en milieu forestier. Supplement, numéro spécial scientifique et technique, December 1979. Bulletin Mensuel de l'Office National de la Chasse: 193–205.

Bonney, G. E. 1987. Logistic regression for dependent binary observations. Biometrics **43**:951–973.

Box, G. E. P., and G. M. Jenkins. 1970. Time series forecasting and control. Holden-Day, San Francisco, California, USA.

Boyce, M. S., and C. M. Perrins. 1987. Optimizing Great Tit clutch size in a fluctuating environment. Ecology **68**: 142–153.

Breslow, N. E., and N. E. Day. 1980. Statistical models in cancer research. Volume 1. The analysis of case-control studies. International Agency of Research on Cancer, Lyon, France.

Brownie, C., D. R. Anderson, K. P. Burnham, and D. S. Robson. 1985. Statistical inference from band recovery data—a handbook. United States 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. 1976. Models allowing for age-dependent survival rates from band-return data. Biometrics **32**:305–323.

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.

———. 1982. A mark–recapture survival analysis. Journal of Animal Ecology **51**:833–847.

Burnham, K. P. 1988. A unified approach to animal release–resampling studies of survival processes and population estimation. North Carolina State University Institute of Statistics Technical Report Series Number **1698.**

———. 1989. Numerical survival rate estimation for capture–recapture models using SAS PROC NLIN. Pages 416–435 *in* L. McDonald, B. Manly, J. Lockwood, and J. Logan, editors. Estimation and analysis of insect populations. Springer-Verlag, New York, New York, USA.

Burnham, K. P., D. R. Anderson, and J. L. Laake. 1980. Estimation of density from line transect sampling of biological populations. Wildlife Monographs **72**:1–202.

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 release–recapture. American Fisheries Society Monograph **5.**

Burnham, K. P., G. C. White, and D. R. Anderson. 1984. Estimating the effect of hunting on annual survival rates of adult mallards. Journal of Wildlife Management **48**:350–361.

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

Charlesworth, B. 1980. Evolution in age structured populations. Cambridge University Press, Cambridge, England.

Chatfield, C. 1988. Problem solving: a statistician guide. Chapman and Hall, London, England.

Clobert, J. 1981. Etude de la dynamique d'une population d'Etourneaux sansonnets. Dissertation. Université catholique de Louvain, Louvain-la-neuve, Belgium.

Clobert, J., and J. D. Lebreton. 1985. Dépendance de facteurs de milieu dans les estimations de taux de survie par capture-recapture. Biometrics **41**:1031–1037.

Clobert, J., and J. D. Lebreton. 1987. Recent models for mark–recapture and mark resighting data: a response to C. Brownie. Biometrics **43**:1019–1022.

Clobert, J., and J. D. Lebreton. 1991. Estimation of demographic parameters in bird population dynamics. Pages 75–104 *in* C. M. Perrins, J. D. Lebreton, and G. Hirons, editors. Bird population studies: their relevance to conservation and management. Oxford University Press, Oxford, England.

Clobert, J., and J. D. Lebreton, and D. Allainé. 1987. A general approach to survival rate estimation by recaptures or resightings of marked birds. Ardea **75**:133–142.

Clobert, J., J. D. Lebreton, M. Clobert-Gillet, and H. Coquillart. 1985. The estimation of survival in bird populations by recaptures or resightings of marked individuals. Pages 197–213 *in* B. J. T. Morgan and P. M. North, editors. Statistics in ornithology. Springer-Verlag, New York, New York, USA.

Clobert, J., J. D. Lebreton, and G. Marzolin. 1990. The estimation of local juvenile survival rates and of age-specific proportions of breeders in bird populations. Pages 199–213 *in* J. Blondel, A. Gosler, J. D. Lebreton, and R. McCleery, editors. Population biology of passerine birds: an integrated approach. NATO ASI Series G, Volume 24. Springer, Berlin, Germany.

Clutton-Brock, T. H., editor. 1988. Reproductive success: studies of individual variation in contrasting breeding systems. University of Chicago Press, Chicago, Illinois, USA.

Clutton-Brock, T. H., F. E. Guinness, and S. D. Albon. 1982. Red deer: behavior and ecology of two sexes. University of Chicago Press, Chicago, Illinois, USA.

Clutton-Brock, T. H., M. Major, and F. E. Guinness. 1985. Population regulation in male and female red deer. Journal of Animal Ecology **54**:831–846.

Conroy, M. J., and B. K. Williams. 1984. A general methodology for maximum likelihood inference from band recovery data. Biometrics **40**:739–748.

Conroy, M. J., J. E. Hines, and B. K. Williams. 1989. Procedures for the analysis of band-recovery data and user instructions for program MULT. United States Fish and Wildlife Service, Resource Publication **175.**

Cook, T. D., and D. T. Campbell. 1979. Quasi-experimentation: design and analysis issues for field settings. Houghton Mifflin, Boston, Massachusetts, USA.

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

———. 1968. The statistics of capture–recapture methods. Oceanography and Marine Biology **6**:455–506.

———. 1979. Models for capture–recapture. Pages 217–255 *in* R. M. Cormack, G. P. Patil, and D. S. Robson, editors. Sampling biological populations. International Cooperative Publishing House, Fairland, Maryland, USA.

———. 1981. Loglinear models for capture–recapture experiments on open populations. Pages 217–235 *in* R. W. Hirons and D. Cooke, editors. The mathematical theory of the dynamics of biological populations. II. Academic Press, London, England.

———. 1985. Examples of the use of GLIM to analyze capture–recapture studies. Pages 243–273 *in* B. J. T. Morgan and P. M. North, editors. Statistics in ornithology. Springer-Verlag, New York, New York, USA.

———. 1989. Loglinear models for capture–recapture. Biometrics **45**:395–413.

Coulson, J. C., and R. D. Wooller. 1976. Differential survival rates among breeding Kittiwake Gulls *Rissa tridactyla* (L.). Journal of Animal Ecology **45**:205–213.

Cox, D. R., and D. Oakes. 1984. Analysis of survival data. Chapman and Hall, London, England.

Cox, D. R., and E. J. Snell. 1989. Analysis of binary data. Chapman and Hall, New York, New York, USA.

Cramp, S., editor. 1985. The birds of the western Palearctic. Volume 4. Oxford University Press, Oxford, England.

Cramp, S., and K. E. L. Simmons, editors. 1977. The birds of the western Palearctic. Volume 1. Oxford University Press, Oxford, England.

Crawley, M. J., and R. M. May. 1987. Population dynamics and plant community structure: competition between annuals and perennials. Journal of Theoretical Biology **125**: 475–489.

Critchley, F., I. Ford, and O. Rijal. 1988. Interval estimation based on the profile likelihood: strong Lagrangian theory with applications to discrimination. Biometrika **75**:21–28.

Crosbie, S. F., and B. F. J. Manly. 1985. Parsimonious modelling of capture–mark–recapture studies. Biometrics **41**:399–410.

Dhondt, A. A., and R. Eyckerman. 1980. Competition and the regulation of numbers in Great and Blue Tit. Ardea **68**: 121–132.

Dijkstra, T. K., editor. 1988. On model uncertainty and its statistical implications. Volume 307. Lecture notes in economics and mathematical statistics. Springer-Verlag, New York, New York, USA.

Draper, N. R., and H. Smith. 1981. Applied regression analysis. John Wiley & Sons, New York, New York, USA.

Drost, F. C., W. C. M. Kallenberg, D. S. Moore, and J. Oosterhott. 1989. Power approximation to multinomial tests of fit. Journal of the American Statistical Association **84**: 130–141.

Durbin, J. 1987. Statistics and statistical science. Journal of the Royal Statistical Society, Series A **150**:177–191.

Eberhardt, L. L. 1978. Appraising variability in population studies. Journal of Wildlife Management **42**:207–238.

Edgington, E. S. 1980. Randomization tests. Marcel Dekker, New York, New York, USA.

Edwards, A. W. F. 1972. Likelihood. Cambridge University Press, Cambridge, England.

Endler, J. A. 1986. Natural selection in the wild. Monographs in Population Biology **21.**

Fienberg, S. E. 1981. The analysis of cross-classified categorical data. MIT Press, Cambridge, Massachusetts, USA.

Finney, D. J. 1971. Probit analysis. Third edition. Cambridge University Press, London, England.

Flack, V. F., and P. C. Chang. 1987. Frequency of selecting noise variables in subset regression analysis: a simulation study. American Statistician **41**:84–86.

Gaillard, J. M. 1988. Contribution à la dynamique des populations de grands mammifères: l'exemple du chevreuil (*Capreolus capreolus*). Dissertation. Université Lyon I, Villeurbanne, France.

Galbraith, H., and S. J. Tyler. 1982. The movements and mortality of the dipper as shown by ringing recoveries. Ringing and Migration **4**:9–14.

Glutz Von Blotzheim, U., and K. Bauer. 1980. Handbuch der Vogel Mitteleuropas. 9. Akademische Verlag, Wiesbaden, Germany.

Glutz Von Blotzheim, U., and K. Bauer. 1985. Handbuch der Vogel Mitteleuropas. 10/11. Aula, Wiesbaden, Germany.

Gnanadesikan, R., editor. 1983. Statistical data analysis. Proceedings of the Symposia in Applied Mathematics. Volume 28. American Mathematical Society, Providence, Rhode Island, USA.

Goodman, L. A. 1984. The analysis of cross-classified data having ordered categories. Harvard University Press, Cambridge, Massachusetts, USA.

Granjon, L. 1987. Evolution allopatrique chez les Muridés: mécanismes écoéthologiques liés au syndrome d'insularité chez *Mastomys* et *Rattus.* Dissertation. Montpellier II Université, Montpellier, France.

Harding, C. M., A. W. Heathwood, R. G. Hunt, and K. L. Q. Read. 1984. The estimation of animal population size by the removal method. Applied Statistics **33**:196–202.

Hayne, D. W. 1978. Experimental designs and statistical analyses in small mammal population studies. Pages 3–10 *in* D. P. Snyder, editor. Populations of small mammals under natural conditions. Special Publication Series, Number 5, Pymatuning Laboratory of Ecology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA.

Hines, J. E., and J. R. Sauer. 1989. Program CONTRAST: a general program for the analysis of several survival or recovery rate estimates. United States Fish and Wildlife Service, Technical Report **24.**

Hosmer, D. W., and S. Lemeshow, Jr. 1989. Applied logistic regression. John Wiley and Sons, New York, New York, USA.

Hudson, D. J. 1971. Interval estimation from the likelihood function. Journal of the Royal Statistical Society, Series B **33**:256–262.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs **54**: 187–211.

Hurvich, C. M., and C. L. Tsai. 1990. The impact of model selection on inference in linear regression. American Statistician **44**(3):214–217.

Jackson, C. H. N. 1933. On the true density of tsetse flies. Journal of Animal Ecology **2**:204–209.

———. 1939. The analysis of an animal population. Journal of Animal Ecology **8**:236–246.

Johnson, A. R. 1983. Etho-Ecologie du Flamant rose (*Phoenicopterus ruber roseus,* Pallas) en Camargue et dans l'Ouest paléarctique. Dissertation. Université Paul Sabatier, Toulouse, France.

Johnson, A. R., R. E. Green, and G. J. M. Hirons. 1991.

Survival rates of Greater Flamingos in the west Mediterranean region. Pages 249–271 *in* C. M. Perrins, J. D. Lebreton, and G. J. M. Hirons, editors. Birds population studies: their relevance to conservation and management. Oxford University Press, Oxford, England.

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

———. 1982. Mark–recapture models with parameters constant in time. Biometrics **38**:301–321.

Jolly, G. M., and J. M. Dickson. 1980. Mark-recapture suite of programs. Pages 570–576 *in* M. M. Barritt and D. Wishart, editors. COMPSTAT 1980: Proceedings in Computational Statistics. Volume 4. Physica-Verlag, Vienna, Austria.

Kalbfleisch, J. D., and D. A. Sprott. 1970. Applications of likelihood methods to models involving large numbers of parameters. Journal of the Royal Statistical Society, Series B **32**:175–208.

Kleinbaum, D. G., and L. L. Kupper. 1978. Applied regression analysis and other multivariate methods. Duxbury Press, North Sciutate, Massachusetts, USA.

Leamer, E. E. 1978. Specification searches: ad hoc inference with nonexperimental data. John Wiley & Sons, New York, New York, USA.

Lebreton, J. D. 1989. Statistical methodology for the study of animal populations. Bulletin of the International Statistical Institute **53**(1):267–282.

Lebreton, J. D., and J. Clobert. 1986. User's guide to program SURGE, Version 2.0. Document polycopié CEFE/CNRS, Montpellier, France.

Lebreton, J. D., G. Hemery, J. Clobert, and H. Coquillart. 1990. The estimation of age-specific breeding probabilities from recaptures or resightings in vertebrate populations. I. Transversal models. Biometrics **46**:609–622.

Linhart, H., and W. Zucchini. 1986. Model selection. John Wiley & Sons, New York, New York, USA.

Loery, G., and J. D. Nichols. 1985. Dynamics of a Black-capped Chickadee population, 1958–1983. Ecology **66**:1195–1203.

Lomnicki, A. 1988. Population ecology of individuals. Monographs in Population Biology **25.**

Manly, B. F. J. 1985. The statistics of natural selection on animal populations. Chapman and Hall, London, England.

Marzolin, G. 1988. Polygynie du Cincle plongeur (*Cinclus cinclus*) dans les côtes de Lorraine. L'Oiseau et la Revue Francaise d'Ornithologie **58**:277–286.

Massot, M. 1987. Influence de la densité sur la démographie d'une population de lézards vivipares du Mont Lozère. Unpublished report. Available from Ecole Normale Supérieure. Paris, France.

McCullagh, P., and J. A. Nelder. 1983. Generalized linear models. Chapman and Hall, New York, New York, USA.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman and Hall, New York, New York, USA.

McCullagh, P., and D. Pregibon. 1985. Discussion comments on the paper by Diaconis and Efron. Annals of Statistics **13**:898–900.

Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. Introduction to the theory of statistics. Second edition. McGraw-Hill, New York, New York, USA.

Moore, D. F. 1987. Modelling the extraneous variance in the presence of extra-binomial variation. Journal of the Royal Statistical Society, Series C **36**:8–14.

Morgan, B. J. T., and S. N. Freeman. 1989. A model with first-year variation for ring-recovery data. Biometrics **45**: 1087–1101.

Nichols, J. D., G. R. Hepp, K. H. Pollock, and J. E. Hines. 1987. The Husting dilemma: a methodological note. Ecology **68**:213–217.

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

Noble, I., and R. O. Slatyer. 1980. The use of vital attributes to predict successional changes in plant communities subject to recurrent disturbances. Vegetatio **43**:5–21.

North, P. M., and B. J. T. Morgan. 1979. Modelling heron survival using weather data. Biometrics **35**:667–681.

Ortega-Rubio, A., T. Pilorge, M. Khodadoost, and L. Arriagal. 1989. Interpopulation home range comparison of a temperate lizard. Herpetology **20**:71–80.

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

Perrins, C. M. 1971. Age of first breeding and adult survival rates in the Swift. Bird Study **18**:61–70.

Peterman, R. M. 1990. The importance of reporting statistical power: the forest decline and acidic deposition example. Ecology **71**:2024–2027.

Pilorge, T. 1988. Dynamique comparée de populations de lézards vivipares: régulation et variabilité intra- et interpopulationnelle. Dissertation. Université Paris VI, Paris, France.

Pilorge, T., J. Clobert, and M. Massot. 1987. Life history variations according to sex and age in *Lacerta vivipara.* Pages 311–315 *in* J. J. van Gelder, H. Strybosch, and P. J. M. Bergers, editors. Proceedings of the Fourth Ordinary General Meeting of the Societas Europaea Herpetologica. Faculty of Sciences, Nijmegen, The Netherlands.

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

———. 1981. Capture–recapture models allowing for age-dependent survival and capture rates. Biometrics **37**:521–529.

———. 1982. A capture–recapture design robust to unequal probability of capture. Journal of Wildlife Management **46**: 757–760.

Pollock, K. H., J. E. Hines, and J. D. Nichols. 1984. The use of auxiliary variables in capture–recapture and removal experiments. Biometrics **40**:329–340.

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., and R. H. K. Mann. 1983. Use of an age-dependent mark–recapture model in fisheries research. Canadian Journal of Fisheries and Aquatic Sciences **40**:1449–1455.

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

Pollock, K. H., S. R. Winterstein, and M. J. Conroy. 1989. Estimation and analysis of survival distributions for radio-tagged animals. Biometrics **45**:99–109.

Pradel, R. 1988. Nouvelles approches statistiques en capture–recapture. Un example: le test de parallélisme. Unpublished report. Available from Montpellier II University, Montpellier, France.

Pradel, R. J., J. Clobert, and J. D. Lebreton. 1990. Recent developments for the analysis of multiple capture–recapture data sets: an example concerning two blue tit populations. Ring, *in press.*

Pregibon, D. 1984. Book review of McCullagh and Nelder, *Generalized linear models.* 1983. Annals of Mathematical Statistics **12**:1589–1596.

Rao, C. S. 1972. Linear statistical inference and its applications. John Wiley & Sons, New York, New York, USA.

Rawlings, J. O. 1988. Applied regression analysis: a research tool. Wadsworth & Brooks/Cole, Pacific Grove, California, USA.

Read, T. R. C., and N. A. C. Cressie. 1988. Goodness-of-

fit statistics for discrete multivariate data. Springer-Verlag, New York, New York, USA.
Rexstad, E. A., K. P. Burnham, and D. R. Anderson. 1990. Design of survival experiments using marked animals: a case study. American Fisheries Society Symposium 7:581–587.
Robson, D. S. 1969. Mark–recapture methods of population estimation. Pages 120–140 *in* N. L. Johnson and H. Smith, Jr., editors. New developments in survey sampling. John Wiley and Sons, New York, New York, USA.
Sakamoto, Y., M. Ishiguro, and G. Kitagawa. 1986. Akaike information criterion statistics. KTK Scientific, Tokyo, Japan.
Sandland, R. L., and P. Kirkwood. 1981. Estimation of survival in marked populations with possibly dependent sighting probabilities. Biometrika **68**:531–541.
Sauer, J. R., and B. K. Williams. 1989. Generalized procedures for testing hypotheses about survival or recovery rates. Journal of Wildlife Management **53**:137–142.
Schwarz, C. J. 1988. Post-release stratification and migration models in band-recovery and capture–recapture model. Dissertation. University of Manitoba, Winnipeg, Manitoba, Canada.
Schwarz, C. J., K. P. Burnham, and A. N. Arnason. 1988. Post-release stratification in band-recovery models. Biometrics **44**:765–785.
Schwarz, G. 1978. Estimating the dimensions of a model. Annals of Statistics **6**:461–464.
Searle, S. R. 1971. Linear models. John Wiley and Sons, New York, New York, USA.
Seber, G. A. F. 1965. A note on the multiple recapture census. Biometrika **52**:249–259.
———. 1982. The estimation of animal abundance and related parameters. Second edition. Macmillan, New York, New York, USA.
———. 1986. A review of estimating animal abundance. Biometrics **42**:267–292.
Shibata, R. 1989. Statistical aspects of model selection. Pages 215–240 *in* J. C. Williams, editor. From data to model. Springer-Verlag, New York, New York, USA.
Soulé, M. E. 1987. Viable populations for conservation. Cambridge University Press, New York, New York, USA.
Spaepen, J. F. 1988. Estimation of the survival rates of Meadow Pipits—a comparison of two different methods. Ringing and Migration **9**:117–128.
Stearns, S. C. 1976. Life history tactics: a review of ideas. Quarterly Review of Biology **51**:3–47.
———. 1980. A new view of life history evolution. Oikos **35**:266–281.
Stokes, S. L. 1984. The Jolly-Seber model applied to age-stratified populations. Journal of Wildlife Management **48**: 1053–1059. (Also see the corrigendum in Journal of Wildlife Management **49**:282.)
Stone, M. 1974. Cross-validation choice and assessment of statistical predictions. Journal of the Royal Statistical Society B **36**:111–133.
Tinbergen, J. M., J. H. van Balen, and H. M. van Eck. 1985. Density-dependent survival in an isolated Great Tit population: Kluijver's data reanalyzed. Ardea **73**:38–48.
Venzon, D. J., and S. H. Moolgavkar. 1988. A method for computing profile-likelihood based confidence intervals. Applied Statistics **37**:87–94.
Wedderburn, R. W. M. 1974. Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. Biometrika **61**:439–447.
Weimerskirch, H., J. Clobert, and P. Jouventin. 1987. Survival in five southern albatrosses and its relations with their life-history. Journal of Animal Ecology **56**:1043–1055.
White, G. C. 1983. Numerical estimation of survival rates from band-recovery and biotelemetry data. Journal of Wildlife Management **47**:716–728.
———. 1986. Program SURVIV user's manual, Version 1.2. Department of Fishery and Wildlife Biology, Colorado State University, Fort Collins, Colorado, USA.
White, G. C., D. R. Anderson, K. P. Burnham, and D. L. Otis. 1982. Capture–recapture and removal methods for sampling closed populations. Los Alamos National Laboratory, LA-8787-NERP, Los Alamos, New Mexico, USA.
White, G. C., and R. A. Garrott. 1990. Analysis of wildlife radio-telemetry data. Academic Press, New York, New York, USA.
Wilkinson, G. N., and C. E. Rogers. 1973. Symbolic description of factorial models for analysis of variance. Applied Statistics **22**:392–399.
Williams, D. A. 1982. Extra-binomial variation in logistic linear models. Applied Statistics **31**:144–148.
Wilson, K. R., J. D. Nichols, and J. E. Hines. 1989. A computer program for sample size computations for banding studies. United States Fish and Wildlife Service, Technical Report **23.**

APPENDIX

Linear constraints in capture–recapture models compared with ANOVA models: the case of two factors, which for capture–recapture models may be age (a) and time (t).*

ANOVA	Model notation	Capture–recapture†	
		Survival‡	Model notation
$E(y_{ijk}) = \mu$	$\cdots$	$f^{-1}(\phi) = c$	ϕ
$E(y_{ijk}) = \mu + \alpha_i$	A	$f^{-1}(\phi) = g_a$	ϕ_a
$E(y_{ijk}) = \mu + \beta_j$	B	$f^{-1}(\phi) = h_t$	ϕ_t
$E(y_{ijk}) = \mu + \alpha_i + \beta$	$A + B$	$f^{-1}(\phi) = g_a + h_t$	ϕ_{a+t}
$E(y_{ijk}) = \mu + c_{ij}$		$f^{-1}(\phi) = u_{at}$	
$= \mu + \alpha_i + \beta_j + \gamma_{ij}$	$A*B = A + B + A \cdot B$	$= g_a + h_t + m_{at}$	$\phi_{a*t} = \phi_{a+t+a \cdot t}$

* The notation used here for linear predictors is described in detail by McCullagh and Nelder (1983:41ff). Main effects of factors and covariates are represented by letters (e.g., a, t, d). Interaction between factors is denoted by a dot (e.g., $a \cdot t$). When an interaction is included in addition to main factors (e.g., $a + t + a \cdot t$), the corresponding quantity may vary freely over all combinations of factors a and t. The formula $a + t + a \cdot t$ is denoted for short by $a * t$. This notation expands readily to more factors. For instance, $a + t + s + t \cdot s = a + t * s$ (where a stands for age, t for time, s for sex) denotes variation which is parallel between age classes over dates *within* each sex, and which is parallel between age classes and across sexes, for each date.

† Capture–recapture models of this type can be fit using built-in multiple regression on dummy variables in SURGE. Details on the use of dummy variables in multiple regression for fitting ANOVA models can be found in Draper and Smith (1981: 241 and 447) and Kleinbaum and Kupper (1978:Chapter 13).

‡ Similar models can be used for recapture. f is an appropriate link function.