

A new method for analysing discrete life history data with missing covariate values

E. A. Catchpole

University of New South Wales at the Australian Defence Force Academy, Canberra, Australia, and University of Kent, Canterbury, UK

and B. J. T. Morgan and G. Tavecchia

University of Kent, Canterbury, UK

[Received October 2005. Final revision September 2007]

Summary. Regular censusing of wild animal populations produces data for estimating their annual survival. However, there can be missing covariate data; for instance time varying covariates that are measured on individual animals often contain missing values. By considering the transitions that occur from each occasion to the next, we derive a novel expression for the likelihood for mark–recapture–recovery data, which is equivalent to the traditional likelihood in the case where no covariate data are missing, and which provides a natural way of dealing with covariate data that are missing, for whatever reason. Unlike complete-case analysis, this approach does not exclude incompletely observed life histories, uses all available data and produces consistent estimators. In a simulation study it performs better overall than alternative methods when there are missing covariate data.

Keywords: Complete-case analysis; Life history data; Maximum likelihood; Missing data; Renewal process; Survival analysis; Time varying individual covariates; Trinomial distribution

1. Introduction

In this paper we provide a new way to deal with missing covariate data when modelling the survival of wild animals. The assessment of how covariates that are measured on individuals may affect survival is essential, and the most important such covariates are usually time varying. Typically, time varying individual covariates contain missing values, and there is therefore a need for a methodology to deal with this.

1.1. Data

We analyse mark–recapture–recovery (MRR) data, in which animals are initially marked, and then on subsequent occasions (usually annual) are either recaptured or resighted (alive) or recovered (dead). For each individual, the life history is a string of values, one for each occasion, with 1 indicating the initial capture or a subsequent recapture, 0 indicating that the animal was not encountered on that occasion and 2 indicating a dead recovery. The data comprise the set of life histories, together with the values of any relevant covariates; see, for example, Williams *et al.* (2002). We take a Jolly–Seber rather than a Cormack–Jolly–Seber approach to the modelling (see, for example, Manly *et al.* (2005)), so the initial capture of an animal is not modelled: the

Address for correspondence: E. A. Catchpole, School of Physical, Environmental and Mathematical Sciences, University of New South Wales at the Australian Defence Force Academy, Canberra, ACT 2600, Australia.
E-mail: e.catchpole@adfa.edu.au

likelihood is conditional on the initial capture. This focuses attention on the estimation of annual mortality, rather than population size, and enables mortality rates to be related to covariates. Covariates are of two types: extrinsic covariates which apply to all the animals, e.g. weather and population density, and intrinsic covariates, which depend on characteristics of individual animals. Intrinsic covariates may be time invariant, such as sex, or time varying, such as weight or reproductive status; for general discussion see Pollock (2002).

1.2. Missing covariate values in mark–recapture–recovery data

There are currently several possible ways of handling covariates with missing values in MRR data.

- (a) A ‘complete-case’ analysis simply omits all animals which have any missing covariate data. However, such an analysis can result in a substantial loss of information and can also result in severe bias (Abraham and Russell, 2004).
- (b) Bonner and Schwarz (2006) used a diffusion model to impute missing covariate values of weight. The accuracy of this approach depends on the goodness of the imputation model.
- (c) A last observation carried forward analysis replaces unknown covariate values with their last available value (see, for example, Catchpole, Fan, Morgan, Clutton-Brock and Coulson (2004)). This imputation method has clear potential for bias if there is a trend in the covariate, as may occur with weight for example; see also Prentice (1982) and Gadbury *et al.* (2003).
- (d) The guide to the widely used computer package MARK (White and Burnham, 1999) recommends unconditional mean imputation (Little and Rubin (2002), page 61), in which missing individual covariate values are replaced by the mean of the variable for the sample that is measured. This imputation method shrinks individual differences, which is undesirable. Disadvantages of this and the previous imputation methods are given in Molenberghs and Verbeke (2005), pages 492–494.
- (e) Catchpole, Morgan and Coulson (2004) used a binomial analysis, including only those individuals that are seen on a particular occasion, and based on the conditional probabilities of being seen alive or dead on that occasion. This method requires separate estimates of the recapture and recovery probabilities on each occasion and, in common with classical use of imputation methods, makes no allowance for the errors in estimating these.

In this paper we propose and evaluate a new conditional method, again using only those individuals that are seen on a particular annual occasion, but now based on the probabilities of the possible outcomes (seen alive, found dead or not seen) on the *following* occasion. The new approach does not exclude incompletely observed life histories and uses all available data. We show that this easily implemented method produces consistent estimators and, by means of a simulation study, that it performs better overall than alternative methods.

2. Model and likelihood

Suppose that we have information on n animals from k MRR occasions, at times t_1, t_2, \dots, t_k . Let $c_i, i = 1, \dots, n$, denote the occasion on which animal i is initially captured and marked, and $l_i, c_i \leq l_i \leq k$, the occasion on which it is last known to be alive, either through being seen alive on that occasion but never seen subsequently, live or dead, or through being found dead on the following occasion. We assume that, if an animal dies in (t_r, t_{r+1}) , then it is either found in this time interval or is never found—it cannot be found dead later. This is a common and valid

assumption in most studies of birds and mammals, but it may not be a good model for shellfish, for example (Catchpole *et al.*, 2001).

The definitions which follow are all for the i th animal. For $r = c_i, \dots, k - 1$, let

$$\begin{aligned} \phi_{i,r} &= \Pr(\text{alive at } t_{r+1} | \text{alive at } t_r), \\ p_{i,r} &= \Pr(\text{recaptured at } t_{r+1} | \text{alive at } t_{r+1}), \\ \lambda_{i,r} &= \Pr\{\text{found dead during } (t_r, t_{r+1}) | \text{died during } (t_r, t_{r+1})\}. \end{aligned}$$

Different models result from making different assumptions regarding the structure of these parameters. The usual focus is on survival, regarding the other model parameters as nuisance parameters. Often survival is related to covariates through a suitable link function, and the logistic function is usually used. We do not in this paper consider model selection but present a general framework within which standard procedures of inference can take place.

The traditional form for the likelihood, conditional on the initial capture event, is (Catchpole *et al.*, 1998; Catchpole, Fan, Morgan, Clutton-Brock and Coulson, 2004; King and Brooks, 2003)

$$L = \prod_{i=1}^n \{(1 - \phi_{i,l_i})\lambda_{i,l_i}\}^{d_i} \chi_{i,l_i}^{1-d_i} \prod_{r=c_i}^{l_i-1} \{\phi_{i,r} p_{i,r}^{w_{i,r}} (1 - p_{i,r})^{1-w_{i,r}}\}. \tag{1}$$

Here d_i indicates the known death of animal i (1 if found dead; 0 otherwise), $w_{i,r}$ is the indicator variable for being seen alive at t_{r+1} , and $\chi_{i,r}$, which is a function of the other parameters, is the probability that animal i is never seen, alive or dead, from t_{r+1} onwards, having been alive at t_r ; see also Skalski *et al.* (1993) and Hoffman and Skalski (1995).

3. A three-state process

3.1. A new expression for the likelihood

Here we present a new expression for the likelihood for MRR data, which is based on a three-state process. It is this expression that is central to the new methodology of the paper and which provides a new approach for dealing with missing covariate data.

To derive the new form for the likelihood, we extend the definition of $\chi_{i,r}$ to

$$\chi_{i,r,s} = \Pr(\text{not found, alive or dead, from } t_{r+1} \text{ to } t_s \text{ inclusively} | \text{alive at } t_r),$$

for $s = r + 1, \dots, k$, with $\chi_{i,r,r} = 1$, so that $\chi_{i,r} = \chi_{i,r,k}$. Then we have the recurrence relation

$$\chi_{i,r,s} = (1 - \phi_{i,r})(1 - \lambda_{i,r}) + \phi_{i,r}(1 - p_{i,r})\chi_{i,r+1,s}, \quad c_i \leq r < s \leq k, \tag{2}$$

which enables $\chi_{i,r,s}$ to be calculated.

Let us now denote by $h_{i,r}$ the life history data entry at t_r for animal i , i.e.

$$h_{i,r} = \begin{cases} 0, & \text{if the animal was not seen at } t_r, \text{ and not previously found dead,} \\ 1, & \text{if the animal was seen alive at } t_r, \\ 2, & \text{if the animal is known to be dead at } t_r. \end{cases}$$

This notation combines the life process and the observation process. Approaches to missing data in which these processes are kept separate are also possible (see, for example, Dupuis (1995)).

We then define

$$\pi_{i,r}(a, b) = \Pr(h_{i,r+1} = b | h_{i,r} = a, h_{i,r-1}, \dots, h_{i,1}), \tag{3}$$

which, for brevity, suppresses the dependence on the history before t_r . It is simple to derive the following expressions for $\pi_{i,r}(a, b)$. In each of the cases where $a = 0$, l denotes the last occasion,

before occasion r , on which the animal was seen alive. We have, for $c_i \leq r \leq k - 1$,

$$\pi_{i,r}(0, 0) = \chi_{i,l,r+1} / \chi_{i,l,r}, \tag{4a}$$

$$\pi_{i,r}(0, 1) = \prod_{s=l}^{r-1} \phi_{i,s} (1 - p_{i,s}) \phi_{i,r} p_{i,r} / \chi_{i,l,r}, \tag{4b}$$

$$\pi_{i,r}(0, 2) = \prod_{s=l}^{r-1} \phi_{i,s} (1 - p_{i,s}) (1 - \phi_{i,r}) \lambda_{i,r} / \chi_{i,l,r}, \tag{4c}$$

$$\pi_{i,r}(1, 0) = \chi_{i,r,r+1}, \tag{5a}$$

$$\pi_{i,r}(1, 1) = \phi_{i,r} p_{i,r}, \tag{5b}$$

$$\pi_{i,r}(1, 2) = (1 - \phi_{i,r}) \lambda_{i,r}, \tag{5c}$$

and $\pi_{i,r}(2, 2) = 1$. Note that the observations form a renewal process, renewing each time that state 1 occurs, and that the conditioning in equation (3) only need go back as far as the last live sighting.

These probabilities enable the likelihood to be constructed directly from the life history data on each of the n animals, in the form

$$L = \prod_{i=1}^n \prod_{r=c_i}^{k-1} \prod_{a=0}^2 \prod_{b=0}^2 \pi_{i,r}^{x_{i,r}(a,b)}(a, b) \tag{6}$$

where $x_{i,r}(a, b)$ is an indicator, which is equal to 1 if $h_{i,r} = a$ and $h_{i,r+1} = b$, and 0 otherwise.

As explained below, terms can be missing from this product if there are missing covariates. When all terms are present, the two likelihood expressions (1) and (6) are equivalent, as we demonstrate in Appendix A. But, when some values are missing, expression (6) has a distinct advantage, as shown in Section 3.3, where we present two ways of dealing with missing individual covariates, neither of which involves imputation.

3.2. Missing covariate data

If covariates are introduced by means of logistic regression of the survival probabilities $\phi_{i,r}$ on the covariates, then the survival probabilities are missing when there are missing individual covariate values. For example, survival from t_r to t_{r+1} might be modelled in terms of an animal's weight at t_r , which is known only if the animal was captured then. However, it is not always true that the covariates are measured when the animal is observed—for example, an animal might be captured but not weighed on a particular occasion. Nor is it always true that covariate values are missing when the animal is not observed—an extrinsic covariate, such as a rainfall, will be recorded whether or not an animal is seen.

If missing covariate data are not imputed, then one way of dealing with individuals with missing covariate values is to omit those individuals from the likelihood and to perform a complete-case analysis. The alternative approach of this paper is to omit transition probabilities from the likelihood if and only if these transitions cannot be calculated.

When modelling animal survival, the focus is different from the standard approach (Rubin, 1976; Little and Rubin, 2002) for missing values in longitudinal data. There, for each unit (person) there is a longitudinal vector of intended observations, $\mathbf{Y} = (\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{mis}})$, which is partitioned into observed and missing values, together with a missingness indicator vector \mathbf{M} . The

likelihood is the probability distribution $f(\mathbf{y}_{\text{obs}}, \mathbf{m})$. The selection model (Rubin, 1976) or selection factorization (Diggle *et al.*, 2007) is $f(\mathbf{y}, \mathbf{m}) = f(\mathbf{y}) f(\mathbf{m}|\mathbf{y})$. The missing value mechanism is informative if the probability distribution $f(\mathbf{m}|\mathbf{y})$ depends on \mathbf{y}_{mis} as well as on \mathbf{y}_{obs} ; the mechanism is data missing at random if $f(\mathbf{m}|\mathbf{y})$ depends only on \mathbf{y}_{obs} (Little and Rubin (2002), page 12).

Joint modelling of \mathbf{M} and \mathbf{Y} was initiated by Wu and Carroll (1988), and in this context a variety of ways of analysing $f(\mathbf{m}|\mathbf{y})$ have been proposed; for instance, Diggle and Kenward (1994) used logistic regression, Henderson *et al.* (2000) employed linked random effects and used an EM algorithm, whereas Guo and Carlin (2004) used Markov chain Monte Carlo methods. Saha and Jones (2005) investigated the asymptotic bias in mixed effects linear models for longitudinal data when there are non-ignorable patterns of missingness which are not properly modelled. For reviews and extensions of this work see for instance Abraham and Russell (2004), Hogan *et al.* (2004) and Diggle *et al.* (2007). In terms of the Little and Rubin framework, we have informative missing values, as we show in the next section, and have modelled the missingness process, in a similar way to Diggle and Kenward (1994), who modelled dropout.

Our work differs from these references in two ways. Firstly, ecological modelling of MRR data uses a three-state vector \mathbf{h} , which is defined in Section 3.1, in place of the indicator missingness vector \mathbf{m} . Secondly, the ecological focus is usually on modelling $\mathbf{h}|\mathbf{Y}_{\text{obs}}$ rather than on modelling \mathbf{Y}_{obs} . Indeed, for simplicity, in this paper we regard \mathbf{y}_{obs} as being known, without error, so that the likelihood is simply $f(\mathbf{h}|\mathbf{y}_{\text{obs}})$. This obviates the need for any assumptions on the missingness process, such as data missing at random.

We now illustrate how the complete-case and three-state likelihoods differ when there are missing covariate values.

3.3. Comparing complete-case and three-state likelihoods

Consider an example with $k = 6$ MRR occasions, and with $\phi_{i,r}$ depending on an intrinsic covariate $V_{i,r}$, which is measured if and only if animal i was captured at t_r . Suppose that this animal has the life history $(h_{i,1}, \dots, h_{i,6}) = (1, 0, 1, 1, 0, 0)$. Dropping the subscript i for clarity, the contribution of this animal to likelihood (1) is

$$\phi_1(1 - p_1)\phi_2p_2\phi_3p_3\chi_4, \tag{7}$$

whereas the contribution to the three-state likelihood (6) is

$$\pi_1(1, 0) \pi_2(0, 1) \pi_3(1, 1) \pi_4(1, 0) \pi_5(0, 0). \tag{8}$$

Because the animal is not seen on occasions 2 and 5, V_2 and V_5 (and hence ϕ_2 , ϕ_5 and χ_4) are unknown, and so the likelihood contribution (7) is unknown. In a complete-case analysis this animal would be deleted from the likelihood. Similarly, in expression (8), $\pi_2(0, 1)$ and $\pi_5(0, 0)$ are unknown. But, because expression (8) is based on conditional (transition) probabilities, we can simply omit these transitions from the likelihood, to leave

$$\pi_1(1, 0) \pi_3(1, 1) \pi_4(1, 0).$$

Thus, when there are missing covariate values, the three-state method that is based on equation (6) uses more of the available information than the complete-case method that is based on the traditional likelihood (1).

Also, in this example, the probability that the animal is missed at t_3 is $(1 - \phi_2)(1 - \lambda_2) + \phi_2(1 - p_2)$. Since ϕ_2 depends on V_2 , which is missing, the missingness depends on unobserved data. Thus the missingness process is data not missing at random.

3.4. A trinomial distribution

Suppose that we include in the three-state likelihood only those transitions from an occasion on which the animal was seen alive. Then the likelihood (6) reduces to the partial likelihood

$$L = \prod_{i=1}^n \prod_{r=c_i}^{k-1} \prod_{b=0}^2 \pi_{i,r}^{x_{i,r}(1,b)}(1, b). \tag{9}$$

This is a simple product trinomial distribution, with probabilities given by equations (5). If animal i being seen or captured at t_r is a necessary and sufficient condition for all the covariates that are relevant to $\phi_{i,r}$, $p_{i,r}$ and $\lambda_{i,r}$ to be known, then this is exactly equivalent to the three-state likelihood. An example of this is given in the previous section. Cases where the more general three-state likelihood is required include where an extrinsic covariate is missing on a particular occasion, or when an intrinsic covariate has values that are missing owing to observer error.

Note that if we have only dropout, rather than intermittent missing values, then equation (9) corresponds to equation 11.3.7 in Diggle *et al.* (1994).

3.5. Consistency

Estimators from the three-state and trinomial likelihoods are consistent, since they are maximum likelihood estimators for proper (conditional) distributions. For example, in the trinomial case, the likelihood can be written as

$$\prod_{i=1}^n \prod_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} \pi_{i,r}(1, h_{i,r+1}). \tag{10}$$

Although we have conditional probabilities, the proof follows standard lines, as in Cox and Hinkley (1974), page 287. Details are given in Appendix B. In the three-state likelihood, the conditioning event changes from $h_{i,r} = 1$ to the covariate value for animal i being known at t_r , but consistency follows in the same way. Note that the above argument does not extend to the complete-case analysis as in that case the corresponding probabilities do not result in a proper distribution.

In the general case, where the covariate \mathbf{Y}_{obs} is also modelled, the consistency of the trinomial estimators can be shown similarly to Diggle *et al.* (1994), section 11.3, who dealt only with dropout rather than intermittent missing values. There they gave the full likelihood $f(\mathbf{y}, \mathbf{m})$, modelling the probability of dropout. We can do the same, until the first 0 (missing value) in the life history. But, since the missingness process ‘renews’ when the next 1 (live recapture) occurs, we can regard this as the beginning of a new record, and then proceed as before.

4. Alternative approaches

4.1. Partial-case analysis

Instead of using the standard likelihood (1), but considering only complete case histories, as in a complete-case analysis, we could consider partial cases by simply omitting all missing terms from likelihood (1). In the example that is given in Section 3.3, leaving out all missing values from the likelihood (7) would leave us with the likelihood contribution $\phi_1(1 - p_1)p_2\phi_3p_3$.

4.2. A binomial distribution

Illius *et al.* (1995), Moorcroft *et al.* (1996) and Milner *et al.* (1999) analysed MRR data by using a binomial approach, taking each annual occasion in turn and looking only at those animals

which are found, either live or dead, on that occasion, and then assuming that the probabilities of being alive or dead are in the ratio $\phi : (1 - \phi)$. Catchpole, Morgan and Coulson (2004) pointed out that this approach is incorrect whenever the recapture and recovery probabilities p and λ are not equal, since the conditional probabilities are in the ratio $\phi p : (1 - \phi)\lambda$. They then used the resulting binomial likelihood to model ϕ , assuming that p and λ , which are in general time varying, are known on each occasion, possibly by having been estimated from a separate MRR analysis using the traditional likelihood (1) without covariate regressions. By excluding on each occasion those animals with missing time varying individual covariates for that occasion, this approach deals naturally with missing individual covariates, in a similar way to the three-state likelihood (6).

4.3. A four-state process

An alternative conditional approach to likelihood (6) is based on the observation that, in the life history (1, 0, 1, 1, 0, 0), the 0 on occasion 2 is qualitatively different from those on occasions 5 and 6. Although on each of these occasions the animal is not seen, on occasion 2 it is known to be alive, since it is seen alive on a subsequent occasion. The same conclusion would follow if the animal had later been found dead, since recoveries are assumed to be immediate, as noted in Section 3. It is natural to want to use this information. We can rewrite the sample life history above as (1, 3, 1, 1, 0, 0) by defining new states

$$h_{i,r}^* = \begin{cases} 0, & \text{if animal } i \text{ is not seen, alive or dead, at } t_s, \text{ for all } r \leq s \leq k, \text{ and has not been} \\ & \text{found dead before } t_r, \\ 1, & \text{if animal } i \text{ is seen alive at } t_r, \\ 2, & \text{if animal } i \text{ is known to be dead at } t_r, \\ 3, & \text{if animal } i \text{ is not seen, alive or dead, at } t_r, \text{ but is seen alive at } t_s, \text{ for some} \\ & r < s \leq k. \end{cases}$$

If we define probabilities $\rho_{i,r}(a, b)$ in a similar way to $\pi_{i,r}(a, b)$ in Section 3.1, then the likelihood can be written in terms of this four-state process as

$$L = \prod_{i=1}^n \prod_{r=c_i}^{k-1} \prod_{a=0}^3 \prod_{b=0}^3 \rho_{i,r}^{x_{i,r}(a,b)}(a, b) \tag{11}$$

where the indicator variables $x_{i,r}(a, b)$ are suitably redefined in terms of the new states. However, the states $h_{i,r}^*$ are defined in terms of future occurrences, and the probabilities $\rho_{i,r}(a, b)$ involve conditioning on these future occurrences. It is easily shown that $\rho_{i,r}(0, 0) = \rho_{i,r}(2, 2) = 1$, and

$$\rho_{i,r}(1, 0) = \chi_{i,r}, \tag{12a}$$

$$\rho_{i,r}(1, 1) = \phi_{i,r} p_{i,r}, \tag{12b}$$

$$\rho_{i,r}(1, 2) = (1 - \phi_{i,r}) \lambda_{i,r}, \tag{12c}$$

$$\rho_{i,r}(1, 3) = \phi_{i,r} (1 - p_{i,r}) (1 - \chi_{i,r+1}), \tag{12d}$$

$$\rho_{i,r}(3, 0) = 0, \tag{13a}$$

$$\rho_{i,r}(3, 1) = \phi_{i,r} p_{i,r} / (1 - \chi_{i,r}), \tag{13b}$$

$$\rho_{i,r}(3, 2) = (1 - \phi_{i,r}) \lambda_{i,r} / (1 - \chi_{i,r}), \tag{13c}$$

$$\rho_{i,r}(3, 3) = \phi_{i,r} (1 - p_{i,r}) (1 - \chi_{i,r+1}) / (1 - \chi_{i,r}). \tag{13d}$$

The four-state process is an alternative way to model time varying individual covariates with missing values. When there are no missing terms in expression (11), the four-state likelihood is

identical to the traditional likelihood, as we show in Appendix A. The four-state process is superficially attractive as it appears to use more information than the three-state process. However, it has disadvantages compared with the three-state process, as the following section illustrates.

4.4. Missing values in the four-state process

Consider again the example of Section 3.3, with the life history now written as (1, 3, 1, 1, 0, 0). The animal is seen, and so the covariates that are required to model subsequent annual survival probability are known, at times t_1 , t_3 and t_4 . So we might expect to be able to calculate the transition probabilities at each of these times, as we can with the three-state process. However, in the four-state process the transition probabilities from $t_1 \rightarrow t_2$ and $t_2 \rightarrow t_3$ both involve $\chi_{i,2}$, which depends on $\phi_{i,2}$, and so are unknown. Also the $t_4 \rightarrow t_5$ transition probability, $\chi_{i,4}$, depends, via the recurrence relation (2), on $\phi_{i,5}$, and so it also is unknown. The net result is that the only known transition probability is $\rho_{i,3}(1, 1)$.

In general, when a covariate determining survival probability is known only when an animal is found alive, then, compared with the case in which the covariate is always known, the four-state process loses

- (a) all the transitions from states 0 and 3,
- (b) all the $1 \rightarrow 3$ transitions and
- (c) all the $1 \rightarrow 0$ transitions except those starting at t_{k-1} .

The three-state and trinomial likelihoods, in contrast, lose only the transitions corresponding to (a). This difference between the three-state and four-state models results from the dependence of the states of the four-state process on future occurrences. This is reflected in, for example, the expressions $\chi_{i,r}$ in equation (12a) as compared with $\chi_{i,r,r+1}$ in equation (5a). This results in the four-state maximum likelihood estimator being biased when there are missing covariate values.

5. Simulation study

We now compare the performances of each of the methods that were discussed above, for simulated data under three scenarios: firstly, where there are no missing covariate data; secondly, where a covariate value is missing if and only if an animal is not captured; and, finally, where the covariate values are missing at random.

We simulate a study with $k = 11$ MRR occasions and 100 animals marked as newborns on each occasion except the last. We consider the model $\{\phi_1, \phi_a(V); p_t; \lambda_t\}$, which has two age categories for survival, with survival probabilities of animals in their first year of life (ϕ_1) being different from annual survival probabilities (ϕ_a) for older animals, and this 'adult' survival probability being regressed linearly, on the logistic scale, on an individual, time varying covariate V . The recapture probabilities p_t and recovery probabilities λ_t are both fully time dependent, with (potentially) different values on each occasion.

We performed 100 simulations, in each of which we generated data from the model, using the parameter values $\phi_1 = 0.5$, $\text{logit}(\phi_a) = 1.386 + V$, $p_t = 0.8$ and $\lambda_t = 0.4$, $t = 1, \dots, 10$ (although there are 11 MRR occasions, only initial marking occurs on the first occasion). We have used constant values for p_t and λ_t for ease of interpretation of the results. The covariate V is simulated from a standard normal distribution, independently for each animal on each occasion. The same values of the covariate were used in all simulations, except that in each simulation only 'known' covariate values were used. We then fitted the model by using each of the expressions (1), (6), (9) and (11) for the likelihood, as well as the binomial method of Catchpole, Morgan and Coulson (2004).

Table 1. Results from 100 simulations of the model $\{\phi_1, \phi_a(V); p_t; \lambda_t\}$, with no missing covariate values, using (a) either the three-state likelihood (6), or the four-state likelihood (11), or the complete-case likelihood (1) or the partial-case method of Section 4.1, which are all equivalent under this scenario (denoted 'standard' below), (b) the trinomial likelihood (9) and (c) the binomial method of Catchpole, Morgan and Coulson (2004)†

Parameter	True value	Results for the following methods:					
		(a) Standard		(b) Trinomial		(c) Binomial	
		Mean	Standard error	Mean	Standard error	Mean	Standard error
ϕ_1	0.500	0.498	0.017	0.496	0.030	0.502	0.023
ϕ_{aC}	1.386	1.378	0.086	1.379	0.190	1.406	0.139
ϕ_{aV}	1.000	1.012	0.097	1.016	0.127	1.011	0.121
p	0.800	0.802	0.042	0.803	0.062	0.802	0.051
λ	0.400	0.400	0.056	0.401	0.068	0.405	0.073

†Shown are the true values of the parameters, the means and standard errors of the first-year survival ϕ_1 , and the intercept ϕ_{aC} and slope ϕ_{aV} of the adult survival. For the recapture probabilities p_t and recovery probabilities λ_t we show the overall mean and the overall standard deviation. The adult survival parameters are shown on a logistic scale: other parameters are shown in the natural (probability) scale.

5.1. No missing values

The first case is when there are no missing covariate values. The three-state and four-state likelihoods (6) and (11), and the complete-case and partial-case approaches to the likelihood (1), are all equivalent in this case, since no covariate data are missing. Such a case would normally occur when there are only extrinsic (or no) covariates. However, for comparison with later results, we retain the intrinsic covariate V and assume that V is known, whether or not the animal is seen.

Table 1 shows the true values of the parameters and the means and standard errors of the 100 estimates. For p_t and λ_t Table 1 shows only the overall mean and standard deviation. Means and standard deviations for the ϕ_a intercept and slope are calculated on the logistic scale; for the other parameters we use the natural (probability) scale. It is clear from Table 1 that all methods produce approximately unbiased estimates, but that the estimates from the three-state, four-state, complete-case and partial-case methods are more precise than those from the trinomial likelihood (9). This is because the trinomial method uses only the transitions from state 1, whereas the other methods use all the data. The binomial method of Catchpole, Morgan and Coulson (2004) is intermediate in precision. Note that this is the first time that the binomial method has been evaluated in a simulation study, when the standard errors reflect the entire process of using estimates of p_t and λ_t obtained from fitting an MRR model by maximum likelihood. In practice, when the binomial method is used on real data, the standard errors of the ϕ -parameters will be underestimated, as no account is taken of the variation due to the estimation of p_t and λ_t .

5.2. Missing when not seen

Our second simulation uses the same model, but now we assume that the covariate V is known for a particular animal for the transition $t_j \rightarrow t_{j+1}$ only if the animal is seen alive at t_j ; otherwise

Table 2. Results from 100 simulations of the model $\{\phi_1, \phi_a(V); p_t; \lambda_t\}$, with the covariate V missing whenever an animal is not seen, using (a) the three-state likelihood (6) and the trinomial likelihood (9), which are equivalent under this scenario, (b) the complete-case method, (c) the partial-case method of Section 4.1, (d) the four-state likelihood (11) and (e) the binomial method of Section 4.2†

Parameter	True value	Results for the following methods:									
		(a) Three state (trinomial)		(b) Complete case		(c) Partial case		(d) Four state		(e) Binomial	
		Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error	Mean	Standard error
ϕ_1	0.500	0.498	0.029	0.477	0.023	0.699	0.019	0.660	0.018	0.502	0.023
ϕ_{aC}	1.386	1.396	0.168	1.422	0.176	2.240	0.135	2.068	0.140	1.393	0.157
ϕ_{aV}	1.000	1.010	0.136	1.031	0.167	1.009	0.140	1.006	0.143	1.006	0.129
p	0.800	0.804	0.062	0.982	0.058	0.813	0.062	0.969	0.093	0.802	0.051
λ	0.400	0.401	0.068	0.940	0.182	0.962	0.117	0.957	0.132	0.405	0.073

†Results are shown in the same way as in Table 1.

the covariate, and hence ϕ_a , is missing. Here the three-state and trinomial likelihoods are equivalent, since the trinomial method considers only transitions from state 1, and in the three-state process all transitions except those from state 1 are missing. The four-state process leaves out more data, as described in Section 4.4, and the complete-case likelihood leaves out more still, since it leaves an animal in the likelihood only if that animal is never missed on any recapture occasion before the last.

Table 2 shows the results of 100 simulations, in a similar way to Table 1. Table 2 shows that the three-state method works well, remaining unbiased, but with decreased precision compared with Table 1 when there were no missing covariate data. The binomial method performs well, with almost identical results to those in Table 1. But the complete-case, partial-case and four-state methods all produce badly biased estimates.

The complete-case, partial-case and four-state methods all delete the ‘trailing 0s’ parts of the data, i.e. those parts of case histories that end 0...0, except for the case of a single 0 at t_k . As a result, they all estimate $\lambda_1 = \dots = \lambda_{k-2} = 1$. The complete-case and four-state methods also do not handle intermediate 0s in the record, and as a result estimate $p_1 = \dots = p_{k-2} = 1$. The partial-case method does include the p_j -terms for intermediate 0s, since they do not depend on a covariate, and so its estimates of the p_j are much less biased.

The reason that the binomial method works so well in this case may be that it is a two-stage process: first the model $\{\phi_{1,t}, \phi_{a,t}; p_t; \lambda_t\}$ is fitted, to estimate p_t and λ_t , and then the binomial method is used to estimate the survival parameters. The first step in this process uses all the data: there are no covariates and so no missing values (and, as a result, the estimates of p_t and λ_t are identical to those which are obtained in the first simulation). The second step uses the reduced data set, where the covariate is known, but in this step there are many fewer parameters to estimate.

5.3. Missing at random

There are many different scenarios which produce different results from each of the methods for dealing with missing covariate values. For example values can be missing at random, owing to

Table 3. Results from 100 simulations of the model $\{\phi_1, \phi_a(V); p_t; \lambda_t\}$, with the covariate V missing at random, using (a) the three-state likelihood (6), (b) the trinomial likelihood (9) and (c) the binomial likelihood of Section 4.2†

Parameter	True value	Results for the following methods:					
		(a) Three state		(b) Trinomial		(c) Binomial	
		Mean	Standard error	Mean	Standard error	Mean	Standard error
ϕ_1	0.500	0.501	0.019	0.504	0.030	0.501	0.023
ϕ_{aC}	1.386	1.391	0.098	1.392	0.197	1.410	0.146
ϕ_{aV}	1.000	1.006	0.102	0.980	0.148	1.013	0.125
p	0.800	0.799	0.042	0.799	0.068	0.802	0.051
λ	0.400	0.402	0.063	0.402	0.079	0.405	0.073

†Results are shown in the same way as in Table 1.

lost records, or missing on one or more occasions through equipment malfunction, for example. In the third simulation we assume that 10% of the covariate values are missing at random but are otherwise known, whether or not an animal is observed. We retain the same individual covariate here, for comparison with the previous simulations rather than for biological realism. We consider here only those methods that produced reasonably unbiased estimates in the previous simulation.

Table 3 shows the results of 100 simulations, in the same way as Tables 1 and 2. The three-state, trinomial and binomial methods again produce approximately unbiased parameter estimates, with the three-state method having the best precision. The reason that the trinomial and three-state methods differ is that the three-state method includes transitions from state 0, when the covariate value is not missing, whereas the trinomial method uses only transitions from state 1.

6. Discussion

The essential difference between the complete-case and three-state likelihoods is that in the latter we construct the likelihood one step at a time, using transition probabilities. This contrasts with the normal procedure, in which the entire history of an animal is used to construct a multinomial likelihood. The three-state and four-state likelihoods differ in that, in the four-state method, we use hindsight from the end of the study to fill in information on 0s occurring during the study. The three-state method, in contrast, uses only the information that is available at the time that each observation is made.

However, this transition probability approach is specifically designed for cases where the parameters ϕ_t , p_t and λ_t depend on covariates that are measured at t . This covers many practically important cases in animal studies, e.g. where survival is modelled in terms of last year's weight. But there are cases that it does not cover, such as when the recapture probability depends on current breeding status. In such situations, where any of the parameters depend on covariates that are measured at $t + 1$, a different approach is needed. Our method depends on forming a conditional distribution, and in this distribution there are no missing data, when all covariates are known at t . This would no longer be the case if parameters depended on covariates that are measured at $t + 1$.

The binomial approach of Catchpole, Morgan and Coulson (2004) is attractive, as it is easily programmed as a generalized linear model. However, it has several disadvantages. It requires a separate MRR analysis to estimate the time varying values of p and λ , it makes no allowance in its estimated standard errors for the errors in estimating p and λ , and it is limiting in that it is not practicable to allow covariates that influence ϕ also to influence p and λ . In our simulations, the good performance of the binomial method may be due to the few (three) parameters that are involved in survival modelling in our simulations, compared with far more (20) in the recapture and recovery probabilities. In cases where there are many parameters in the survival modelling, we would expect the performance of the binomial method to decline relatively to that of the other methods.

It is impressive that in our simulations the three-state approach outperforms the binomial method overall, without having to import separate estimates of p_t and λ_t . The three-state likelihood is identical to the complete-case likelihood when there are no missing covariate values, and it performs better than the complete-case likelihood when there are missing values. It is a simple plug-in replacement for the traditional likelihood, it covers models with and without covariates and is easily coded so that it deals automatically with missing covariate data. For instance, it could easily be incorporated in standard computer packages such as MARK (White and Burnham, 1999) and M-SURGE (Choquet *et al.*, 2004), where it would provide an automatic way of dealing with missing values of individual covariates. Neither package can currently deal with time varying individual covariates with missing values. Our method is also easy to program in stand-alone code, and MATLAB code is available from

<http://www.blackwellpublishing.com/rss>

The new formulation of the likelihood applies directly to the formation of the posterior distribution if Bayesian analysis is used.

In our simulations, the parameter values and numbers of animals that are used are realistic for large mammal studies. However, the results are readily interpretable, and we expect that our conclusions will hold quite generally. Although we have found in the complete-case analyses that the bias in p_t and λ_t has not led to any large bias in the survival probabilities, we note that in our simulations ϕ_a was regressed only on an individual covariate. If it had been regressed on a time varying environmental covariate instead (or as well), we speculate that this would have led to biased estimators of survival.

Our simulations have used independent and identically distributed values for the covariate. Very similar results occurred in simulations using correlated values, where the covariate was simulated

- (a) from an auto-regressive moving average model in time (within each animal), with the same fixed mean for each animal, and
- (b) with independent and identically distributed values within each animal, but with the animal means being different (generated at random).

Also, all simulations have used values of p and λ which, although time varying, do not depend on an individual covariate. Simulations in which they depended on the same covariate as ϕ gave similar results to those shown.

In this paper we have analysed MRR data incorporating both recapture and recovery information. The approach can also be used if only recovery or recapture information is available, although we have not yet evaluated its efficacy in these cases. Of course we cannot expect good performance if there are individual time varying covariates and the only data come from dead

recoveries. Similarly, we have made no distinction between resighting and recapture; such a distinction is clearly possible, and our method can be extended to cover this case. If there are more than three states, as, for instance, with multisite MRR models, then we could construct different conditional models, all of which would be acceptable, and, subject to not being parameter redundant, all of which would result in consistent estimators. Some would be more efficient than others.

Diggle *et al.* (1994), page 210, ‘know of no well-developed methodology for dealing with informative, intermittent missing values in longitudinal data’. This paper represents a step in that direction.

We are engaged in further research on the three-state process, through its practical application, Bayesian implementation and comparison with imputation methods. Future work will explore applications to multistate MRR models, possible human applications and cases in which the covariate is measured with error.

Acknowledgements

The work of GT was supported by a Biotechnology and Biological Sciences Research Council research grant, and that of BJTM was supported by a Leverhulme Fellowship. We thank Wenyang Zhang for the proof that is shown in Appendix B, and the Joint Editor, an Associate Editor and two referees for their comments, which greatly improved the paper. We thank Geert Molenberghs for discussions concerning consistency and human applications.

Appendix A: Proof of the equivalence of the three-state, four-state and traditional likelihoods when there are no missing covariate data

Since each of the three-state, four-state and traditional likelihoods is written as a product over the animals, we need only to prove their equivalence for a single, arbitrary life history, say for the *i*th animal. We consider a subsequence of the life history data for an animal beginning on an occasion (possibly the initial tagging occasion) on which the animal was seen alive. There are three possible cases:

- (a) the next time that the animal is seen, it is alive,
- (b) the next time that the animal is seen, it is dead or
- (c) the animal is never seen again.

The proof is based on the fact that each of the likelihoods renews when the animal is seen alive, i.e. when the data entry is 1. By ‘renews’, we mean that the likelihood in each of these cases has no dependence on any previous entries in the life history. If we prove that the likelihoods are equal for each of these subsequences, then, since the three cases exhaust the possibilities, this is sufficient to complete the proof.

A.1. Case (a)

Suppose that animal *i* is seen on occasion *l*, where this may be the initial tagging occasion, and that the animal is subsequently seen alive, after *z* ($z \geq 0$) occasions on which it was not seen, i.e. we consider a subsequence of the life history of the form $(\dots, 1, 0, \dots, 0, 1, \dots)$, where there are *z* 0s between the two 1s.

First note that, if $z = 0$, then, from equations (5b) and (12b), the three-state and four-state likelihoods for this subsequence are simply $\pi_{i,l}(1, 1) = \rho_{i,l}(1, 1) = \phi_{i,l}p_{i,l}$, and this is just the traditional form of the likelihood for this subsequence. If $z > 0$, the three-state likelihood is

$$\pi_{i,l}(1, 0) \prod_{s=l+1}^{l+z-1} \pi_{i,s}(0, 0) \pi_{i,l+z}(0, 1),$$

where, in the case $z = 1$, the central product as taken as 1. From equations (4) and (5), the likelihood becomes

$$\chi_{i,l,l+1} \prod_{s=l+1}^{l+z-1} \frac{\chi_{i,l,s+1}}{\chi_{i,l,s}} \prod_{s=l+1}^{l+z-1} \phi_{i,s}(1-p_{i,s}) \frac{\phi_{i,l+z} p_{i,l+z}}{\chi_{i,l,l+z}} = \prod_{s=l+1}^{l+z-1} \phi_{i,s}(1-p_{i,s}) \phi_{i,l+z} p_{i,l+z},$$

which is just the traditional form of the likelihood for this subsequence.

In a similar manner, from equations (12) and (13), the four-state likelihood is

$$\begin{aligned} \rho_{i,l}(1, 3) \prod_{s=l+1}^{l+z-1} \rho_{i,s}(3, 3) \rho_{i,l+z}(3, 1) &= \phi_{i,l}(1-p_{i,l})(1-\chi_{i,l+1}) \prod_{s=l+1}^{l+z-1} \phi_{i,s}(1-p_{i,s}) \frac{1-\chi_{i,s+1}}{1-\chi_{i,s}} \frac{\phi_{i,l+z} p_{i,l+z}}{1-\chi_{i,l+z}} \\ &= \prod_{s=l+1}^{l+z-1} \phi_{i,s}(1-p_{i,s}) \phi_{i,l+z} p_{i,l+z}, \end{aligned}$$

as above. Case (b) is similar; the details are omitted.

A.2. Case (c)

Finally we consider the case in which the animal is never seen again, live or dead, after having been seen alive on occasion l , so that the subsequence of the life history is $(\dots, 1, 0, \dots, 0)$, where the 1 is followed by z 0s, $z \geq 0$, finishing on occasion k .

If $z = 0$ there is nothing to prove, as there is no contribution to the likelihood. If $z > 0$, then, using equations (4) and (5) as usual, the three-state likelihood is

$$\pi_{i,l}(1, 0) \prod_{s=l+1}^{k-1} \pi_{i,s}(0, 0) = \chi_{i,l,l+1} \prod_{s=l+1}^{k-1} \frac{\chi_{i,l,s+1}}{\chi_{i,l,s}} = \chi_{i,l,k} = \chi_{i,l},$$

which is just the traditional form of the likelihood for this subsequence.

Similarly, from equations (12) and (13), the four-state likelihood is

$$\rho_{i,l}(1, 0) \prod_{s=l+1}^{k-1} \rho_{i,s}(0, 0) = \chi_{i,l} \times 1.$$

Appendix B: Proof of consistency

The trinomial likelihood (10) can be written

$$L(\theta) = \prod_{i=1}^n \prod_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} \pi_{i,r}(1, h_{i,r+1}|\theta), \tag{14}$$

where θ is the vector of all unknown parameters. Let $\hat{\theta}$ be the maximizer of likelihood (14). It is easy to see that

$$\frac{1}{n} \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} \log\{\pi_{i,r}(1, h_{i,r+1}|\theta)\} \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} E[\log\{\pi_{i,r}(1, h_{i,r+1}|\theta)\} | h_{i,r} = 1]. \tag{15}$$

Let θ_0 be the true θ . For any $\theta \neq \theta_0$,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} E[\log\{\pi_{i,r}(1, h_{i,r+1}|\theta)\} | h_{i,r} = 1] &- \frac{1}{n} \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} E[\log\{\pi_{i,r}(1, h_{i,r+1}|\theta_0)\} | h_{i,r} = 1] \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} E \left[\log \left\{ \frac{\pi_{i,r}(1, h_{i,r+1}|\theta)}{\pi_{i,r}(1, h_{i,r+1}|\theta_0)} \right\} \middle| h_{i,r} = 1 \right] \\ &< \frac{1}{n} \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} \log \left[E \left\{ \frac{\pi_{i,r}(1, h_{i,r+1}|\theta)}{\pi_{i,r}(1, h_{i,r+1}|\theta_0)} \middle| h_{i,r} = 1 \right\} \right] = 0, \end{aligned}$$

because

$$E\left\{\frac{\pi_{i,r}(1, h_{i,r+1}|\theta)}{\pi_{i,r}(1, h_{i,r+1}|\theta_0)} \middle| h_{i,r} = 1\right\} = 1.$$

So, we have

$$\frac{1}{n} \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} E[\log\{\pi_{i,r}(1, h_{i,r+1}|\theta)\} | h_{i,r} = 1] < \frac{1}{n} \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} E[\log\{\pi_{i,r}(1, h_{i,r+1}|\theta_0)\} | h_{i,r} = 1].$$

This, together with equation (15), leads to

$$\lim_{n \rightarrow \infty} \left(\Pr \left[\sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} \log\{\pi_{i,r}(1, h_{i,r+1}|\theta_0)\} > \sum_{i=1}^n \sum_{\substack{r=c_i \\ h_{i,r}=1}}^{k-1} \log\{\pi_{i,r}(1, h_{i,r+1}|\theta)\} \right] \right) = 1.$$

So $\hat{\theta}$ is consistent.

References

- Abraham, W. T. and Russell, D. W. (2004) Missing data: a review of current methods and applications in epidemiological research. *Curr. Opin. Psychiatr.*, **17**, 315–321.
- Bonner, S. and Schwarz, C. (2006) An extension of the Cormack-Jolly-Seber model for continuous covariates with application to *Microtus pennsylvanicus*. *Biometrics*, **62**, 142–149.
- Catchpole, E. A., Fan, Y., Morgan, B. J. T., Clutton-Brock, T. H. and Coulson, T. (2004) Sexual dimorphism, survival and dispersal in red deer. *J. Agric. Biol. Environ. Statist.*, **9**, 1–26.
- Catchpole, E. A., Freeman, S. N., Morgan, B. J. T. and Harris, M. P. (1998) Integrated recovery/recapture data analysis. *Biometrics*, **54**, 33–46.
- Catchpole, E. A., Freeman, S. N., Morgan, B. J. T. and Nash, W. J. (2001) Abalone I: analysing mark-recapture-recovery data, incorporating growth and delayed recovery. *Biometrics*, **57**, 469–477.
- Catchpole, E. A., Morgan, B. J. T. and Coulson, T. (2004) Conditional methodology for individual case history data. *Appl. Statist.*, **53**, 123–131.
- Choquet, R., Reboulet, A.-M., Pradel, R., Gimenez, O. and Lebreton, J.-D. (2004) M-SURGE: new software specifically designed for multistate capture-recapture models. *Anim. Biodiversity Conservn.*, **27**, 207–215.
- Cox, D. R. and Hinkley, D. V. (1974) *Theoretical Statistics*. London: Chapman and Hall.
- Diggle, P., Farewell, D. and Henderson, R. (2007) Analysis of longitudinal data with drop-out: objectives, assumptions and a proposal. *Appl. Statist.*, **56**, 499–529.
- Diggle, P. and Kenward, M. G. (1994) Informative drop-out in longitudinal data analysis (with discussion). *Appl. Statist.*, **43**, 49–93.
- Diggle, P. J., Liang, K.-Y. and Zeger, S. L. (1994) *Analysis of Longitudinal Data*. Oxford: Oxford University Press.
- Dupuis, J. A. (1995) Bayesian estimation of movement and survival probabilities from capture-recapture data. *Biometrika*, **82**, 761–772.
- Gadbury, G. L., Coffey, C. S. and Allison, D. B. (2003) Modern statistical methods for handling missing repeated measurements in obesity trial data: beyond LOCF. *Obesity Rev.*, **4**, 175–184.
- Guo, X. and Carlin, B. P. (2004) Separate and joint modelling of longitudinal and event time data using standard computer packages. *Am. Statist.*, **58**, 16–24.
- Henderson, R., Diggle, P. and Dobson, A. (2000) Joint modelling of longitudinal measurements and event time data. *Biostatistics*, **1**, 465–480.
- Hoffman, A. and Skalski, J. R. (1995) Inferential properties of an individual-based survival model using release-recapture data: sample size, validity and power. *J. Appl. Statist.*, **22**, 579–595.
- Hogan, J. W., Roy, J. and Korkontzelou, C. (2004) Handling drop-out in longitudinal studies. *Statist. Med.*, **23**, 1455–1497.
- Illius, A. W., Albon, S. D., Pemberton, J. M., Gordon, I. J. and Clutton-Brock, T. H. (1995) Selection for foraging efficiency during a population crash in Soay sheep. *J. Anim. Ecol.*, **64**, 481–492.
- King, R. and Brooks, S. P. (2003) Closed-form likelihoods for Arnason-Schwarz models. *Biometrika*, **90**, 435–444.
- Little, R. J. A. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*, 2nd edn. Hoboken: Wiley.
- Manly, B. F. J., McDonald, T. L. and Amstrup, S. C. (2005) Introduction to the handbook. In *Handbook of Capture-Recapture Analysis* (eds S. C. Amstrup, T. L. McDonald and B. F. J. Manly). Princeton: Princeton University Press.
- Milner, J. M., Elston, D. A. and Albon, S. D. (1999) Estimating the contributions of population density and climatic fluctuations to interannual variation in survival of Soay sheep. *J. Anim. Ecol.*, **68**, 1235–1247.
- Molenberghs, G. and Verbeke, G. (2005) *Models for Discrete Longitudinal Data*. New York: Springer.

- Moorcroft, P. R., Albon, S. D., Pemberton, J. M., Stevenson, I. R. and Clutton-Brock, T. H. (1996) Density-dependent selection in a fluctuating ungulate population. *Proc. R. Soc. Lond. B*, **263**, 31–38.
- Pollock, K. H. (2002) The use of auxiliary variables in capture-recapture modelling: an overview. *J. Appl. Statist.*, **29**, 85–102.
- Prentice, R. L. (1982) Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, **69**, 331–342.
- Rubin, D. B. (1976) Inference and missing data. *Biometrika*, **63**, 581–592.
- Saha, C. and Jones, M. P. (2005) Asymptotic bias in the linear mixed effects model under non-ignorable missing data mechanisms. *J. R. Statist. Soc. B*, **67**, 167–182.
- Skalski, J. R., Hoffman, A. and Smith, S. G. (1993) Testing the significance of individual and cohort-level covariates in animal survival studies. In *Marked Individuals in the Study of Bird Populations* (eds J.-D. Lebreton and P. M. North), pp. 9–28. Basel: Birkhäuser.
- White, G. C. and Burnham, K. P. (1999) Program MARK: survival estimation from populations of marked animals. *Bird Study*, **46**, suppl., 120–139.
- Williams, B., Nichols, J. D. and Conroy, M. J. (2002) *Analysis and Management of Animal Populations*. San Diego: Academic Press.
- Wu, M. C. and Carroll, R. J. (1988) Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, **44**, 175–188.