

## Modeling Heterogeneity in Nest Survival Data

Ranjini Natarajan

Division of Biostatistics, Department of Statistics, University of Florida,  
1600 S.W. Archer Road, Room C2-22, Gainesville, Florida 32610, U.S.A.  
*email:* ranjini@stat.ufl.edu

and

Charles E. McCulloch

Biometrics Unit, Department of Statistical Science, Cornell University,  
434 Warren Hall, Ithaca, New York 14853, U.S.A.

**SUMMARY.** Current statistical methods for estimating nest survival rates assume that nests are identical in their propensity to succeed. However, there are several biological reasons to question this assumption. For example, experience of the nest builder, number of nest helpers, genetic fitness of individuals, and site effects may contribute to an inherent disparity between nests with respect to their daily mortality rates. Ignoring such heterogeneity can lead to incorrect survival estimates. Our results show that constant survival models can seriously underestimate overall survival in the presence of heterogeneity. This paper presents a flexible random-effects approach to model heterogeneous nest survival data. We illustrate our methods through data on redwing blackbirds.

**KEY WORDS:** Daily survival rate; Heterogeneity; Likelihood ratio test; Mayfield's method; Nest success; Numerical integration; Overdispersion; Random effects.

### 1. Introduction

Statistical models for analyzing nest data commonly assume a constant probability of survival across nests (Mayfield, 1961; Mayfield, 1975; Bart and Robson, 1982; Pollock and Cornelius, 1988; Bromaghin and McDonald, 1993). However, there are several biological reasons to believe that daily mortality rates differ among nests and not necessarily in relation to any variable the investigator is able to measure (Johnson, 1979; Klett and Johnson, 1982). For instance, the ability of nests to succeed may vary due to nonidentifiable sources of heterogeneity such as genetic fitness of individuals or experience of the nest builder. Identifiable sources of variation can arise from differences in nest site, geographical location, etc.

Current approaches to account for variability in nest data are limited to factors that impact survival for all nests in a similar fashion. Pollock and Cornelius (1988) develop models where daily survival varies with the age of a nest but is assumed to be constant across nests of the same age. Three limitations of this method are that (i) it cannot easily incorporate multiple factors (since sample sizes would become small with increasing stratification); (ii) it cannot include nest-level covariates; and (iii) the number of parameters can increase drastically with long nesting periods (as in waterfowl). A different approach to incorporate heterogeneity in survival has been proposed by Burnham and Rexstad (1993) using ul-

trastructure models superimposed on traditional models for band-recovery data.

This paper describes a flexible random-effects modeling approach to analyzing nest survival data in the presence of tangible and intangible variation between nests. Our methods are applicable for estimating survival in any stage of the nest history, i.e., incubation or nestling. We first present a simple formulation to model "pure" heterogeneity, or nonidentifiable sources of variation. This model can serve as a useful and quick diagnostic tool to evaluate the adequacy of a constant survival assumption. Next, we present a general random-effects model to estimate overall survival in the presence of covariates and multiple sources of heterogeneity. All notation for these models are initially presented assuming nests are found at the beginning of the stage. Extensions to the more typical setting of encounter sampling in wildlife studies will also be discussed using the method of Bromaghin and McDonald (1993). Two advantages of our random-effects approach are that it

- (1) allows calculation of predicted probabilities of survival for individual nests and
- (2) allows accommodation of a rich class of nest-level and time-varying covariate patterns, which facilitates a natural building of models for comparison (e.g., comparing average survival for nests in two different habitats).

This paper is organized as follows. In Section 2, we describe a beta-binomial-type model to capture pure heterogeneity in survival rates. A likelihood ratio statistic to test for inhomogeneities between nests is presented. In Section 3, we describe a random-effects model to account for multiple sources of heterogeneity and arbitrary structure of nest-level covariates. These methods are illustrated through data on redwing blackbirds. Our results suggest that constant survival models can seriously underestimate overall survival in the presence of heterogeneity. Extensions to encounter sampling are discussed in Section 4.

## 2. A Simple Model for Pure Heterogeneity

Ornithologists visit nests periodically and monitor their survival status. Once found, nests are visited either until they fail or succeed (i.e., hatch or fledge, depending on the stage under study). The observed data for nest  $i$  is the pair  $\{y_i, \ell_i\}$ , where  $y_i$  is a binary indicator of the ultimate fate of the nest, coded as one for success and zero for failure, and  $\ell_i$  is the number of time units the nest is under observation and active. For the moment, we focus on the situation where all nests are found immediately after initiation, i.e.,  $\ell$  denotes the length of life of a nest. Thus, nests that succeed are observed for  $\ell_i$  time units, while those that fail are observed for  $\ell_i + 1$  time units. (Throughout this paper, we assume the absence of confounding temporal factors, which allows us to realign the cohort so all have the same time zero of initiation.) Then the contribution of nest  $i$  to the observed-data likelihood is

$$f(y_i, \ell_i | p_i) = \begin{cases} p_i^{\ell_i} & \text{if } y_i = 1, \\ p_i^{\ell_i}(1 - p_i) & \text{if } y_i = 0, \end{cases} \quad (1)$$

where  $f(\cdot)$  denotes generic probability density or mass functions and  $p_i$  is the nest-specific survival probability for a single time unit. The above formulation can be recognized as Mayfield's (1961) natural density, with the constant probability of success  $p$  replaced by nest-level survival rates  $p_i$ . The specification in (1) is completed by postulating a distribution for the  $p_i$ . A flexible and convenient assumption for the  $p_i$  is the beta distribution,

$$f(p_i | \alpha, \beta) = \frac{1}{B(\alpha, \beta)} p_i^{\alpha-1} (1 - p_i)^{\beta-1}, \quad \alpha > 0, \beta > 0, \quad (2)$$

with  $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha + \beta)$ , mean  $\mu_p = \alpha/(\alpha + \beta)$ , and variance  $v_p = \mu_p(1 - \mu_p)/(\alpha + \beta + 1)$  (Griffiths, 1973; Williams, 1975). An important modeling consequence of the specifications in (1) and (2) is the (nonnegative) correlation induced between the survival status of a nest at the repeated visits. Kahn and Raftery (1996) describe a similar formulation for studying hospital variation in the discharge of hip-fracture patients to skilled nursing facilities.

The estimation problem lies in calculating the probability of surviving a period of  $J$  (say) time units, namely,  $\mathcal{P} = E[p_i^J]$ , where the expectation is over the survival distribution  $f(p_i | \alpha, \beta)$ . Upon noting that the marginal probability of surviving time unit  $j$  ( $j \geq 1$ ) given existence at the start of  $j$  is  $(\alpha + j - 1)/(\alpha + \beta + j - 1)$ , we have

$$\mathcal{P} = \prod_{j=1}^J \frac{(\alpha + j - 1)}{(\alpha + \beta + j - 1)}. \quad (3)$$

An estimator of survival and its asymptotic variance (obtained by the delta method) under the heterogeneous model are

$$\hat{\mathcal{P}} = \prod_{j=1}^J \frac{(\hat{\alpha} + j - 1)}{(\hat{\alpha} + \hat{\beta} + j - 1)},$$

$$v(\hat{\mathcal{P}}) = \hat{\mathbf{d}}^T(\hat{\alpha}, \hat{\beta}) \hat{\mathbf{V}}(\hat{\alpha}, \hat{\beta}) \hat{\mathbf{d}}(\hat{\alpha}, \hat{\beta}),$$

where  $\hat{\alpha}, \hat{\beta}$  are the maximum likelihood estimates of  $\alpha$  and  $\beta$ ,  $\hat{\mathbf{V}}$  is the associated variance matrix, and  $\hat{\mathbf{d}}$  the derivative of (3) with respect to  $\alpha$  and  $\beta$  evaluated at the maximum likelihood estimates. Expressions for  $\hat{\mathbf{d}}$  and  $\hat{\mathbf{V}}$  are presented in the Appendix.

In some instances, the predicted probabilities of survival for individual nests  $p_i^* = E[p_i | y_i, \ell_i, \alpha, \beta]$  may also be of interest, and these can be straightforwardly estimated by

$$\hat{p}_i^* = \frac{B(\hat{\alpha} + \ell_i + 1, \hat{\beta} - y_i + 1)}{B(\hat{\alpha} + \ell_i, \hat{\beta} - y_i + 1)}$$

$$= \frac{(\hat{\alpha} + \ell_i)}{(\hat{\alpha} + \ell_i + \hat{\beta} - y_i + 1)}.$$

### 2.1 Estimation of $\alpha$ and $\beta$

Estimates of the shape and scale parameters of the survival probability distribution are obtained by maximizing the logarithm of the observed-data likelihood derived from (1) and (2). More specifically, we maximize the logarithm of

$$L(\alpha, \beta | \mathbf{y}, \ell) = \prod_{i=1}^n f(y_i, \ell_i | \alpha, \beta)$$

$$= \prod_{i=1}^n \int_0^1 f(y_i, \ell_i | p_i) f(p_i | \alpha, \beta) dp_i$$

$$= \prod_{i=1}^n \left( \frac{B(\alpha + \ell_i, \beta)}{B(\alpha, \beta)} \right)^{y_i} \left( \frac{B(\alpha + \ell_i, \beta + 1)}{B(\alpha, \beta)} \right)^{1-y_i} \quad (4)$$

where  $n$  is the number of nests in the sample,  $\mathbf{y} = \{y_1, \dots, y_n\}^T$  and  $\ell = \{\ell_1, \dots, \ell_n\}^T$ . Numerical algorithms such as Newton-Raphson may be used to maximize (4) and the asymptotic variance matrix obtained by the inverse of the negated observed information.

It is important to note that the probability of survival obtained from the heterogeneity model is larger than that under the independence model, i.e.,  $(\alpha/(\alpha + \beta))^J$ , since for each  $j (\geq 1)$  we have  $(\alpha + j)/(\alpha + \beta + j) > \alpha/(\alpha + \beta)$ . Thus, we would typically expect constant survival models (which assume independence of survival status at repeated visits) to underestimate overall survival in the presence of pure heterogeneity. We investigated the impact of the average daily survival rate  $\mu_p$ , degree of heterogeneity  $v_p$ , and the sample size  $n$  on the magnitude of bias through a series of simulation studies. Our findings suggest that the bias is exacerbated in low surviving populations with high degrees of heterogeneity; further, it does not diminish very quickly with increasing sample size. Analysis of nest data must therefore carefully explore the possibility of inhomogeneities between nests before fitting constant survival models. In the next

section, we develop a likelihood ratio statistic to test for the presence of pure heterogeneity.

### 2.2 Likelihood Ratio Test for Heterogeneity

In order to formulate a hypothesis test for the presence of heterogeneity, we consider the following reparameterization of the model in (4):  $\mu_p = \alpha/(\alpha + \beta)$  and  $\theta = 1/(\alpha + \beta + 1)$ . Then the variance of the survival probabilities  $p_i$  can be expressed as  $v_p = \theta\mu_p(1 - \mu_p)$ . The null hypothesis of homogeneity is  $H_0: \theta = 0$  against the alternative  $\theta > 0$ . The likelihood function in the reparameterized scale is

$$L^*(\mu_p, \theta, | \underline{y}, \underline{\ell}) = L(\mu_p(1/\theta - 1), (1 - \mu_p)(1/\theta - 1) | \underline{y}, \underline{\ell})$$

and the likelihood ratio test statistic given by

$$\Lambda = \frac{L^*(\hat{\mu}_{op}, 0 | \underline{y}, \underline{\ell})}{L^*(\hat{\mu}_p, \hat{\theta} | \underline{y}, \underline{\ell})},$$

where  $\hat{\mu}_{op}$  is the maximum likelihood estimator of  $\mu_p$  under the null and  $\hat{\mu}_p$  and  $\hat{\theta}$  are the estimators under the heterogeneous model. Self and Liang (1987) prove that, under the null hypothesis of homogeneity,  $-2 \ln \Lambda$  is asymptotically distributed as a 50:50 mixture of a chi-square distribution with 1 d.f. ( $\chi_1^2$ ) and the constant zero. Thus, the  $p$ -value corresponding to a test based on a  $\chi_1^2$  is halved to obtain a  $p$ -value for this one-sided test on  $\theta$ . The null hypothesis of homogeneity is rejected at the 5% level of significance if the resulting  $p$ -value is smaller than 0.05.

We illustrate our method through an example. All calculations were performed in the matrix language GAUSS (Aptech Systems, 1994) using programs developed by the first author.

### 2.3 Hundred Acre Cove Redwings

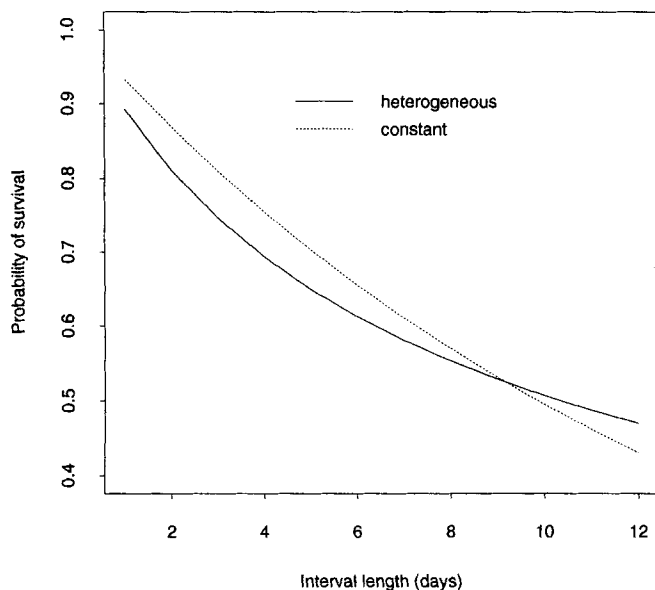
We analyze nest survival data collected by Steven E. Reinert from a salt marsh population of redwing blackbirds in Barrington, Rhode Island, between 1982 and 1985 (Reinert, Golet, and DeRagon, 1981). We focus on the nestling stage and only use the 53 nests for which the hatching date was observed. A nest success was recorded if at least one young fledged. The data are summarized in Table 1.

There does not appear to be much support for a constant survival assumption since the proportion of survivors across the observation period does not display a decreasing exponential trend. We fit a constant survival model (see Bart and Robson, 1982, equation (5)) and heterogeneous model (equation (4)). Estimated daily survival from the constant model is 0.93, while that from the heterogeneous model is 0.89. Maximum likelihood estimates of  $\alpha$  and  $\beta$  are 5 and 0.6, respectively. Estimated survival at 12 days is 46.9% (standard error of 0.068) for the heterogeneous model and 43.0% (standard error of 0.069) for the constant survival model. Figure 1 displays the probability of survival for varying interval lengths using both models. The survival curve from the heterogeneous model is much flatter, especially for the larger interval lengths, compared to the negative exponential curve of the constant survival model. The heterogeneous model yields a log-likelihood of  $-97.496$  with one additional parameter, compared with  $-98.709$  from the constant survival model. The likelihood ratio statistic to test the null hypothesis of homogeneity is 2.426, which corresponds to a  $p$ -value of 0.050, as explained in Section 2.2. Thus, there appears to be

**Table 1**  
*Survival data for nestling redwing blackbirds organized by the length of the observation period (day).*

Year	Day	Number of		
		Nests	Deaths	Nests fledged
1982	1	12	1	0
	2	11	0	0
	3	11	1	0
	4	10	1	0
	5	9	0	0
	6	9	1	0
	7	8	0	0
	8	8	1	0
	9	7	0	0
	10	7	0	4
	11	3	0	2
	12	1	1	0
Total		—	6	6
1983	1	12	3	0
	2	9	1	0
	3	8	1	0
	4	7	0	0
	5	7	0	0
	6	7	0	0
	7	7	1	0
	8	6	0	0
	9	6	0	0
	10	6	0	3
	11	3	1	2
Total		—	7	5
1984	1	13	1	0
	2	12	1	0
	3	11	0	0
	4	11	0	0
	5	11	0	0
	6	11	0	0
	7	11	0	0
	8	11	0	0
	9	11	0	0
	10	11	0	3
	11	8	0	6
	12	2	0	2
Total		—	2	11
1985	1	16	2	0
	2	14	2	0
	3	12	2	0
	4	10	0	0
	5	10	2	0
	6	8	1	0
	7	7	0	0
	8	7	0	0
	9	7	1	0
	10	6	2	2
	11	2	0	2
Total		—	12	4

evidence of heterogeneity among the nests. It is conceivable that some of this heterogeneity may be explained by adjusting for potentially important covariates; Section 3.2 investigates exactly such an analysis.



**Figure 1.** Estimated probability of nest survival (nestling stage) for redwing blackbird nests using the heterogeneous and constant survival model.

A primary advantage of the beta framework described in Section 2 is that variability between nests is modeled directly through the success probabilities  $p_i$ . However, it suffers from the limitation that it does not explain sources of heterogeneity. In Section 3, we formulate a logistic normal model to estimate nest survival in this setting.

### 3. Logistic Normal Model for Nest Survival

In this section, we describe a very general class of random-effect models to estimate nest success in the presence of one or more sources of heterogeneity. Conditional on the survival probabilities  $p_i$ , the contribution of nest  $i$  to the observed likelihood is the same as in (1). However, the nest-specific probabilities  $p_i$  are modeled as

$$\text{logit}\{p_i\} = \mathbf{x}_i^T \boldsymbol{\gamma} + v_i, \quad (5)$$

where  $\mathbf{x}_i$  is a  $p \times 1$  vector of covariates for nest  $i$ ,  $\boldsymbol{\gamma}$  is a  $p \times 1$  vector of unknown parameters, and  $v_i$  is a nest-specific intercept. (Note that a slight change in the formulation of (5) can incorporate time-varying covariates as described in equation (9).) The parameter  $\boldsymbol{\gamma}$  captures the effect of the covariates on survival for the average nest in the population, while  $v_i$  modifies the average response to make it specific to nest  $i$ . The specification in (5) is completed by postulating

$$v_i = \mathbf{z}_i^T \mathbf{b} + u_i, \quad (6)$$

where  $\mathbf{z}_i$  is a  $q \times 1$  design vector for the random effects  $\mathbf{b}$  and  $u_i$  is random error. The random effects  $\mathbf{b}$  can denote variations due to site, geographical location, year of data collection, or other tangible sources of heterogeneity. (It is prudent to entertain such models only when there are a reasonable number of levels of the source of heterogeneity. Opinions vary, but generally five or more levels are considered sufficient to investigate such models.) Typical distributional assumptions are  $\mathbf{b} \sim N_q(0, \mathbf{D})$  independent of  $u_i \sim N(0, \sigma^2)$ . Thus, equation (6) decomposes the variation between nests

into pure heterogeneity (i.e.,  $\sigma^2$ ) and that explained by differences in identifiable factors such as site (i.e.,  $\mathbf{D}$ ). This decomposition induces a correlation between survival status for a nest on repeated occasions as before but, in addition, it also induces correlation between nests that share a random effect. Models such as these have been studied extensively for analyzing spatially correlated data and are encompassed within the realm of spatial hierarchical models (see Ghosh et al., 1998).

Often, in practice, ornithologists may be interested in estimating survival in the presence of a single explainable source of heterogeneity, say  $k$  plots. Then equation (6) resembles the formulation for an analysis of variance, namely,

$$v_{ij} = b_i + u_{ij}, \quad i = 1, \dots, k, \quad j = 1, \dots, r_i, \quad (7)$$

where  $i$  indexes plots,  $j$  indexes nest within plots,  $b_i$  is a plot-specific random intercept distributed as  $b_i \sim N(0, \theta)$ , and  $u_{ij} \sim N(0, \sigma^2)$ . Equation (7) arises from equation (6) by defining  $\mathbf{Z} = \mathbf{I} \otimes \mathbf{1}$ , where  $\mathbf{Z}$  is the  $n \times q$  matrix with rows  $\mathbf{z}_i$ ,  $\mathbf{I}$  the identity matrix,  $\mathbf{1}$  a vector of ones, and  $\otimes$  the direct product operator. A variety of other heterogeneity patterns may be modeled by appropriate choices of  $\mathbf{Z}$ .

Conditional on the random effects  $\mathbf{b}$ , nests are statistically independent with likelihood given by (1). Unconditionally, the likelihood is given by

$$\begin{aligned} L(\boldsymbol{\gamma}, \sigma^2, \mathbf{D} \mid \mathbf{y}, \ell) \\ \propto \int \left\{ \prod_{i=1}^n \int f(y_i, \ell_i \mid u_i, \mathbf{b}) \frac{\exp\left(-\frac{u_i^2}{2\sigma^2}\right)}{\sigma} du_i \right\} \\ \times \frac{\exp\left(-\frac{1}{2} \mathbf{b}^T \mathbf{D}^{-1} \mathbf{b}\right)}{|\mathbf{D}|^{1/2}} d\mathbf{b}. \end{aligned} \quad (8)$$

Closed-form expressions do not exist for the likelihood in (8). However, for simple random-effect structures (i.e.,  $\mathbf{Z}$ ), numerical integration methods may be used to provide an estimate. (Hedeker and Gibbons [1994] have developed several programs to estimate integrals of the form in (8).) For complicated models, the Monte Carlo method may be used to provide a simulation-based estimate of the likelihood function. (McCulloch [1997] has studied the quality of various Monte Carlo approximants for related problems.) This estimated likelihood can then be maximized to obtain maximum likelihood estimates of the regression parameters  $\boldsymbol{\gamma}$  and the components of variance  $\sigma^2$  and  $\mathbf{D}$ .

As in Section 2, the inferential goal lies in estimating the unconditional (integrated over the random effects) probability of surviving  $J$  days, which in this context is calculated for specific covariate strata  $\mathbf{x}$ , i.e.,

$$\begin{aligned} \mathcal{P}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{q+1}{2}}} \int \left\{ \int p_i^J \frac{\exp\left(-\frac{u_i^2}{2\sigma^2}\right)}{\sigma} du_i \right\} \\ \times \frac{\exp\left(-\frac{1}{2} \mathbf{b}^T \mathbf{D}^{-1} \mathbf{b}\right)}{|\mathbf{D}|^{1/2}} d\mathbf{b}, \end{aligned}$$

where  $p_i$  is given by equations (5) and (6). The above expression involves calculations similar to those required for likelihood evaluation. Asymptotic standard errors may be obtained by the delta method.

Often, interest may also focus on predicted probabilities of survival for individual nests, which is given by  $E[(1 + \exp(-\mathbf{x}_i^T \underline{\gamma} - \mathbf{z}_i^T \underline{\mathbf{b}} - u_i))^{-J} \mid \mathbf{y}, \underline{\ell}]$ , with the expectation over the conditional distribution  $f(\underline{\mathbf{b}}, \underline{\ell} \mid \mathbf{y}, \underline{\ell})$ . Predicted values of individual random effects may also be calculated as  $\hat{\mathbf{b}} = E[\underline{\mathbf{b}} \mid \mathbf{y}, \underline{\ell}]$  and  $\hat{u}_i = E[u_i \mid \mathbf{y}, \underline{\ell}]$ .

In the next section, we will describe some of the calculations for a single source of explainable heterogeneity, i.e., for  $\mathbf{Z} = \mathbf{1} \otimes \mathbf{1}$ .

### 3.1 Single Source of Explainable Heterogeneity

For a design with  $k$  levels of a source of heterogeneity and  $r_i$  nests per level, the likelihood in (8) reduces to

$$L(\underline{\gamma}, \theta, \sigma^2 \mid \mathbf{y}, \underline{\ell}) \propto \prod_{i=1}^k \int \left\{ \prod_{j=1}^{r_i} \int f(y_{ij}, \ell_{ij} \mid u_{ij}, b_i) \frac{\exp\left(-\frac{u_{ij}^2}{2\sigma^2}\right)}{\sigma} du_{ij} \right\} \times \frac{\exp\left(-\frac{1}{2\theta} b_i^2\right)}{\sqrt{\theta}} db_i.$$

An estimate of the above likelihood may be obtained very accurately using quadrature methods (Abramowitz and Stegun, 1964), even for  $k$  as large as 100, and is given by the expression

$$\exp \left( \sum_{i=1}^k \ln \left\{ \sum_h w_h \exp \left[ \sum_{j=1}^{r_i} \ln \left\{ \sum_g w_g p_{hg}^{\ell_{ij}} \times (1 - p_{hg})^{1-y_{ij}} \right\} \right] \right\} \right),$$

where the probability

$$p_{hg} = \left( 1 + \exp \left( -\mathbf{x}_{ij}^T \underline{\gamma} - \sqrt{\theta} a_h - \sqrt{\sigma^2} a_g \right) \right)^{-1},$$

and  $a, w$  are the abscissae and weights associated with the Gauss-Hermite quadrature. (Note that we calculate logarithms and then exponentiate the resulting expressions in order to avoid underflow errors.) Overall unconditional survival for a particular covariate strata  $\underline{\mathbf{x}}$  is

$$\mathcal{P}(\underline{\mathbf{x}}) \approx \frac{1}{(2\pi)^{\frac{k+1}{2}}} \sum_h w_h \sum_g w_g p_{hg}^J.$$

The null hypothesis of homogeneity for the oneway random-effects model is  $H_0: \theta = 0$  versus the alternative  $\theta > 0$ . As in Section 2.2, a likelihood ratio statistic to perform this test is

given by

$$\Lambda = \frac{L(\hat{\gamma}_o, \hat{\sigma}_o^2, 0 \mid \mathbf{y}, \underline{\ell})}{L(\hat{\gamma}, \hat{\sigma}^2, \hat{\theta} \mid \mathbf{y}, \underline{\ell})},$$

where  $\hat{\gamma}_o, \hat{\sigma}_o^2$  are the maximum likelihood estimates under the null and  $\hat{\gamma}, \hat{\sigma}^2$ , and  $\hat{\theta}$  are the unrestricted estimates. The null hypothesis of homogeneity is rejected for large positive values of  $-2 \ln \Lambda$ , as compared to a 50:50 mixture of a  $\chi_1^2$  and the constant zero.

### 3.2 Redwing Blackbirds Revisited

In order to illustrate the use of logit-normal models, we explore an analysis of the redwing data to (i) adjust for potential effects of age on daily survival and (ii) account for suspected variations across the four data collection years. More specifically, in year  $i$  ( $i = 1, \dots, 4$ ) for nest  $j$  at age  $t$ , we postulate the following model:

$$\text{logit} \{p_{ijt}\} = \gamma_0 + \gamma_1 I(t \leq 2) + \gamma_2 I(3 \leq t \leq 6) + \gamma_3 I(t \geq 7) + b_i + u_{ij}, \quad (9)$$

where  $I(\cdot)$  is the indicator function,  $b_i$  is a year-specific random intercept, and  $u_{ij}$  is a nest-within-year specific intercept. The above formulation allows daily survival to vary as a function of nest age, which is grouped into three intervals. (We explored other choices for the age intervals, including finer and coarser divisions, but did not notice any substantial differences in our results.) Table 2 compares estimated probability of 12-day survival rates and their standard errors for five models, including a constant survival model, the Pollock-Cornelius full model (using 12 age-specific failure probabilities, one for each day in the nestling period), an independence logit model with age effects (i.e., equation (9) without the random effects  $b_i$  and  $u_{ij}$ ), a logit-normal model with age effects and nest intercepts alone (i.e., equation (9) without the year-specific intercepts  $b_i$ ), and the full logit-normal model in equation (9). By incrementally adding the nest and year intercepts, we hope to isolate their individual contributions and develop a parsimonious model to best explain the variations in these data. We used the estimated likelihood and survival estimates to guide us in our choice of a model.

Estimated 12-day survival from the constant model is 43% and 45% from the independence logit model compared with 50% from both the random-effect models and the Pollock-Cornelius model. It is important to note that the survival

**Table 2**  
Comparison of estimated 12-day survivals and their standard errors for the redwing blackbird data for five models

Model	No. of parameters	Survival	Standard error	Log-likelihood
Constant survival model	1	0.430	0.069	-98.709
Logit model with age effects	3	0.456	0.072	-97.922
Logit-normal model with age and nest effects	4	0.504	0.082	-96.225
Logit-normal model with age, nest, and year effects	5	0.490	0.095	-95.760
Pollock-Cornelius model	12	0.490	0.070	-95.798

estimates from the random-effect models are interpreted as average survival for an entire population of redwing blackbirds, while those from the constant, independence logit, and Pollock-Cornelius models pertain to the specific nests under observation. Addition of nest-specific intercepts to the independence logit model (Model III in Table 2) results in a significant increase in the likelihood ( $p = 0.031$ ). This suggests that discrepancies in survival propensity remain over and above age influences, which, if not accounted for, can lead to models that underestimate overall survival. Inclusion of year-specific intercepts in the model adjusted for age and nest-level variation (Model IV in Table 2) does not provide a significant improvement in fit ( $p = 0.16$ ), suggesting the absence of correlation in survival within years. (This may be a manifestation of the small number of years under study.) It is thus reasonable to pool the data over years and consider a model with age and nest effects alone. A goodness of fit test of this reduced model with the saturated Pollock-Cornelius model shows that the latter (despite having eight additional parameters and making no assumptions about the functional form of survival probabilities) does not provide a significantly better fit ( $p = 0.99$ ). Thus, by carefully accounting for inhomogeneities between nests, we have developed a parsimonious model that can capture the effect of covariates as well as correlations induced at various levels. Average daily survival estimates (and standard errors) for each of the three age intervals from the logit-normal model with age and nest effects are 0.880 (0.035), 0.772 (0.073), and 0.724 (0.067).

#### 4. Extensions to Encounter Sampling

Typically, in practice, nests are found at various stages of their development. Thus, the observed data is the pair  $\{y_i, t_i\}$ , where  $y_i$  is as defined in Section 2 and  $t_i$  is the number of time units a nest is under observation and active. This observed interval is often smaller than the true length of life  $\ell_i$ . However, assuming that the time required for a nest to succeed is a known constant  $\mathcal{L}$ , these observed data contain information on  $\ell_i$ . Thus, e.g.,  $\ell_i = \mathcal{L}$  for nests that succeed and  $t_i \leq \ell_i \leq (\mathcal{L} - 1)$  for nests that fail. Ignoring this additional information and computing survival estimates based only on the observation time can result in conservative estimates of nesting success. Modeling extensions to encompass such encounter sampling is straightforward both for the pure heterogeneity model and the logistic-normal models.

##### 4.1 Pure Heterogeneity Model for Encounter Sampling

Conditional on the nest-specific probabilities  $p_i$ , the contribution of nest  $i$  to the observed likelihood is

$$f(y_i, t_i | p_i) = \begin{cases} f(t_i | \ell_i = \mathcal{L}, \varphi) p_i^{\mathcal{L}} & \text{if } y_i = 1, \\ \sum_{\ell=t_i}^{\mathcal{L}-1} f(t_i | \ell_i = \ell, \varphi) p_i^{\ell} (1 - p_i) & \text{if } y_i = 0. \end{cases} \quad (10)$$

The distribution  $f(t_i | \ell_i, \varphi)$  in equation (10) is determined by the search strategy used to sample nests. The sampling schemes commonly used in practice are discussed by Bromaghin and McDonald (1993); for illustration purposes, we consider the systematic sampling scheme, i.e.,

$$f(t_i | \ell_i, \varphi) = \frac{\varphi(1 - \varphi)^{\ell_i - t_i}}{(1 - (1 - \varphi)^{\ell_i})},$$

where  $\varphi$  is the daily probability of detection, which is assumed constant across nests. Bromaghin and McDonald (1993) also

discuss the issue of weighting the joint distribution of survival and lifetimes  $f(y_i, \ell_i)$  to account for the probability sampling inherent in nest data; these ideas can be applied in our context as well. It is easy to see that the unconditional observed-data likelihood is  $(\prod_{i=1}^n (f(t_i | \ell_i = \mathcal{L}, \varphi) B(\alpha + \mathcal{L}, \beta))^{\mathcal{L}} (\sum_{\ell=t_i}^{\mathcal{L}-1} f(t_i | \ell_i = \ell, \varphi) B(\alpha + \ell, \beta + 1))^{1 - \mathcal{L}}) / B(\alpha, \beta)$ . Estimation in this setting can proceed as before by numerically maximizing the likelihood function to obtain maximum likelihood estimates of the regression parameters, components of variance, and detection parameters.

##### 4.2 Logistic-Normal Model for Encounter Sampling

Conditional on the random effects  $\mathbf{b}$  and pure heterogeneity  $u_i$ , the contribution of nest  $i$  to the observed likelihood is as given in equation (10) with  $p_i$  of the form given by (5) and (6). The unconditional observed-data likelihood  $L(\gamma, \varphi, \mathbf{D}, \sigma^2 | \mathbf{y}, \mathbf{t})$  is proportional to

$$\int \left\{ \prod_{i=1}^n \int f(y_i, t_i | \mathbf{b}, u_i) \frac{\exp(-\frac{1}{2\sigma^2} u_i^2)}{\sigma} du_i \right\} \times \frac{\exp(-\frac{1}{2} \mathbf{b}^T \mathbf{D}^{-1} \mathbf{b})}{|\mathbf{D}|^{1/2}} d\mathbf{b},$$

and can be estimated in a similar fashion as the logistic-normal models.

#### 5. Conclusion

This paper presents a flexible random-effects framework to accommodate heterogeneity in nest survival data. We recommend that analysis of nest data begin with an examination of models that account for pure heterogeneity. If data has been collected from several different locations (or multiple years), models that induce correlation between nests from the same location (or year) should also be considered. If there appears to be insufficient evidence of disparities based on these analyses, only then should one resort to constant survival models.

#### ACKNOWLEDGEMENTS

We thank Steven E. Reinert, Medical Computing, LifeSpan, Rhode Island, for providing the redwing data. We also thank an associate editor and anonymous referee for several comments that greatly improved the manuscript. The work of the first author was partially supported by National Science Foundation grant DMS-9709091 and the second author was supported by National Science Foundation grant DMS-9625476.

#### RÉSUMÉ

Les méthodes statistiques habituellement utilisées pour estimer la survie des nichées sont basées sur l'hypothèse que tous les nids sont équivalents en terme de capacité de réussite. Cependant il y a de multiples raisons biologiques pour remettre en question cette hypothèse. Par exemple, l'expérience du constructeur du nid, le nombre de ses aides, les aptitudes génétiques individuelles, les caractéristiques du site peuvent contribuer à une hétérogénéité naturelle des taux quotidiens de mortalité entre les nichées (Klett et Johnson, 1982, *The Auk* **99**, 77-87). Ignorer cette hétérogénéité peut conduire à des estimations incorrectes des taux de survie.

Nos résultats montrent que les modèles de survie constante peuvent sous-estimer gravement la survie globale lorsqu'une telle hétérogénéité existe. Ce papier présente une approche flexible avec effets aléatoires pour modéliser des données de survie des nids en présence d'hétérogénéité. Nous illustrons ces méthodes par des données portant sur la variété des merles Redwing.

## REFERENCES

- Abramowitz, M. and Stegun, I. (1964). *Handbook of Mathematical Functions*. Washington, D.C.: U.S. Government Printing Office.
- Aptech Systems, Inc. (1994). Maple Valley, Washington: Natarajan.
- Bart, J. and Robson, D. S. (1982). Estimating survivorship when the subjects are visited periodically. *Ecology* **63**, 1078–1090.
- Bromaghin, J. F. and McDonald, L. L. (1993). Weighted nest survival models. *Biometrics* **49**, 1164–1172.
- Burnham, K. P. and Rexstad, E. A. (1993). Modeling heterogeneity in survival rates of banded waterfowl. *Biometrics* **49**, 1194–1208.
- Ghosh, M., Natarajan, K., Stroud, T. W. F., and Carlin, B. P. (1998). Generalized linear models for small area estimation. *Journal of the American Statistical Association* **93**, 273–282.
- Griffiths, D. A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. *Biometrics* **29**, 637–648.
- Hedeker, D. and Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics* **50**, 933–944.
- Johnson, D. H. (1979). Estimating nest success: The Mayfield method and an alternative. *The Auk* **96**, 651–661.
- Kahn, M. J. and Raftery, A. E. (1996). Discharge rates of Medicare stroke patients to skilled nursing facilities: Bayesian logistic regression with unobserved heterogeneity. *Journal of the American Statistical Association* **91**, 29–41.
- Klett, A. T. and Johnson, D. H. (1982). Variability in nest survival rates and implications to nesting studies. *The Auk* **99**, 77–87.
- Mayfield, H. (1961). Nesting success calculated from exposure. *Wilson Bulletin* **73**, 255–261.
- Mayfield, H. (1975). Suggestions for calculating nest success. *Wilson Bulletin* **87**, 456–466.
- McCulloch, C. E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association* **92**, 162–170.
- Pollock, K. H. and Cornelius, W. L. (1988). A distribution-free nest survival model. *Biometrics* **44**, 397–404.
- Reinert, S. E., Golet, F. C., and DeRagon, W. R. (1981). Avian use of ditched and unditched salt marshes in southeastern New England: A preliminary report. *Proceedings of the Northeast Mosquito Control Association* **27**, 1–23.
- Self, S. G. and Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605–610.
- Williams, D. A. (1975). The analysis of binary responses from toxicological experiments involving reproduction and teratogenicity. *Biometrics* **31**, 949–952.

Received May 1997. Revised August 1998.

Accepted August 1998.

## APPENDIX

Expression for  $d(\alpha, \beta)$

$$d(\alpha, \beta) = \mathcal{P} \left( \frac{\beta \sum_{j=1}^J ((\alpha + J - j)(\alpha + \beta + J - j))^{-1}}{-\sum_{j=1}^J (\alpha + \beta + J - j)^{-1}} \right)$$

where  $\mathcal{P}$  is defined in equation (3).

Expression for  $V(\alpha, \beta)$

The asymptotic variance matrix  $V(\alpha, \beta) = I^{-1}(\alpha, \beta)$ , where  $I(\cdot)$  is the negated observed information,

$$I(\alpha, \beta) = - \begin{pmatrix} i_{11} & i_{12} \\ i_{12} & i_{22} \end{pmatrix}$$

and

$$\begin{aligned} i_{11} &= \sum_{i=1}^n \left[ y_i \left\{ - \sum_{j=1}^{\ell_i} \frac{1}{(\alpha + \ell_i - j)^2} \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^{\ell_i} \frac{1}{(\alpha + \beta + \ell_i - j)^2} \right\} \right. \\ &\quad \left. + (1 - y_i) \left\{ - \sum_{j=1}^{\ell_i} \frac{1}{(\alpha + \ell_i - j)^2} \right. \right. \\ &\quad \left. \left. + \sum_{j=0}^{\ell_i} \frac{1}{(\alpha + \beta + \ell_i - j)^2} \right\} \right], \\ i_{12} &= \sum_{i=1}^n \left[ y_i \left\{ \sum_{j=1}^{\ell_i} \frac{1}{(\alpha + \beta + \ell_i - j)^2} \right\} \right. \\ &\quad \left. + (1 - y_i) \left\{ \sum_{j=0}^{\ell_i} \frac{1}{(\alpha + \beta + \ell_i - j)^2} \right\} \right], \\ i_{22} &= \sum_{i=1}^n \left[ y_i \left\{ \sum_{j=1}^{\ell_i} \frac{1}{(\alpha + \beta + \ell_i - j)^2} \right\} \right. \\ &\quad \left. + (1 - y_i) \left\{ - \frac{1}{\beta^2} + \sum_{j=0}^{\ell_i} \frac{1}{(\alpha + \beta + \ell_i - j)^2} \right\} \right]. \end{aligned}$$