Studies in Avian Biology No. 34:145–148

MODELING NEST-SURVIVAL DATA: RECENT IMPROVEMENTS AND FUTURE DIRECTIONS

JAY ROTELLA

Abstract. Studies of nesting birds commonly seek to estimate nest success and to evaluate relationships between nest-survival rates and hypothesized influential factors. Recently, a number of advances have been made with regard to the analysis of nest-survival data, and improved methods now exist for relaxing assumptions and accounting for potentially important sources of variation in nest-survival data. Methods have been developed that allow diverse covariates of nest-survival rate to be incorporated into analyses of either discrete survival data or failure times. Analysis of binomial data for nest fates over discrete periods has dominated the nest-survival literature and been the subject of many recent advances that extend possible analyses beyond that of the Mayfield method. Recent papers that describe the use of generalized linear mixed models to incorporate covariate effects on nest survival, including some examples that employed a random-effects framework, illustrate the benefits that can be gained from using such models when they are appropriate. Noteworthy examples of the use of analysis of failure times also exist and illustrate the key elements of this type of analysis, which can accommodate censoring, heterogeneity in survival, staggered entry of subjects into the study, and continuous and categorical covariates of survival times. The new analytical approaches should allow avian ecologists to evaluate a broad variety of competing models. By using the various methods interchangeably, future analyses should provide new insights into the nesting ecology of birds.

Key Words: daily survival rate, logistic regression, mixed models, Mayfield, nest success, nest survival.

MODELANDO DATOS DE SOBREVIVENCIA DE NIDO: MEJORAS RECIENTES Y DIRECCIONES FUTURAS

*Resumen.*Comúnmente los estudios de anidación de aves buscan estimar el éxito de anidación y evaluar las relaciones entre las tasas de sobrevivencia de nido, así como hipotetizar los factores que influyen. Recientemente un número de avances han sido desarrollados respecto al análisis de datos de sobrevivencia de nido, y existen ahora métodos mejorados para suavizar las suposiciones, así como el conteo de potenciales fuentes importantes de variación en datos de sobrevivencia de nido. También han sido desarrollados métodos los cuales permiten que diversas tasas de covariantes de sobrevivencia de nido sean incorporadas ya sea a análisis de datos discretos de sobrevivencia, como a veces fallidas. El análisis de datos binomiales para destino de nido sobre periodos discretos ha dominado la literatura respecto a sobrevivencia de nido, y ha sido el tema de varios avances recientes que amplían posibles análisis más allá del método de Mayfield. Artículos recientes los cuales describen la utilización de modelos generalizados lineares mezclados para incorporar efectos covariables en sobrevivencia de nido, incluidos algunos ejemplos que emplearon un marco de efectos al azar, ilustran los beneficios que pueden ser obtenidos al utilizar dichos modelos cuando son apropiados. Existen ejemplos significativos de la utilización de análisis de veces fallidas que ilustran los elementos clave de este tipo de análisis, los cuales pueden adecuar la censura, heterogeneidad en la sobrevivencia, escalonar la entrada de temas en el estudio, y continuas y categóricas covariantes de veces de sobrevivencia. Los nuevos enfoques deberían permitir a los ecólogos de aves evaluar una amplia variedad de modelos competentes. Al utilizar los métodos intercambiablemente, análisis futuros deberían proveer nuevas incursiones en la ecología de anidación de aves.

Studies of nesting birds are widespread in the avian literature. For example, several hundred such papers were published in the year 2004 alone. Studies commonly seek to estimate nest success (the probability that a nest survives from initiation to completion and has at least one offspring leaves the nest) and to evaluate relationships between nest-survival rates and hypothesized influential factors. Accordingly, methods for estimating nest-survival rate have received considerable attention (Mayfield 1961, Johnson 1979, Bart and Robson 1982, Natarajan and McCulloch 1999, Farnsworth et al. 2000, Dinsmore et al. 2002). Williams et al. (2002), Johnson (*this volume*), and Heisey et al. (*this volume*) provided recent and useful reviews of historical development, available approaches, and estimation programs.

The Mayfield method, either in its original form or as expanded by Johnson (1979) and Bart and Robson (1982), requires the assumption of a constant daily survival rate for all nests in a sample over the time period being considered (for further details of the method, its assumptions, and history, see Johnson, *this volume*). However, heterogeneity in daily survival rates among members of the study population can cause estimates of nest success and, in some cases, daily survival rate to be biased (Johnson 1979). Thus, nest-survival data are frequently divided into groups for analysis with the Mayfield method, e.g., stratified by stage of the nesting cycle, season, and habitat conditions (Heisey and Fuller 1985). But, stratification can commonly lead to small samples for many strata if multiple covariates are used to classify data, because most nesting studies investigate how daily survival rates of nests vary in relation to multiple explanatory variables.

To allow greater flexibility in modeling nest-survival data in the presence of heterogeneity, numerous publications have presented methods for relaxing assumptions and accounting for potentially important sources of variation (Dinsmore et al. 2002). Some of the recent improvements have received considerable attention in the avian ecology literature (Dinsmore et al. (2002) had already been cited by 21 publications by the end of 2005, while other advances have received less attention; He et al. (2001) had only been cited twice by the end of 2005). Such differences may have to do with the ease with which new approaches can be implemented in readily available software: Dinsmore et al.'s (2002) approach is implemented in program MARK (White and Burnham 1999) with excellent supporting materials; whereas the approach developed by He et al. (2001) allows great flexibility in modeling but has not yet been accompanied by readily accessible software or code for implementation. Still other methods are simply too new to have yet received attention by the majority of avian ecologists, (Nur et al. 2004, Etterson and Bennett 2005).

Given the diversity of important developments that have recently been made with respect to analysis of nest-survival data, the goal of this paper is to briefly review the latest advances and to comment on areas of future research that would further improve analysis of nest-survival data. The excellent and detailed reviews by Johnson (*this volume*), Heisey et al. (*this volume*) provide much greater detail on the plethora of analysis options that are currently available.

GENERAL APPROACHES

Many of the recent advances can be placed into several broad analytical categories. Here, following the recent treatment of the topic by Williams et al. (2002), two broad classes are used: the analysis of discrete survival data and the analysis of failure times. Heisey et al. (*this volume*) examine these two classes in detail and discuss how they relate to one another.

ANALYSIS OF DISCRETE SURVIVAL DATA

Analysis of binomial data for nest fates over discrete periods has dominated the nest-survival literature and been the subject of many recent advances. Specifically, generalized linear models have been used in a number of recent publications that have extended the analysis of nest-survival data beyond that of the Mayfield method (Dinsmore et al. 2002, Rotella et al. 2004, Shaffer 2004a). As used for nest-survival data, generalized linear models usually employ a logit link between daily survival rate and the covariates of interest, while allowing visitation intervals to vary among observations and making no assumptions about when nest failure occurs.

The recent use of generalized linear mixed models to incorporate covariate effects on nest survival in a random-effects framework takes further advantage of modeling advances (Natarajan and McCulloch (1999); also see reviews by Rotella et al. (2004, *this volume*). Shaffer (2004a), Winter et al. (2005a), and Stephens et al. (2005) employed methods that allow incorporation of random effects along with fixed effects, i.e., mixed models. Several benefits can be gained from using mixed models when they are appropriate. In some situations, the precision of estimates will be increased. Further, when models containing random effects are supported by data, impetus is provided for considering what is responsible for the overdispersion being modeled by the random effect. Such an effort can improve future studies if it leads to the inclusion of new covariates in the fixed effects that reduce the overdispersion. Finally, incorporation of random effects can allow one to make broader inferences, e.g., to a population of study sites rather than just the specific study sites used.

However, one must be cautious with interpretation of estimates obtained in the presence of random effects. In typical studies of nest survival, data are left-truncated because some nests that fail early are not included in the sample (Heisey et al., *this volume*). Under these circumstances, the usual assumption that the mean of a random effect is zero is inappropriate if the design is not balanced (Rotella et al., *this volume*), i.e., if sample sizes are unequal across levels of the covariate being treated as a random factor (e.g., study sites). All else being equal, if care is not taken to balance the sampling design, sample sizes will be larger for those covariate levels (e.g., study sites) that are associated with higher survival rates simply because nests in such settings are expected to survive longer and thus, have a greater chance of entering the sample. When the sample sizes are positively correlated with survival rates, estimates of survival will be biased high to some extent because nests in the sample over represent nests with higher underlying survival rates (Heisey et al., *this volume*). Simulation work completed to date indicates that balanced designs (equal numbers of nests found across levels of the covariate being treated as a random factor) effectively deal with this potential problem (Rotella et al., *this volume*). Thus, given that one will not typically know prior to data analysis whether or not random effects will exist in the data, it seems prudent to adjust search effort such that balanced samples are achieved. The issue of bias from left truncation has received little attention, and more work is needed to determine the magnitude of the problem under typical sampling scenarios.

More information on the use of generalized linear mixed models for nest-survival data can be found in Natarajan and McCulloch (1999), Rotella et al. (2004, *this volume*), and Heisey et al. (*this volume*). Also, the statistics literature contains numerous in-depth treatments of the topic from a more fundamental perspective. As succinctly stated by Williams et al. (2002: 349), the complexity of the computations may limit the ability of many biologists to apply a random-effects approach. However, random-effects modeling is a reasonable and natural way to view nest survival. Williams et al. (2002) believe that the approach will see increasing use, especially when computations are simplified or made more accessible with, for example, Markov chain Monte Carlo methods. The prediction of increasing use may prove correct quite quickly. Biologists are becoming more aware of the benefits of such models and the use of Markov chain Monte Carlo methods due to recent articles explaining the benefits of the approach (Link et al. 2002). Further, Bayesian approaches to modeling nest survival (He et al. 2001, He 2003), which have also recently been extended to include diverse spatio-temporal covariates (J. Cao and C. He, pers. comm.), have proved useful for obtaining parameter estimates.

Although the linear-logistic-modeling approach makes no assumption about the timing of nest failures that occur between two nest visits (but see Aebischer 1999), it is important to consider that the method does require the assumption that nests can be aged correctly. Implicit in this is the assumption that the day of hatching, or fledging, can also be determined correctly. In some studies, uncertainty will exist about nest ages and when transitions among nest stages occur (Williams et al. 2002). For some species, nest age will be a covariate of interest but be unknown for many nests (Stanley 2004a). Also, typical assumptions about the distributions of hatching and fledging events may be violated in some studies (for details, see Etterson and Bennett 2005). Under such circumstances, it will also be difficult to know the exact fledging date for nests and to time final nest checks such that nest fates can be unambiguously determined (Manolis et al. 2000). Several publications have presented methods for dealing with ambiguities in aging and determining fate (Manolis et al. 2000; Stanley 2000, 2004a; Etterson and Bennett 2005). However, these advances have not yet been integrated into models containing complex sets of covariates despite the fact that these circumstances will occur regularly for some species of interest.

Analysis of Failure Times

In contrast to the general analysis approach described above, which focuses on the number of nests surviving over a fixed time period, this approach focuses on time until failure (nest loss) or censoring (Williams et al. 2002). The analysis of failure times has been used in many fields, notably medical science and engineering, and thus, has received a great deal of statistical development and can readily be executed in many statistical packages. Accordingly, diagnostics for analysis of failure times are quite extensive (Nur et al. 2004). Analysis of failure times is compared and contrasted with analysis of discrete survival data by Williams et al. (2002) and Heisey et al. (*this volume*). Analysis of failure times can accommodate censoring (ultimate nest fate need not be known), heterogeneity in survival, staggered entry of subjects into the study, and continuous and categorical covariates of survival times. Accordingly, it should not be surprising that the method was recently applied to the analysis of nest survival by Renner and Davis (2001) and Nur et al. (2004). Nur et al. (2004), Heisey et al. (*this volume*), and Johnson (*this volume*) provide excellent treatments of the subject with respect to the analysis of nest survival. Non-parametric (Kaplan-Meier estimation), semi-parametric (proportional hazards model), and parametric (e.g., Weibull regression) alternatives to analysis of failure time exist. However, for reasons given in Heisey et al. (*this volume*), non-parametric methods have limited utility in most studies of nest survival. Both the semi-parametric and parametric analyses allow continuous and categorical covariates of survival times to be incorporated. Shochat et al. (2005a, b) recently used the proportional-hazards model to successfully analyze nest-survival data of diverse species as functions of multiple covariates. Further, Pankratz et al. (2005) recently provided methods for conducting variance component analyses under general random-effects proportional-hazards models, which makes it feasible to handle correlated time-to-event data, but the applicability of their approach to nest-survival data has not yet been fully evaluated.

As explained by Nur et al. (2004), it is important to realize that estimates of the age of a nesting attempt upon discovery are required for survival-time analysis. However, this requirement exists for the discrete time analyses discussed above as well, unless the analyst is willing to assume constant survival. A further assumption of the analysis of failure times as presented by Nur et al. (2004), although not a general assumption for the method and one that is not necessary with discrete time analysis, is that the date of nest failure is accurately obtained. Thus, short intervals between nest visits are necessary with this method.

Future Directions

The methods discussed here that allow complex sets of covariates to be incorporated in models of nest-survival data do not consider detection probability for nests with different characteristics as do some other methods (Pollock and Cornelius 1988, Bromaghin and McDonald 1993a, McPherson et al. 2003). Accordingly, the methods reviewed here provide estimates that are conditional on the data set. That is, they only represent the population of interest to the extent that the sample of nest data is representative of the population of nest data. A better understanding of how well samples represent populations of interest under various circumstances is needed, e.g., see discussion of random effects above. Information on age-specific nest encounter probabilities can provide information about survival probabilities prior to encounter. The utility of such information has been presented by Williams et al. (2002) and McPherson et al. (2003), and it would be useful if encounter probability could be incorporated into regression models of nest survival. Given the flexibility of the Bayesian approaches (He et al. 2001, He 2003), it would be beneficial if analysis programs and supporting documentation for implementing Bayesian analyses could be made readily available.

Goodness-of-fit tools now exist for models of discrete survival data that include individual covariates and/or random effects (Sturdivant et al., *this volume*) and are available in diverse forms for parametric and semi-parametric analysis of failure times (Lawless 1982). However, estimation of overdispersion remains problematic for analyses of discrete survival data unless random effects are incorporated (see Rotella et al., *this volume*). Further work on this topic would be helpful, but it should be readily apparent that goodness-of-fit and overdispersion are much lesser issues for the complex models now available than they were for simple Mayfield analyses where survival rate is assumed constant over all observations analyzed.

It is clear that new analytical tools allow avian ecologists to evaluate a broader variety of covariates and competing models than was previously possible. The available approaches can be used interchangeably as best suits a particular problem. However, to take full advantage of the approaches, sizeable samples of nests across gradients relevant to the hypotheses of interest will be needed. Interesting questions and well-thought-out sampling designs should, when combined with recent analytical advances, provide new insights into the nesting ecology of birds.

ACKNOWLEDGMENTS

I thank S. L. Jones and G. R. Geupel for organizing the symposium and inviting my participation. This paper benefited from my interactions with S. J. Dinsmore, D. H. Johnson, S. L. Jones, T. L. Shaffer, M. L. Taper, and G. C. White.