

## CORRECTING NESTING-SUCCESS ESTIMATES FOR OBSERVER EFFECTS: MAXIMUM-LIKELIHOOD ESTIMATES OF DAILY SURVIVAL RATES WITH REDUCED BIAS

JAY J. ROTELLA,<sup>1</sup> MARK L. TAPER, AND ANDREW J. HANSEN  
*Biology Department, Montana State University, Bozeman, Montana 59717, USA*

**ABSTRACT.**—We develop a statistical method that simultaneously estimates daily survival rate and observer effect. We used Monte Carlo simulation to (1) evaluate the performance of the model, (2) compare model performance with models that ignore observer effects, and (3) evaluate methods of choosing between competing models of survival. When observer effects were absent, all models produced unbiased estimates of daily survival rate. In the presence of observer effects, however, models that ignore these effects underestimated daily survival rate. In such cases, estimates of nesting success were strongly affected even when observer effects were relatively small. In contrast, estimates of daily survival rate and nesting success produced by the model that considers observer effects consistently had little bias. However, estimates of daily survival rate from this model were less precise than those from the simpler model. Objective criteria for choosing between competing models did not perform well with sample sizes of 150 to 600 because subtle but important observer effects are difficult to detect. Likelihood-ratio tests had low power for rejecting the null hypothesis of no observer effect over a wide range of levels of observer effect and with sample sizes of 150 to 600. Estimates of daily survival rate from models selected based on Akaike's Information Criterion (AIC) had higher bias than estimates from the model that estimates observer effect when observer effect was present. Estimates from AIC-selected models had lower mean squared error than estimates from the model that estimates observer effect when observer effects were small, but the pattern reversed as effects increased. We recommend that researchers estimate observer effects using the more complex model when observer effects are possible and decide whether to use estimates of daily survival from the simpler or more complex model based on analysis results and simulation or analytic results for relevant sample sizes, daily survival rates, and observer effects. To illustrate use of the analytical techniques, we analyzed field data from Dusky Flycatcher (*Empidonax oberholseri*) nests monitored during the nestling stage. The observer effect was estimated to be 1.003 (95% CI 0.866 to 1.162); thus, point estimates of daily survival were very similar from the simpler (0.971; 95% CI 0.957 to 0.985) and more complex model (0.970; 95% CI 0.925 to 1.000). In this case, analysis results and simulation results indicate that the simpler model is adequate and provides an estimate of daily survival rate with small potential bias and increased precision compared with an estimate from the more complex model. Received 30 July 1998, accepted 7 May 1999.

ESTIMATES OF NESTING SUCCESS are vital to many studies of avian populations. To estimate nesting success, researchers typically visit nests periodically to monitor nest contents (Johnson 1979, Bart and Robson 1982, Klett et al. 1986, Pollock and Cornelius 1988). However, nest visits may provide predators and/or brood parasites with cues for finding nests (Nichols et al. 1984, Westmoreland and Best 1985) or, in contrast, they may deter some predators from visiting nests (MacIvor et al. 1990). Thus, field activities used for estimating nesting success may lead to biased estimates of success. The level of bias may vary with visitation

schedule, observation methods, predator assemblage, and habitat characteristics. Moreover, variable and unknown levels of bias can confound comparisons of nesting-success estimates from different studies (Hannon et al. 1993) and yield misleading conclusions regarding population viability.

Although most studies of nesting birds have not assessed whether nest visits affect nesting success (Hannon et al. 1993, Mayer-Gross et al. 1997), the topic has received considerable attention. Götmark (1992) reviewed the literature and identified 225 studies that addressed the effects of observer disturbance on nesting birds. Many of these studies simply provided anecdotal reports of observer effects. Studies

<sup>1</sup> E-mail: rotella@montana.edu

that have evaluated the effects of nest visits more formally typically have tested for a statistical difference in the fates of nests that received different rates of human visitation (e.g. Nichols et al. 1984).

The results and interpretation of tests of the null hypothesis that nest visits do not affect nesting success have been varied (Götmark 1992). Studies that failed to detect a significant difference between treatment groups commonly have concluded that nest visits do not affect success. More recently, some researchers have acknowledged that they may have failed to detect an important difference because statistical power was inadequate to detect small effects of nest visits (Hannon et al. 1993, Mayer-Gross 1997). This failure can have important consequences, because small changes in daily survival rate yield large changes in estimated nesting success (e.g.  $0.96^{26} = 0.346$  vs.  $0.95^{26} = 0.264$ , where nesting success equals daily survival rate raised to the power of the number of nest days). Studies that have detected a significant difference between groups typically have made cautionary statements and recommended ways to reduce the effects of nest visits on nest fate. However, until robust statistical methods are available for reducing observer-induced bias in estimates of nesting success, researchers must be cautious when making inferences from studies that involve nest visits.

One approach to reducing observer-induced bias is to simultaneously estimate observer effects and survival rates of nests from the same data set. A maximum-likelihood approach to this problem was suggested by Bart and Robson (1982) but has not been developed in the literature. Sedinger (1990) attempted to estimate both parameters using a regression approach but concluded that the approach was imprecise and thus was inefficient for detecting observer effects.

We developed a two-variable model that is a generalized form of the one-variable model typically used to estimate daily survival rate and nesting success for nests that are visited periodically (Mayfield 1961, 1975; Johnson 1979; Hensler and Nichols 1981; Bart and Robson 1982). The approach emphasizes accurate estimation of nesting success in light of possible observer effects rather than testing the null hypothesis of no observer effect. To evaluate the performance of the model, we conducted

simulations that evaluated bias and precision of the competing estimators under various combinations of daily survival rate, observer effect on daily survival rate, sample size, and visitation schedule. We also evaluated the performance of various methods of choosing which estimator to use when the true underlying model was known. Finally, we applied the competing estimators to nest data collected for Dusky Flycatchers (*Empidonax oberholseri*) to illustrate application of the observer-effects model to real data.

#### METHODS

*Estimation.*—We estimated daily survival rate for each set of nest data using three estimators: (1) Mayfield's (1961, 1975) heuristically derived ad-hoc estimator (Mayfield model), (2) a maximum-likelihood estimator (MLE) of daily survival rate (Johnson 1979, Hensler and Nichols 1981, Bart and Robson 1982; survival model), and (3) a generalized form of the survival model that is a MLE of daily survival rate and observer effects (observer-effects model). The Mayfield model estimates daily survival rate as one minus the number of nests that fail per nest day (Mayfield 1975). A nest day is recorded for each nest under study beginning the day the nest is found and until the nesting attempt is completed. The Mayfield model accounts for the fact that many nests are not under observation from the day of initiation. The survival model is a special case of the observer-effects model in which observer effects are assumed to be absent. In the observer-effects model, the probability of surviving an interval of  $t_i$  days between observations is:

$$P(\delta_i = 1 \mid t_i) = h \cdot p^{t_i}, \quad (1)$$

where  $\delta_i$  is an indicator variable that takes the value 1 if the nest survives an interval and 0 otherwise,  $h$  is the observer effect on the survival probability that occurs shortly after a nest visit ( $h < 1$  and  $h > 1$  reduce and increase, respectively, a nest's survival probability for a short period after a visit), and  $p$  is the probability that a nest survives natural mortality each day. The observer effect occurs after a nest visit and during a length of time less than or equal to the minimum interval length used in the study. The model is sufficient for observer effects that occur within minutes of a visit or over the course of a day or more (up to the length of the minimum rechecking interval). This model explicitly estimates the observer effect on daily survival rate. When  $h$  equals 1.0, the observer-effects model is identical to the survival model.

The following transformation of equation 1 indicates heuristically how our procedure is able to extract information about both  $h$  and  $p$  from survival

data for multiple interval lengths. When equation 1 is transformed logarithmically, the equation becomes a linear regression equation:

$$[\log(P(\delta_i = 1 \mid t_i)) = \log(h) + \log(p) \cdot t_i], \quad (2)$$

where  $\log(h)$  is an intercept term (assumed to be zero in the survival model),  $\log(p)$  is a slope term, and  $t_i$  is the independent variable. However, we do not estimate using this transformation and least-squares regression because heteroscedasticity among interval lengths will cause this method to be less efficient than methods using nonlinear optimization or generalized linear models (see below). This log-linear transformation will be useful in deriving analytic approximations for the bias and mean squared error of estimates derived from the survival and the observer-effects models (see below).

In addition to equation 1, we make three further assumptions: (1) any observer effects are assumed to influence survival only during the day following the visit, (2) survival rates and human effects are constant throughout the period for which data are analyzed, and (3) all individuals have identical survival probabilities and human effects. Later, we discuss ways in which these assumptions might be relaxed. Given equation 1 and the above assumptions, the number of nests surviving an interval of given length is binomially distributed. Thus, the likelihood for a set of data with observations with a variety of interval lengths can be written as:

$$L(h, p \mid \mathbf{X}) = \prod_{i=1}^T \binom{N_i}{Z_i} (h \cdot p^{t_i})^{Z_i} (1 - h \cdot p^{t_i})^{N_i - Z_i}, \quad (3)$$

Where  $\mathbf{X}$  is a data set,  $i$  indexes the number of distinct visitation-interval lengths,  $T$  is the number of different interval lengths,  $N_i$  is the number of nest observations for intervals of length  $t_i$ , and  $Z_i$  is the number of these observations that survives the interval. Parameter estimates for  $h$  and  $p$  are taken to be the pair of values that maximizes the likelihood in equation 3. The maximization of the likelihood or equivalently of the log-likelihood can be accomplished numerically using any statistics package capable of nonlinear optimization. Parameter estimation can also be accomplished using any generalized linear model package with a binomial distribution for error terms and a log-link function (e.g. PROC GENMOD of SAS 1990; see Appendix 1). One should note that unless data are available for at least two interval lengths, the parameters  $h$  and  $p$  will be confounded and not independently estimable. We calculated estimates from the Mayfield, survival, and observer-effects models using code written by MLT in program Mathcad8 (MathSoft 1998). For the observer-effects model, we constrained estimates of  $p$  between 0.0 and 1.0 and the product of estimates of  $p$  and  $h$  between 0.0 and 1.0. We did not constrain values of  $h$  directly because observers can have positive or negative effects on nest survival (Götmark 1992).

*Inference.*—The observer-effects model can have its parameters estimated via the MLE method. Thus, comparisons of  $p$ ,  $h$ , or both  $p$  and  $h$  can be made among sites, times, or species using standard likelihood-ratio tests (Casella and Berger 1990). Approximate confidence intervals may be constructed using the asymptotic variance/covariance matrix of the parameters (Casella and Berger 1990). The asymptotic covariance matrix is provided by most nonlinear regression or generalized linear model routines that one might use to estimate the parameters of the model. However, because the primary region of interest is near the boundary of the parameter space for the parameter  $p$ , the asymptotic confidence intervals may sometimes be less than satisfactory, with the confidence interval around  $p$  extending beyond a survival probability of one. Better confidence intervals can be constructed by inverting the likelihood-ratio test (Casella and Berger 1990). Contours of the joint confidence intervals can be obtained by inverting tests based on the likelihood given in equation 3. Individual confidence intervals for each of the parameters can be found by inverting tests based on the marginal likelihood or profile likelihood of each parameter (McCullagh and Nelder 1989).

*Simulations.*—We used Monte Carlo simulation to create 1,000 replicate data sets for each combination of the following parameter values:  $h$  set to 1.0, 0.96, 0.92, 0.88, 0.84, or 0.8;  $p$  set to 0.99, 0.98, 0.96, 0.94, 0.92, or 0.9; and  $n$  set to 150, 300, or 600. Because published estimates of observer effect are lacking, we chose a range of values of  $h$  in hopes of bracketing the range of negative observer effects that may exist in real studies. Additional studies will be needed before we can know whether the range of observer effects that we used is appropriate.

We used interval lengths of 2, 3, and 4 days to provide information for interval lengths typically used in studies of nesting passerines (Martin and Geupel 1993). For studies with longer intervals, e.g. studies of nesting ducks (Klett et al. 1986), bias resulting from observer effects will be less for a given level of observer effect. For example, in a study using 7-day intervals, the probability of a nest under observation surviving an interval is  $hp^7$ , whereas in a study using 3-day intervals, the probability is  $hp^3$ .

In each simulation, samples were divided equally among the three interval lengths. Thus, a simulation with a sample size of 300 contained binary survival data for 100 intervals of each length. The sample size does not represent a number of nests; rather, it is the total number of intervals for which survival data were gathered summed over all interval lengths. In a real study, each active nest found and monitored would provide data for at least one interval and could yield data for multiple intervals and interval lengths. In our simulations, we generated independent nest fates (survive or fail) for each interval and

nest from a Bernoulli distribution whose probability of success was  $hp^i$  (see equation 1).

*Evaluation of estimator performance.*—For each set of simulations run ( $n = 1,000$  replicate data sets), we measured the bias and mean squared error of estimates generated from each estimator. For each combination of values of  $h$ ,  $p$ , and sample size, we calculated bias and mean squared error to evaluate trends in estimator performance under varying conditions. Bias is the difference between the average estimate of a parameter (from 1,000 simulations) and the true parameter value. Mean squared error considers both the variance and the bias of an estimator and thus provides a useful measure for comparing competing models (Bain and Engelhardt 1992). We used mean squared error as a primary basis of evaluation because it is the sum of bias<sup>2</sup> and variance, which are the quantities we wished to minimize (Bain and Engelhardt 1992). We did not conduct statistical comparisons of estimator performance. Rather, we considered the average performance over 1,000 simulations to be an accurate representation of actual performance because standard errors of the estimated statistics typically were less than 0.001.

*Allocation of nest-checking effort.*—We investigated the effects of allocation of nest-checking effort by creating 1,000 replicate data sets for each combination of  $h$  and  $p$  values for interval lengths of 1 and 4 days using sampling intensities of 100 randomly generated fates per interval length ( $n = 200$ ) and 150 fates per interval length ( $n = 300$ ). Estimates generated from these simulations were compared with those generated from simulations using 100 fates per interval length for intervals of 2, 3, and 4 days ( $n = 300$ ).

*Evaluation of model-selection procedures.*—Because the survival model and the observer-effects model provide competing models of the process that generates data on nest survival, we (1) explored how well different model-selection procedures performed at choosing the correct model and (2) compared the mean squared error of estimates resulting from various model-selection procedures under various circumstances of  $h$ ,  $p$ , and sample size. We used likelihood-ratio tests to evaluate rates of Type I and II statistical errors for tests performed at significance levels of 0.05 and 0.10. Likelihood-ratio tests, which evaluate the null hypothesis of no observer effect, are possible because the survival model is nested within the observer-effects model. We also used two popular information criteria to choose among models: AIC (Akaike 1973, 1985; Burnham and Anderson 1998) and Schwarz's Information Criterion (SIC; Schwarz 1978). AIC "attempts to select a parsimonious approximating model as a basis for inference about the population sampled" and thus strives to minimize the mean squared error of estimates (Burnham and Anderson 1998:164). In contrast, SIC emphasizes detecting the dimension of the true under-

lying model and places less emphasis on estimation errors. Thus, we did not expect SIC to perform well but evaluated it because of its frequent use in model selection. We compared the mean squared error of estimates of daily survival rate from the survival model, observer-effects model, AIC-selected model, and SIC-selected model for all scenarios simulated.

*Application to real data.*—To demonstrate the use of the observer-effects model with real data, we used nest-survival data from the nestling stage collected during 1998 for Dusky Flycatchers in mature aspen (*Populus tremuloides*) forests in Gallatin County, Montana, and Fremont County, Idaho (J. Rotella and A. Hansen unpubl. data). To reduce observer effects, we did not check a nest if corvids were in view, and we avoided creating trails to nests. We analyzed the data using a generalized linear models approach using PROC GENMOD (SAS 1990; Appendix 1).

## RESULTS

*Estimator performance.*—The Mayfield model and the survival model performed similarly in all simulated scenarios (Appendices 2 and 3). However, the survival model's performance was always slightly better. Therefore, in the remainder of the paper, we present results only for the survival model. Both the survival model and the observer-effects model produced estimates of daily survival rate with negligible bias in the absence of observer effects ( $h = 1$ ; Fig. 1, Appendix 2). When observer effects were present, however, estimates of daily survival rate from the survival model were biased low by an amount that increased as the level of observer effect increased (Fig. 1, Appendix 2). In contrast, the observer-effects model consistently produced estimates with low levels of bias over the range of scenarios evaluated (Fig. 1, Appendix 2).

The average level of bias in estimates of daily survival rate from the survival model yielded highly biased estimates of nesting success even when  $h$  was near 1. For example, with true  $p = 0.96$ ,  $n = 600$  (divided equally among interval lengths of 2, 3, and 4 days), and  $h = 0.92$ , the average level of bias for estimates from the survival model was  $-0.033$ . Using the true  $p$ , nesting success for a 26-day nesting period was  $0.96^{26}$ , or 0.346. However, using the survival model's average estimate of  $p$ , which was 0.927, nesting success was 0.139. For the same scenario, the observer-effects model had an estimated bias of  $-0.0003$ , which yields an average estimate of nesting success of 0.343.

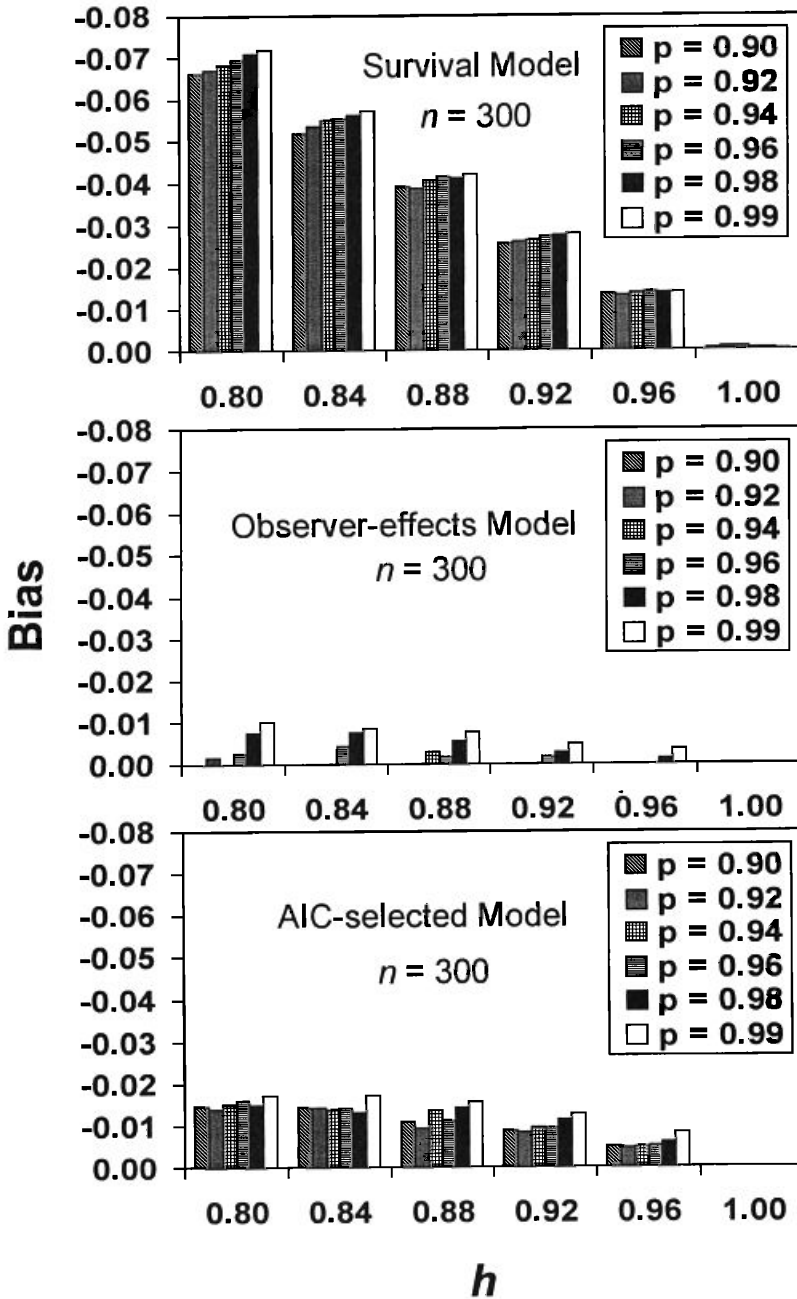


FIG. 1. Bias in estimates of daily survival rate produced by a survival model (upper panel), an observer-effects model (middle panel), and models selected by AIC (lower panel). Bias was calculated from results of 1,000 simulations for each combination of daily survival rate and observer effect ( $h$ ). In each simulation, sample size ( $n$ ) was allocated equally among intervals of 2, 3, and 4 days.

Mean squared error for estimates of daily survival rate from both the survival model and the observer-effects model increased as observer effect increased (Fig. 2, Appendix 2). How-

ever, the mean squared error results for the two models differed in important ways. The mean squared error associated with estimates from the survival model was more strongly associ-

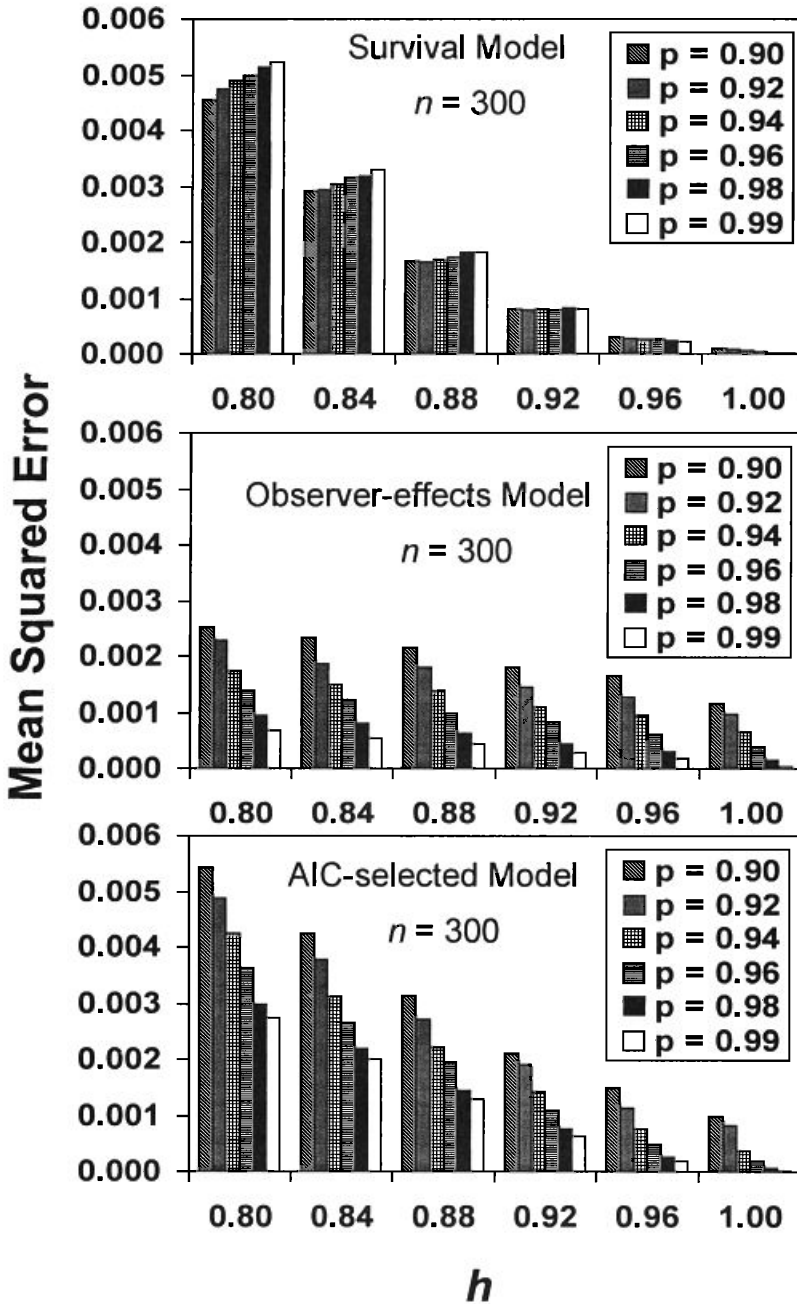


FIG. 2. Mean squared error for estimates of daily survival rate produced by a survival model (upper panel), an observer-effects model (middle panel), and models selected by AIC (lower panel). Mean square error was calculated from results of 1,000 simulations for each combination of daily survival rate and observer effect ( $h$ ). In each simulation, sample size ( $n$ ) was allocated equally among intervals of 2, 3, and 4 days.

ated with increasing observer effects and increased due to changes in bias. In contrast, increases in mean squared error for estimates from the observer-effects model were only weakly associated with the level of observer effect and were associated with changes in variance but not bias.

*Sample size and allocation of nest-checking effort.*—Not surprisingly, estimator performance was better when sample sizes were larger (Appendix 2). This effect was more pronounced for the observer-effects model, which estimates two parameters, than for the survival model, which estimates only one parameter.

The interval lengths between simulated nest visits also affected estimate efficiency. Simulations with 150 nest fates each for 1- and 4-day intervals ( $n = 300$ ) produced a substantially lower mean squared error than did those with the same number of samples allocated evenly among 2-, 3-, and 4-day intervals (Appendices 2 and 3). For example, when  $h$  was 0.96, 0.88, and 0.80, the mean squared error for estimates of daily survival rate were 0.000215, 0.000353, and 0.000512, respectively, when 1- and 4-day intervals were used versus 0.000593, 0.000926, and 0.001423, respectively, when 2-, 3-, and 4-day intervals were used. Furthermore, simulations with 75 nest fates each for 1- and 4-day intervals ( $n = 150$ ) also produced a lower mean squared error than did simulations with a sample size of 300 allocated evenly among 2-, 3-, and 4-day intervals. Although the effect of visitation design has not been fully investigated, it is clear that a combination of short and long intervals will make the most effective use of observations.

*Evaluation of model-selection procedures.*—The power of likelihood-ratio tests to reject the null hypothesis of no observer effect was low when  $h$  was 0.88 to 0.96 (Fig. 3). Even when  $h$  was 0.80 or 0.84, power was still low (typically less than 50%) for most sample sizes and levels of  $p$ . The AIC performed better than likelihood-ratio tests at choosing between the survival model and the observer-effects model (Figs. 3 and 4). For example, for a sample size of 300 and with  $h = 0.96$ , likelihood-ratio tests rejected the survival model in 11 to 18% of simulations depending on daily survival rate, whereas for the same scenario, AIC selected the observer-effects model in 18 to 24% of simulations. However, AIC selected the wrong model a large

proportion of the time, especially when  $h$  was 0.88 to 0.96, or at any level of  $h$  when sample size was below 600. As a consequence, when observer effects were present, estimates of daily survival rate from the model selected by AIC were biased low (Fig. 1) and were more biased than estimates from the observer-effects model (Fig. 5).

The mean squared error of estimates of daily survival rate from AIC-selected models decreased as observer effect approached 1.0 (i.e. no effect) and as daily survival rate increased (Fig. 2). As a result, mean squared error was lower for estimates of daily survival rate from models selected by AIC than from the observer-effects model for some combinations of observer effect and daily survival rate but not for others (Figs. 2 and 6). For large sample sizes ( $n = 600$ ), differences in mean squared error were small except when observer effects were quite large (Fig. 6). With more moderate sample sizes ( $n = 150$ ), differences in mean squared error increased.

As expected, SIC performed poorly under all circumstances, performing worse than AIC at selecting the correct model and in terms of mean squared error. Thus, we do not report numeric results for properties of estimates from SIC-selected models.

*Application to real data.*—During 1998, we monitored 74 Dusky Flycatcher nests and collected survival data for 176 intervals during the nestling stage. Three nests failed during 38 two-day intervals, 2 failed during 55 three-day intervals, and 11 failed during 83 four-day intervals (Appendix 1). The survival model estimated the daily survival rate as 0.971 (95% CI: 0.957 to 0.985). The observer-effects model estimated the observer effect as 1.003 (asymptotic normal 95% CI 0.866 to 1.162) and daily survival rate as 0.970 (asymptotic normal 95% CI 0.925 to 1.000). A likelihood-ratio test between the models failed to reject the survival model ( $\chi^2 = 0.002$ ,  $P = 0.97$ ). Similarly, the AIC value for the survival model (107.84) was lower than that for the observer-effects model (109.83), which suggests that the survival model was more parsimonious. The difference between the AIC values was less than 2.0 (1.99 units); thus, the estimate from the observer-effects model could receive further consideration when making inferences (Sakamoto et al. 1986, Burnham and Anderson 1998). Concepts of

model-selection uncertainty and model averaging in the context of parameter estimation are presented by Burnham and Anderson (1998) but are beyond the scope of this paper.

#### DISCUSSION

*Estimator performance.*—Our simulations indicate that when observer effects exist, the survival model typically produces highly precise wrong answers. The bias created by moderate to small observer effects can have strong consequences when biased daily survival rates are used to calculate nesting success. Thus, the survival model should be used with caution when nest visits may have affected nest survival. In contrast, the observer-effects model consistently yields estimates with low bias. However, the consequence of using the more complex model is reduced precision of the estimates (Burnham and Anderson 1998). Attaining adequate sample sizes and using efficient visitation schedules can circumvent this problem. This will not, however, always be possible for all species and situations.

Besides reducing bias in daily survival rates, the observer-effects model also directly estimates the effects of observers on daily survival rate. We believe that this is more informative than simply conducting a significance test (with a subjectively chosen  $\alpha$  level) of whether an observer effect exists (The Wildlife Society 1995, Steidl et al. 1997). Given a method for estimating observer effects, researchers can evaluate the effects of different study protocols and choose methods that have minimal effects on the animals under study. It is possible that current methods have no effects on nesting success. If current methods do affect nesting success, then the nest-finding cues, e.g. scent or obstructive cover, used by relevant predators and brood parasites should be considered, and researchers should seek methods that do not alter natural levels of nest-finding cues. If methods still affect nesting success, then estimates of daily survival rate from the observer-effects model can be used, but ethical considerations will remain.

The observer-effects model is an extension of the survival model first presented by Johnson (1979). Pollock and Cornelius (1988) presented an alternative method of estimating daily survival rate and stated that simulations should be

done to compare the quality of estimates from their model and the survival model. If future simulation work is done, it should also investigate model forms that estimate observer effects.

*Sample size and allocation of nest-checking effort.*—Clearly, the precision of estimates from the observer-effects model improved with increasing sample size. If smaller sample sizes are used, researchers should be aware that estimate quality from either model and the performance of model-selection procedures will be worse than that depicted in our results. Sampling error occasionally will cause further difficulties when estimating observer effects and daily survival rate for smaller data sets. For example, in simulations with 2-, 3-, and 4-day intervals, a daily survival rate of 0.98, and no observer effect, the observer-effects model produced estimates for 100, 100, and 98.9% of simulated data sets when sample sizes were 600, 300, and 150, respectively. When sample size was set to 75, the observer-effects model failed to converge on estimates for 11.5% of simulated data sets. Thus, sampling needs should be considered before data are collected, and adequate sample sizes should be achieved if use of the observer-effects model is intended.

The nest-visitation schedule also affected precision. We urge others to use simulations or analytic formulas (Appendix 4) for bias and mean squared error to evaluate alternative sample sizes and visitation schedules that may be more relevant to their situations. Simulations and analytic approximations provide very similar results. Including some short intervals in the visitation schedule may lead to a dramatic increase in efficiency. By simulating a variety of situations, nest-checking effort can be used efficiently to achieve the sample sizes needed to attain desired levels of precision.

*Evaluation of model-selection procedures.*—Given the low power of likelihood-ratio tests under most circumstances we simulated, we do not recommend their use for selecting between the survival and the observer-effects model unless adequate (i.e. very large) samples are obtained. We philosophically agree with the use of information criteria for model selection in complex ecological problems where the true underlying model typically is unknown. However, the procedures that we investigated were not able to



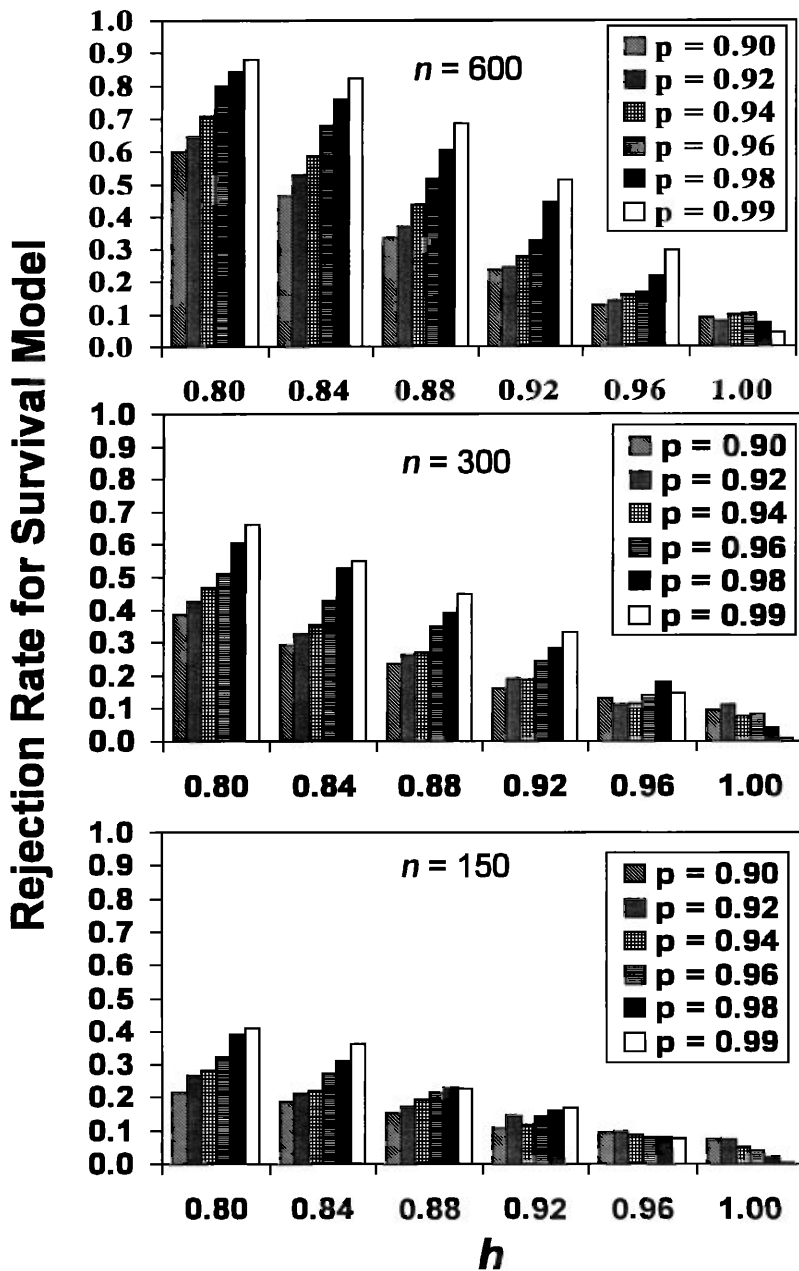


FIG. 3. Proportion of times the survival model was rejected by a likelihood-ratio test ( $P > 0.10$ ) in favor of the observer-effects model. Each proportion was calculated from results of 1,000 simulations for each combination of daily survival rate and observer effect ( $h$ ). In each simulation, sample size ( $n$ ) was allocated equally among intervals of 2, 3, and 4 days.

reliably detect observer effects, even when these effects were pronounced. Thus, when observer effects exist, AIC-selected models generate estimates with greater bias than estimates

from the observer-effects model. Further, the level of bias present is important when biased estimates are converted to estimates of nesting success by raising them to the appropriate

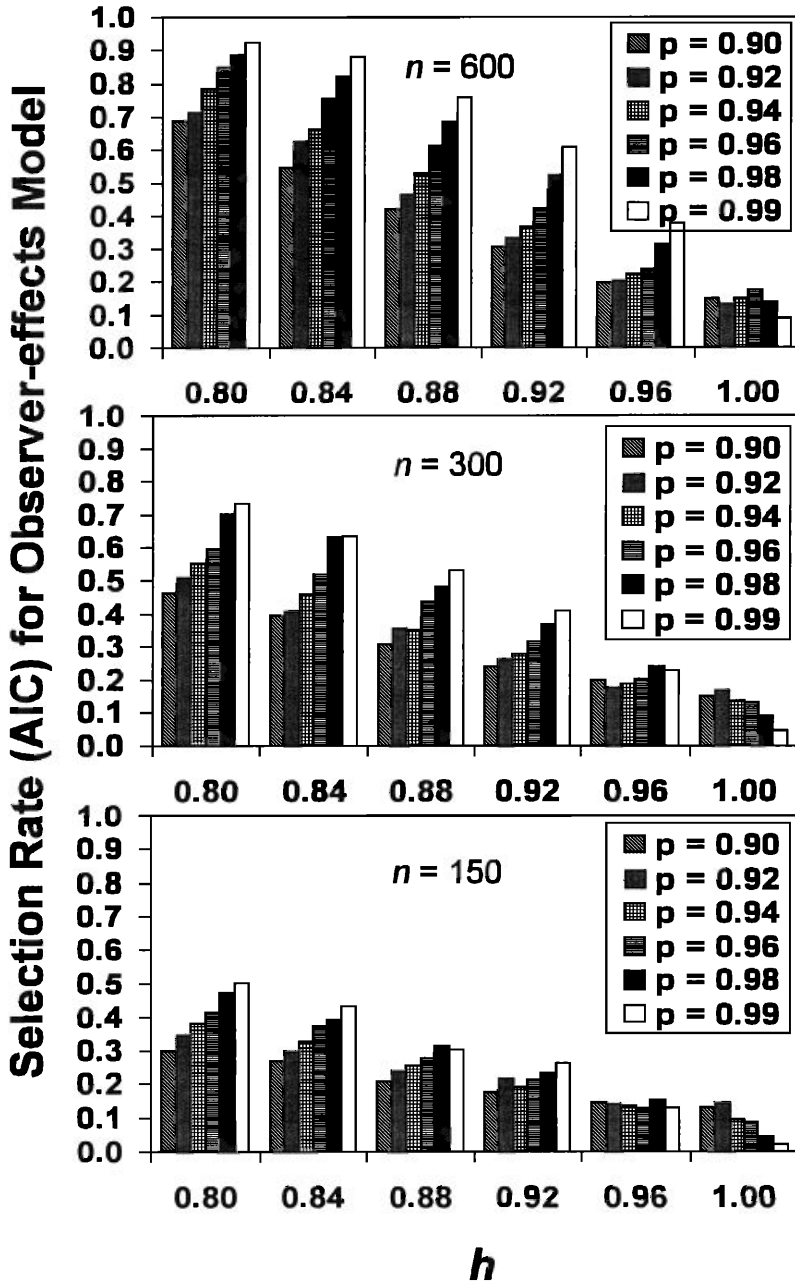


FIG. 4. Proportion of times that the observer-effects model was selected over the survival model based on Akaike's Information Criterion (AIC). Each proportion was calculated from results of 1,000 simulations for each combination of daily survival rate and observer effect ( $h$ ). In each simulation, sample size ( $n$ ) was allocated equally among intervals of 2, 3, and 4 days.

power (days required for a successful nesting attempt).

Of course, AIC does not merely attempt to minimize bias. Rather, it is a data-dependent

selection method concerned with parsimony and minimizing bias and variance (Burnham and Anderson 1998). However, the variance (i.e. mean squared error) of estimates from

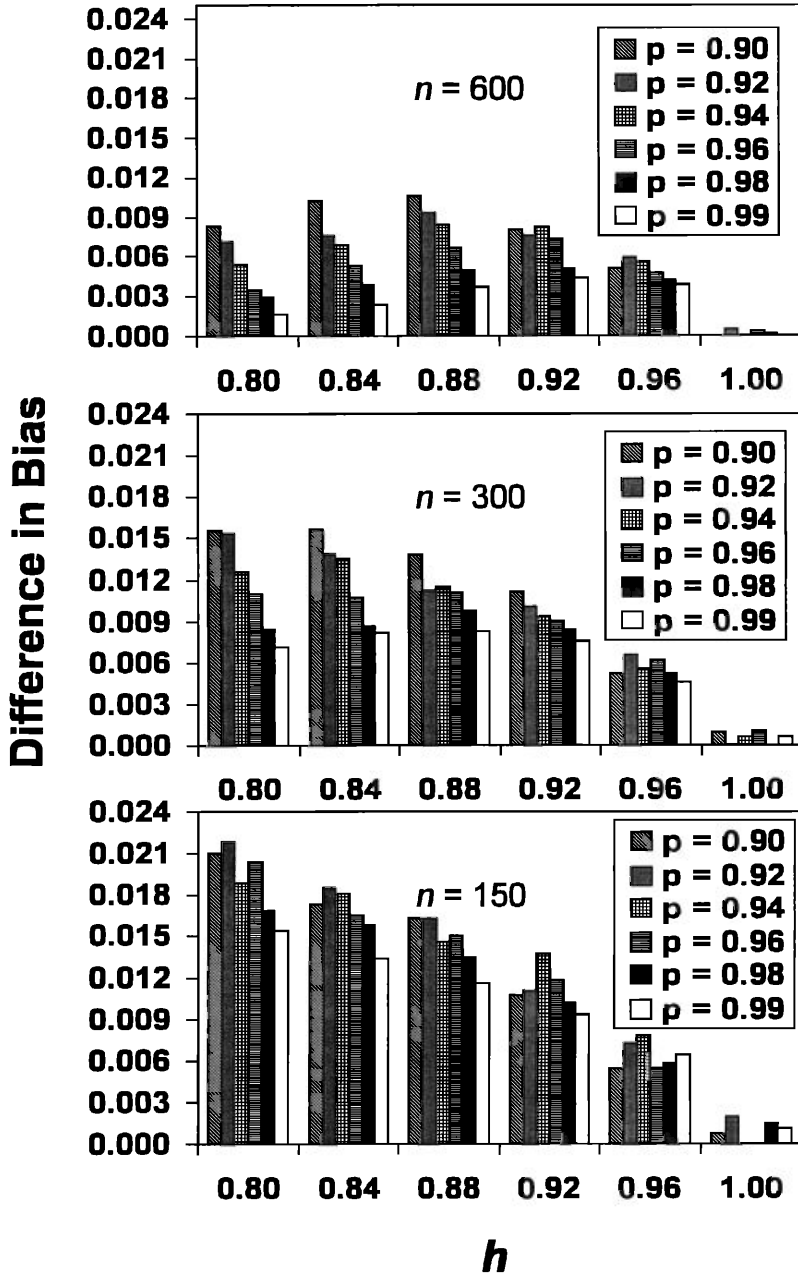


FIG. 5. Average difference in bias between estimates of daily survival rate produced by an observer-effects model and models selected by AIC (bias of observer-effects model minus bias of AIC-selected model). The average difference was calculated from results of 1,000 simulations for each combination of daily survival rate and observer effect ( $h$ ). In each simulation, sample size ( $n$ ) was allocated equally among intervals of 2, 3, and 4 days. Because bias in  $p$  was negative for simulations with  $h < 1.0$ , positive values for the difference in bias indicate that estimates from the AIC-selected model had greater negative bias than did estimates from the observer-effects model.

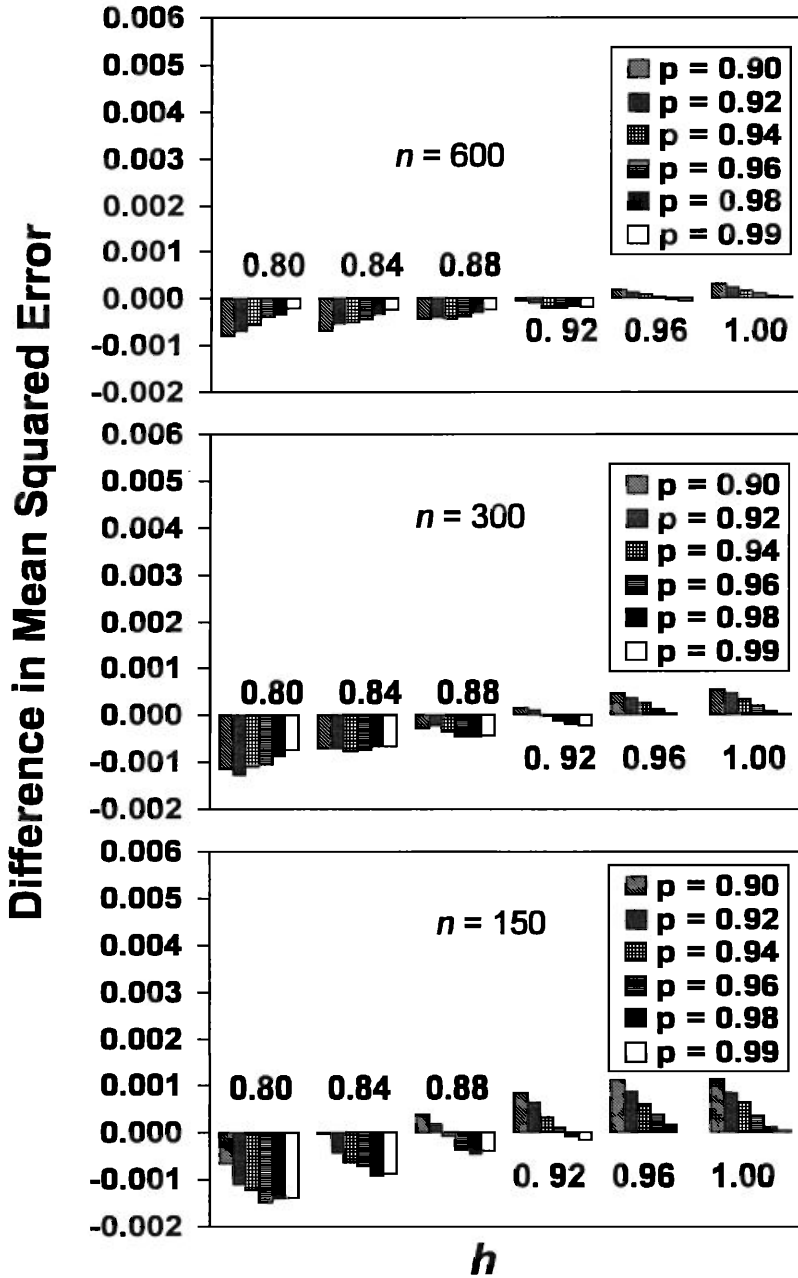


FIG. 6. Average difference in mean squared error between estimates of daily survival rate produced by an observer-effects model and models selected by AIC (mean squared error of observer-effects model minus mean squared error of AIC-selected model). The average difference was calculated from results of 1,000 simulations for each combination of daily survival rate and observer effect ( $h$ ). In each simulation, sample size ( $n$ ) was allocated equally among intervals of 2, 3, and 4 days.

AIC-selected models was higher than that of estimates from the observer-effects model for many of the scenarios simulated. Further, it is not currently possible to know what observer effects are likely to be in most studies because they have not yet been estimated. Researchers should consider carefully whether minimizing bias or mean squared error is more important for their application. If minimizing bias is critical, the observer-effects model should be used. However, if minimizing mean squared error is paramount, we cannot recommend a simple model-selection strategy for all situations that researchers may encounter. Instead, we argue that researchers should (1) estimate observer effects on daily survival rate if they may be present, (2) calculate AIC values for the observer effects model and the survival model, (3) report parameter estimates from both models, and (4) explain how and why they chose among models as they did. Model-selection procedures should be justified based on simulation or analytic results indicating that the procedure used provides reliable estimates, i.e. minimizes mean squared error. If necessary, simulations can be conducted using code available from the authors, and analytic results can be obtained using approximations derived in Appendix 4.

*Application to real data.*—Several features of our analysis of Dusky Flycatcher data are noteworthy. The observer effect (1.003) estimated by the observer-effects model was very close to 1.0. Thus, estimates of daily survival rate are similar for both models, although the estimate from the survival model is more precise than that from the observer-effects model (95% CI 0.957 to 0.985 vs. 0.925 to 1.000, respectively). Neither a likelihood-ratio test nor a comparison of AIC values provided evidence that the observer-effects model should be selected in favor of the simpler survival model. Simulation results suggest that the bias and mean squared error of estimates from AIC-selected models are as small or smaller than they would be from the observer-effects model for the scenario in question. Thus, it seems most appropriate to use the estimate of daily survival rate from the survival model.

*More complex models.*—Under the observer-effects model, a nest visit affects the daily survival rate for a short period following the visit. This scenario seems reasonable if observer ef-

fects occur because predators find nests by watching observers visit nests or by following fresh human scent to nests. Of course, other models of the process are plausible. Observer effects could be assumed to last longer, and terms could be added for the effect on the second day's survival and so on. Perhaps more effectively, observer effects could be modeled as a process that decays over multiple days after a nest visit. Effects could also be modeled as a cumulative process, e.g. multiple visits create trails or reduce incubation and feeding efficiency. The observer-effects model can be modified to accommodate these more complex scenarios if decay or accumulation terms are added to the model in ways suggested by Bart and Robson (1982) for the survival model. Finally, the observer-effects model can be made covariate-dependent to evaluate the need for site-dependent (e.g. habitat-specific) or stage-dependent (different nest stages) estimates of  $h$  and  $p$ . Available software such as PROC GENMOD (SAS 1990) can be used to accomplish such analyses. However, more complex models will require large sample sizes and will impose more constraints on visitation schedules to achieve adequate precision. If one accepts individual variability in survival rates or tolerance for humans, the problem becomes much more difficult. However, it should not be intractable. Work that has been done on overdispersed binomial distributions should be adaptable to this problem (McCullagh and Nelder 1989).

*Related applications.*—Observer effect has also been included as a covariate in analyses of population trends using Breeding Bird Survey data (James et al. 1996). James et al. (1996) employed competing models of population trend; one controlled for potential observer effects and the other did not. They noted that a significant loss in precision occurred when observer effect was incorporated in the model, but they used estimates from the more complex model because of evidence that observer ability to detect birds had changed with time and could be a source of substantial bias if ignored. Covariates have also been used to improve estimates of population vital rates in analyses of data on recaptures or resightings of marked animals (Otis et al. 1978, Lebreton et al. 1992, Nichols et al. 1994). Authors emphasize the importance of considering and estimating nuisance parameters (e.g. capture/recapture probabilities),

which, like observer effects in nesting studies, can lead to biased estimates of the parameters of interest if ignored. Papers on mark-recapture analyses also stress the importance of considering competing models of the process under study and of choosing among models (see Burnham and Anderson 1998).

## ACKNOWLEDGMENTS

We thank J. Bart, D. Johnson, K. Podruzny, and G. White for helpful comments on the manuscript and S. Lele and R. Boik for discussions of statistical approaches. We also thank the numerous technicians that helped with nest searching, especially M. Krasaka.

## LITERATURE CITED

- AKAIKE, H. 1973. Information theory and an extension of the maximum likelihood principle. Pages 267–281 in *Proceedings of the 2nd International Symposium on Information Theory* (B. N. Petran and F. Csáki, Eds.). Akadémiai Kiadó, Budapest, Hungary.
- AKAIKE, H. 1985. Prediction and entropy. Pages 1–24 in *A celebration of statistics: The ISI centenary volume* (A. C. Atkinson and S. E. Fienberg, Eds.). Springer-Verlag, New York.
- BAIN, L. J., AND M. E. ENGELHARDT. 1992. *Introduction to probability and mathematical statistics*, 2nd ed. Duxbury Press, Belmont, California.
- BART, J., AND D. S. ROBSON. 1982. Estimating survivorship when the subjects are visited periodically. *Ecology* 63:1078–1090.
- BURNHAM, K. P., AND D. R. ANDERSON. 1998. *Model selection and inference: A practical information-theoretic approach*. Springer-Verlag, New York.
- CASELLA, G., AND R. L. BERGER. 1990. *Statistical inference*. Wadsworth and Brooks/Cole, Pacific Grove, California.
- GÖTMARK, F. 1992. The effects of investigator disturbance on nesting birds. *Current Ornithology* 9: 63–104.
- HANNON, S. J., K. MARTIN, L. THOMAS, AND J. SCHIECK. 1993. Investigator disturbance and clutch predation in Willow Ptarmigan: Methods for evaluating impacts. *Journal of Field Ornithology* 64:575–586.
- HENSLER, G. L., AND J. D. NICHOLS. 1981. The Mayfield method of estimating nest success: A model, estimators and simulation results. *Wilson Bulletin* 93:42–53.
- JAMES, F. C., C. E. MCCULLOCH, AND D. W. WIENFELD. 1996. New approaches to the analysis of population trends in land birds. *Ecology* 77:13–27.
- JOHNSON, D. H. 1979. Estimating nest success: The Mayfield method and an alternative. *Auk* 96: 651–661.
- KLETT, A. T., H. F. DUEBBERT, C. A. FAANES, AND K. F. HIGGINS. 1986. *Techniques for studying nest success of ducks in upland habitats in the prairie pothole region*. United States Fish and Wildlife Service Resource Publication No. 158.
- LEBRETON, J.-D., K. P. BURNHAM, J. CLOBERT, AND D. R. ANDERSON. 1992. Modeling survival and testing biological hypotheses using marked animals: A unified approach with case studies. *Ecological Monographs* 62:67–118.
- LEHMAN, E. L. 1983. *Theory of point estimation*. John Wiley and Sons, New York.
- MACIVOR, L. H., S. M. MELVIN, AND C. R. GRIFFIN. 1990. Effects of research activity on Piping Plover nest predation. *Journal of Wildlife Management* 54:443–447.
- MARTIN, T. E., AND G. R. GEUPEL. 1993. Nest-monitoring plots: Methods for locating nests and monitoring success. *Journal of Field Ornithology* 64:507–519.
- MATHSOFT, INC. 1998. *Mathcad 8 user's guide*. MathSoft, Inc., Cambridge, Massachusetts.
- MAYER-GROSS, H., H. Q. P. CRICK, AND J. J. D. GREENWOOD. 1997. The effect of observers visiting the nests of passerines: An experimental study. *Bird Study* 44:53–65.
- MAYFIELD, H. F. 1961. Nesting success calculated from exposure. *Wilson Bulletin* 73:255–261.
- MAYFIELD, H. F. 1975. Suggestions for calculating nest success. *Wilson Bulletin* 87:456–466.
- MCCULLAGH, P., AND J. A. NELDER. 1989. *Generalized linear models*, 2nd ed. Chapman and Hall, New York.
- NETER, J., W. WASSERMAN, AND M. H. KUTNER. 1990. *Applied linear statistical models*, 3rd ed. R. D. Irwin, Homewood, Illinois.
- NICHOLS, J. D., H. F. PERCIVAL, R. A. COON, M. J. CONROY, G. L. HENSLER, AND J. E. HINES. 1984. Observer visitation frequency and success of Mourning Dove nests: A field experiment. *Auk* 101:398–402.
- NICHOLS, J. D., J. E. HINES, K. H. POLLOCK, R. L. HINZ, AND W. A. LINK. 1994. Estimating breeding proportions and testing hypotheses about costs of reproduction with capture-recapture data. *Ecology* 73:2052–2065.
- OTIS, D. L., K. P. BURNHAM, G. C. WHITE, AND D. R. ANDERSON. 1978. *Statistical inference from capture data on closed animal populations*. Wildlife Monographs No. 62.
- POLLOCK, K. H., AND W. L. CORNELIUS. 1988. A distribution-free nest survival model. *Biometrics* 44:397–404.
- SAKAMOTO, Y., M. ISHIGURO, AND G. KITAGAWA. 1986. *Akaike information criterion statistics*. KTK Scientific, Tokyo.
- SAS INSTITUTE. 1990. *SAS/STAT user's guide*, ver-

- sion 6, 4th ed. SAS Institute, Inc., Cary, North Carolina.
- SCHWARZ, G. 1978. Estimating the dimension of a model. *Annals of Statistics* 6:461–464.
- SEDINGER, J. S. 1990. Effects of visiting Black Brant nests on egg and nest survival. *Journal of Wildlife Management* 54:437–443.
- STEIDL, R. J., J. P. HAYES, AND E. SCHAUBER. 1997. Statistical power analysis in wildlife research. *Journal of Wildlife Management* 61:270–279.
- THE WILDLIFE SOCIETY. 1995. Journal news. *Journal of Wildlife Management* 59:196–198.
- WESTMORELAND, D., AND L. B. BEST. 1985. The effect of disturbance on Mourning Dove nesting success. *Auk* 102:774–780.
- Associate Editor: J. D. Brawn

APPENDIX 1. Code for estimating daily survival rate with the survival model and the observer-effects model using PROC GENMOD (SAS 1990). Data are from 1998 for Dusky Flycatchers ( $n=74$  nests).

---

\*SAS code to estimate daily survival rates from the (1) survival model (Johnson 1979, Hensler and Nichols 1981, Bart and Robson 1982) and the (2) observer-effects model.<sup>a</sup> Data are for Dusky Flycatcher nests in the nestling stage. The "noint" option creates the survival model. WALDCI requests CIs based on asymptotic normality. LRCI requests CIs base on the profile likelihood.;

---

```
data survival;
  0 input days status$ count;
  cards;
2 lived 35
2 died 3
3 lived 53
3 died 2
4 lived 72
4 died 11
;
* Generate estimates for the survival model;
proc genmod data = survival;
  freq count;
  model status = days / noint LINK=LOG DIST=BINOMIAL WALDCI LRCI;
run;
* Generate estimates from the observer-effects model;
proc genmod data = survival;
  freq count;
  model status = days / LINK=LOG DIST=BINOMIAL WALDCI LRCI;
run;
```

---

<sup>a</sup> The observer-effects model as presented in SAS code is unconstrained. Estimates should be checked for the following conditions before being accepted: (1) If estimated values of  $h$  and  $p$  yield  $hp \geq 1$ , recode the values for the variable "days" to be equal to days - 1. Next, use the survival model to estimate  $p$  and use the estimate of  $p$  to estimate  $h$  as  $1/p$ . (2) If the estimated value of  $p$  is greater than 1, run the analysis again without "days" in the model statement. Do not use the "noint" option. This is an intercept-only model that fixes the value of  $p$  at 1.0 and attributes all mortality to  $h$ .

APPENDIX 2. Performance of three estimators (Mayfield, survival, and observer-effects models) of daily nest-survival rate based on results of Monte Carlo simulations. For each combination of daily survival rate, observer effect ( $h$ ), and sampling intensity, 1,000 simulations were conducted using sampling intensities of 200, 100, or 50 fates per interval length and interval lengths of 2, 3, and 4 days. Results are presented for a representative subset of daily survival rates used in simulations.

$h$	$n$	Bias			Mean squared error		
		Mayfield	Survival	Observer effects	Mayfield	Survival	Observer effects
<b>Daily survival rate = 0.99</b>							
1.00	600	-0.0002	-0.0002	0.0003	0.000005	0.000005	0.000036
0.96	600	-0.0142	-0.0139	-0.0016	0.000215	0.000206	0.000116
0.92	600	-0.0281	-0.0274	0.0026	0.000810	0.000771	0.000172
0.88	600	-0.0428	-0.0416	-0.0042	0.001866	0.001756	0.000245
0.84	600	-0.0582	-0.0561	-0.0040	0.003416	0.003179	0.000281
0.80	600	-0.0746	-0.0716	-0.0065	0.005617	0.005167	0.000394
1.00	300	-0.0005	-0.0004	0.0010	0.000012	0.000011	0.000052
0.96	300	-0.0140	-0.0137	-0.0038	0.000222	0.000213	0.000182
0.92	300	-0.0287	-0.0280	-0.0051	0.000866	0.000822	0.000298
0.88	300	-0.0432	-0.0419	-0.0073	0.001924	0.001809	0.000467
0.84	300	-0.0591	-0.0570	-0.0089	0.003587	0.003332	0.000597
0.80	300	-0.0748	-0.0717	-0.0099	0.005693	0.005232	0.000730
1.00	150	-0.0012	-0.0012	0.0009	0.000025	0.000025	0.000076
0.96	150	-0.0139	-0.0136	-0.0034	0.000245	0.000234	0.000222
0.92	150	-0.0282	-0.0274	-0.0081	0.000886	0.000840	0.000489
0.88	150	-0.0436	-0.0423	-0.0143	0.002025	0.001899	0.000934
0.84	150	-0.0596	-0.0574	-0.0151	0.003708	0.003438	0.001139
0.80	150	-0.0754	-0.0722	-0.0162	0.005880	0.005394	0.001360
<b>Daily survival rate = 0.96</b>							
1.00	600	-0.0013	-0.0005	0.0015	0.000026	0.000022	0.000249
0.96	600	-0.0146	-0.0132	-0.0010	0.000245	0.000205	0.000382
0.92	600	-0.0287	-0.0266	-0.0002	0.000865	0.000747	0.000452
0.88	600	-0.0437	-0.0407	0.0001	0.001957	0.001701	0.000591
0.84	600	-0.0585	-0.0545	0.0006	0.003490	0.003032	0.000636
0.80	600	-0.0749	-0.0697	0.0000	0.005677	0.004913	0.000795
1.00	300	-0.0015	-0.0007	0.0031	0.000046	0.000041	0.000400
0.96	300	-0.0155	-0.0141	0.0008	0.000302	0.000255	0.000593
0.92	300	-0.0294	-0.0273	-0.0004	0.000950	0.000821	0.000821
0.88	300	-0.0443	-0.0413	0.0000	0.002056	0.001787	0.000926
0.84	300	-0.0595	-0.0554	-0.0033	0.003652	0.003171	0.001197
0.80	300	-0.0748	-0.0696	-0.0047	0.005736	0.004961	0.001423
1.00	150	-0.0021	-0.0012	0.0018	0.000106	0.000096	0.000560
0.96	150	-0.0158	-0.0144	-0.0028	0.000379	0.000325	0.000889
0.92	150	-0.0304	-0.0282	-0.0034	0.001092	0.000945	0.001214
0.88	150	-0.0440	-0.0410	-0.0064	0.002144	0.001862	0.001588
0.84	150	-0.0598	-0.0557	-0.0060	0.003825	0.003316	0.001936
0.80	150	-0.0772	-0.0718	-0.0078	0.006264	0.005406	0.002136
<b>Daily survival rate = 0.90</b>							
1.00	600	-0.0047	-0.0003	-0.0016	0.000082	0.000052	0.000701
0.96	600	-0.0177	-0.0123	-0.0003	0.000390	0.000217	0.000831
0.92	600	-0.0322	-0.0256	0.0012	0.001123	0.000730	0.001000
0.88	600	-0.0461	-0.0383	0.0006	0.002215	0.001545	0.001100
0.84	600	-0.0618	-0.0526	0.0014	0.003919	0.002850	0.001235
0.80	600	-0.0755	-0.0650	0.0014	0.005817	0.004322	0.001349
1.00	300	-0.0051	-0.0007	0.0016	0.000162	0.000117	0.001215
0.96	300	-0.0189	-0.0134	0.0003	0.000500	0.000302	0.001680
0.92	300	-0.0320	-0.0254	0.0023	0.001191	0.000788	0.001788
0.88	300	-0.0467	-0.0388	0.0031	0.002350	0.001652	0.002097
0.84	300	-0.0610	-0.0519	0.0012	0.003919	0.002855	0.002195
0.80	300	-0.0768	-0.0662	0.0010	0.006114	0.004558	0.002687
1.00	150	-0.0058	-0.0013	0.0023	0.000300	0.000230	0.002128
0.96	150	-0.0193	-0.0138	0.0004	0.000670	0.000442	0.002621



APPENDIX 2. Continued.

<i>h</i>	<i>n</i>	Bias			Mean squared error		
		Mayfield	Survival	Observer effects	Mayfield	Survival	Observer effects
0.92	150	-0.0327	-0.0260	-0.0011	0.001376	0.000938	0.002977
0.88	150	-0.0476	-0.0397	-0.0011	0.002648	0.001892	0.003528
0.84	150	-0.0636	-0.0542	-0.0040	0.004421	0.003250	0.004234
0.80	150	-0.0785	-0.0677	-0.0054	0.006627	0.004968	0.004760

APPENDIX 3. Performance of three estimators (Mayfield, survival, and observer-effects models) of daily nest-survival rate based on results of Monte Carlo simulations. For each combination of daily survival rate, observer effect (*h*), and sampling intensity, 1,000 simulations were conducted using sampling intensities of 150 or 100 fates per interval length and interval lengths of 1 and 4 days.

<i>h</i>	<i>n</i>	Bias			Mean squared error		
		Mayfield	Survival	Observer effects	Mayfield	Survival	Observer effects
<b>Daily survival rate = 0.96</b>							
1.00	300	-0.0014	-0.0006	-0.0005	0.000062	0.000056	0.000157
0.96	300	-0.0183	-0.0167	-0.0001	0.000414	0.000354	0.000215
0.92	300	-0.0349	-0.0325	-0.0001	0.001328	0.001153	0.000271
0.88	300	-0.0525	-0.0490	-0.0014	0.002889	0.002511	0.000353
0.84	300	-0.0721	-0.0670	0.0001	0.005378	0.004652	0.000417
0.80	300	-0.0906	-0.0840	-0.0012	0.008407	0.007224	0.000512
1.00	150	-0.0015	-0.0007	-0.0005	0.000089	0.000081	0.000306
0.96	150	-0.0183	-0.0168	-0.0001	0.000465	0.000399	0.000381
0.92	150	-0.0354	-0.0330	-0.0001	0.001433	0.001244	0.000505
0.88	150	-0.0536	-0.0499	-0.0014	0.003078	0.002673	0.000635
0.84	150	-0.0724	-0.0673	0.0001	0.005528	0.004775	0.000755
0.80	150	-0.0918	-0.0851	-0.0012	0.008723	0.007489	0.000828

APPENDIX 4. Analytic approximation derivations for bias and mean squared error of parameter estimates.

Let

$$Y_i = \ln\left(\frac{Z_i}{N_i}\right). \tag{4}$$

It can be easily shown that the expected value of  $Y_i$  can be approximated as:

$$E(Y_i) \approx \ln(h) + \ln(p) \cdot t_i. \tag{5}$$

Further, by the delta method (Lehman 1983: theorem 5.2) the variance of  $Y_i$  can be approximated as:

$$V(Y_i) \approx \frac{1 - h \cdot p^{t_i}}{N_i \cdot h \cdot p^{t_i}}. \tag{6}$$

The approximations in equations 5 and 6 will both be good, as long as  $h \cdot p^{t_i}$  is "close" to one and  $N_i$  is "large." The parameters  $\ln(h)$  and  $\ln(p)$  can be estimated with weighted linear regression (Neter et al. 1990). Construct a diagonal weight matrix  $\mathbf{W}$  as  $\mathbf{W}_{ii} = 1/V(Y_i)$ ,  $\mathbf{W}_{ij} = 0$ ;  $i \neq j$ . Construct the design matrix  $\mathbf{X}$  as  $\mathbf{X}_{i1} = 1$ ,  $\mathbf{X}_{i2} = t_i$ . Then,

$$\hat{\beta} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{Y}. \tag{7}$$

$\hat{\beta}$  is a multivariate normal random vector whose first element estimates  $\ln(h)$  and whose second element estimates  $\ln(p)$ .  $\hat{\beta}$  has a variance/covariance matrix given by:

$$\sigma^2(\hat{\beta}) = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}. \tag{8}$$

Estimates of the original parameters are arrived at by back transforming.

$$\hat{p} = \exp(\hat{\beta}_2). \tag{9}$$

The bias and mean squared error (MSE) of  $\hat{p}$  can be calculated from the properties of the lognormal distribution (Casella and Berger 1990). For convenience let

$$\gamma = \exp\left(\frac{\sigma^2(\hat{\beta})_{2,2}}{2}\right). \tag{10}$$

Then we can write

$$\text{bias}(\hat{p}) \approx p \cdot (\gamma - 1) \tag{11}$$

$$\text{MSE}(\hat{p}) \approx p^2 \cdot (\gamma^4 - 2 \cdot \gamma + 1). \tag{12}$$

Bias and MSE for the survival model can be estimated as above if the design matrix  $\mathbf{X}$  is constructed just as a vector of interval lengths ( $t_i$ ). Bias and MSE for  $h$  have the same form as equations 11 and 12 only substituting  $h$  for  $p$  and redefining  $\gamma$  as  $\exp(\sigma^2(\hat{\beta})_{1,1}/2)$ .