Journal of Applied Statistics, Vol. 29, Nos. 1-4, 2002, 103-106

Discussion comments on: the use of auxiliary variables in capture-recapture modelling. An overview

GARY C. WHITE, *Department of Fishery and Wildlife Biology, Colorado State University, USA*

The incorporation of auxiliary variables into capture-recapture models provides the biologist the opportunity to explore many biological hypotheses, and thus provides a powerful technique for extending the analysis of encounter history data. However, cause-and-effect relationships cannot be gained without experimental manipulation of the system. That is, even though an auxiliary variable in a capture-recapture model may provide a good predictor of survival, unless the auxiliary variable is part of a manipulative investigation, the relationship is only correlative and lacks the stronger frame of inference provided by experimental manipulation. Opportunities to incorporate covariates into designed, manipulative experiments should not be declined.

The Achilles' heel of using auxiliary variables in capture-recapture modelling is assessing goodness-of-fit. With the procedures presented by Burnham & Anderson (1998), quasi-likelihood approaches are used for model selection and for adjustments to the variance of the estimates to correct for over-dispersion of the capture-recapture data. An estimate of the over-dispersion parameter, c, is necessary to make these adjustments. However, no general, robust, procedures are currently available for estimating c. Although much of the goodness-of-fit literature concerns testing the hypothesis of lack of fit, I instead view the problem as estimation of c.

Logistic regression also suffers from the detriment of no generally robust methodology to estimate c. The goodness-of-fit procedure derived from likelihood theory suggests the deviance of the model is chi-square distributed with degrees of freedom equal to sample size of the data minus the number of estimated parameters

Correspondence: Department of Fishery and Wildlife Biology, Colorado State University, Fort Collins, CO 80523, USA.

ISSN 0266-4763 print; 1360-0532 online/02/010103-04
DOI: 10.1080/02664760120108476

(McCullagh & Nelder, 1989). The chi-square distribution follows because the deviance is the likelihood ratio test statistic of a saturated model, with the number of parameters equal to the sample size versus the fitted model. However, the deviance statistic is known not to be chi-square distributed for small samples (McCullagh & Nelder, 1989), and the resulting estimate $\hat{c}$ = deviance/df is a biased estimate of the true overdispersion. The Hosmer-Lemeshow test (Hosmer & Lemeshow, 1989) provides an alternative procedure for goodness-of-fit in logistic regression, but suffers from the subjectivity of breaking a continuous variable into discrete classes. Although apparently a useful approach for logistic regression, the test has not been extended to handle capture-recapture data.

The parametric bootstrap procedure currently implemented in Program MARK (White & Burnham, 1999) does not provide an unbiased estimate of c, based on simulations with the Cormack-Jolly-Seber model. Simulated data with $c = 2$ were generated for 5, 10 and 15 occasions, ϕ equals 0.5 and 0.8, p equals 0.5, with 100 releases on each occasion. The 100 sets of simulated data for each of the six scenarios included extra binomial variation by simulating a single encounter history, but recording two animals as experiencing that history. Results shown in Fig. 1 suggest a negative bias for the bootstrap procedure, whereas the goodness-of-fit tests provided in Program RELEASE (Burnham *et al.*, 1987) provide reasonable estimates of c near the expected value of 2.

Another difficulty with incorporating covariates into capture-recapture models that allow estimation of population size, N, is to model N directly as a function of covariates, e.g. habitat characteristics. For likelihood-based estimators, $\hat{N} \geqslant M_{t+1}$, i.e. the estimate must be the number of marked animals in the population. Typically, to implement this constraint, numerical optimization is performed on the quantity $\hat{f}_0 = \hat{N} - M_{t+1}$, i.e. the estimated number of animals never captured. Incorporation of covariates to model $\hat{N}$ is not possible directly in software packages

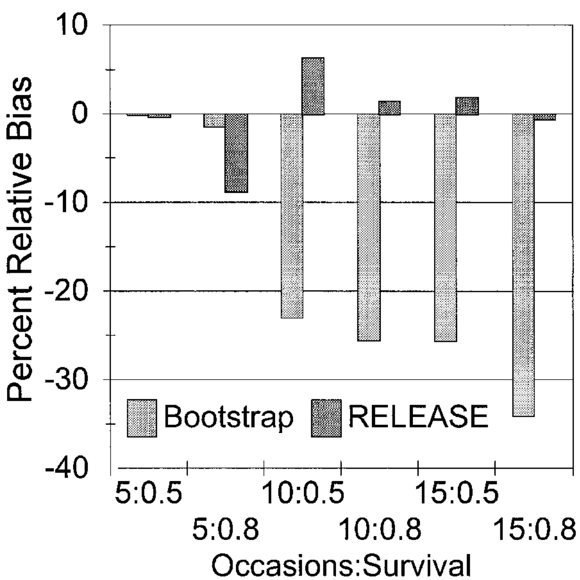

FIG. 1. Simulation results for estimates of the overdispersion parameter (c) from chi-square tables of Program RELEASE (Burnham *et al.*, 1987) and the parametric bootstrap procedure of Program MARK (White & Burnham, 1999). Each scenario consists of 100 replications.

such as MARK (White & Burnham, 1999) because f_0 is being modelled instead of N.

Another issue concerns the loss of efficiency of models that do not include N in the likelihood and condition on the animals observed (e.g. Huggins, 1989, 1991; Alho, 1990) and models that include N in the likelihood (Otis *et al.*, 1978, and references therein). The Huggins models provide a much more flexible framework to incorporate auxiliary information, particularly individual-specific information, into the model.

To evaluate the difference in bias and efficiency of the Huggins and unconditional likelihood models, I conducted a small simulation study with 2000 replications of the M_0 estimator of Otis *et al.* (1978) with $N = 500$, for 5 occasions, and 9 values initial capture and recapture rates of $p = c = 0.0690$, 0.0970, 0.1294, 0.1674, 0.2140, 0.2752, 0.3690, 0.4507 and 0.6019. These values correspond to the fraction of the population seen during the 5 occasions $[p^\star = 1 - (1 - p)^5]$ of 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95 and 0.99. Results for percentage relative bias and standard deviation of $\hat{N}$ are shown in Figs 2 and 3, respectively. Both estimators become biased high as the proportion of the population decreases, although this bias is greater for the Huggins estimator. As expected from likelihood theory, the Huggins estimator is also less efficient, although not markedly so. Basically, neither estimator performs well for small capture probabilities.

In summary, three difficulties with incorporating auxiliary variables into capture-recapture models have been considered: estimation of over-dispersion, constraints on $\hat{N}$, and the performance of estimators when the likelihood is conditioned on only the animals observed. The general estimation of over-dispersion is an unsolved issue. Constraints on $\hat{N}$ may be possible with more robust software. Lastly, the Huggins estimator performs reasonably well with data where at least 60% of the population was captured.

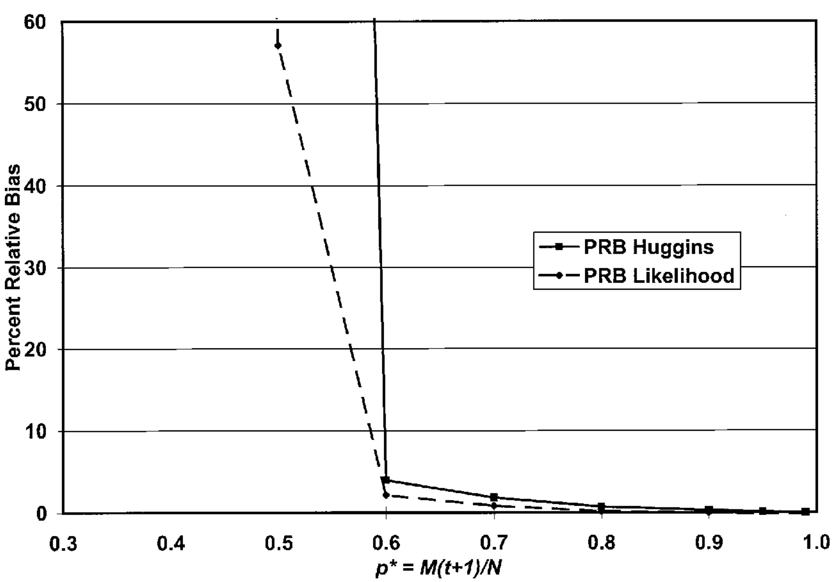

FIG. 2. Percentage relative bias of the Huggins (1989, 1991) and unconditional likelihood (Otis *et al.*, 1978) estimators as a function of the proportion of the population captured. Each scenario was simulated 2000 times.

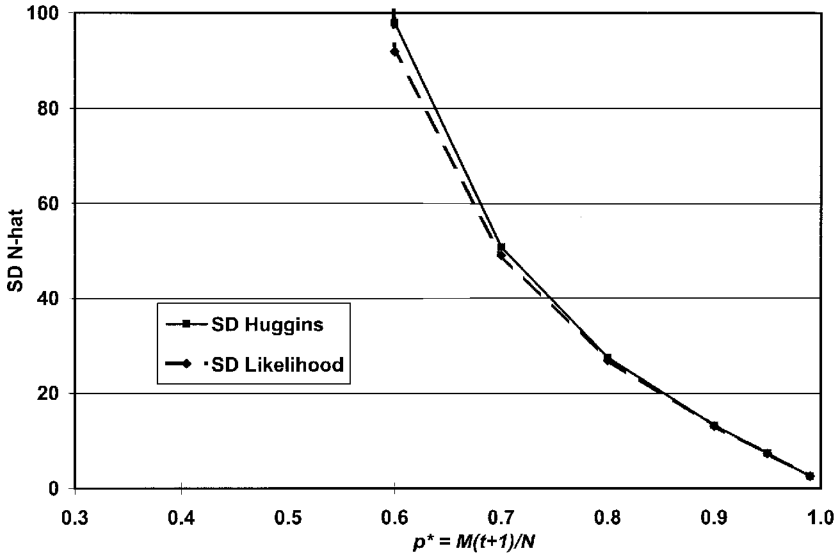

FIG. 3. Standard deviation of the estimates of population size for the Huggins (1989, 1991) and unconditional likelihood (Otis *et al.*, 1978) estimators as a function of the proportion of the population captured. Each scenario was simulated 2000 times.

REFERENCES

ALHO, J. M. (1990) Logistic regression in capture-recapture models, *Biometrics*, 46, pp. 623-635.

BURNHAM, K. P., ANDERSON, D. R., WHITE, G. C., BROWNIE, C. & POLLOCK, P. H. (1987) Design and analysis methods for fish survival experiments based on release-recapture. American Fisheries Society Monograph No. 5, Bethesda, Maryland, USA, 437 pp.

BURNHAM, K. P. & ANDERSON, D. R. (1998) *Model Selection and Inference: A Practical Information-theoretic Approach* (New York, Springer) 353 pp.

HOSMER, D. W., JR. & LEMESHOW, S. (1989) *Applied Logistic Regression* (New York, Wiley) 307 pp.

HUGGINS, R. M. (1989) On the statistical analysis of capture-recapture experiments, *Biometrika*, 76, pp. 133-140.

HUGGINS, R. M. (1991) Some practical aspects of a conditional likelihood approach to capture experiments, *Biometrics*, 47, pp. 725-732.

MCCULLAGH, P. & NELDER, J. A. (1989) *Generalized Linear Models*, 2nd edn (New York, Chapman and Hall), 511 pp.

OTIS, D. L., BURNHAM, K. P., WHITE, G. C. & ANDERSON, D. R. (1978) Statistical inference from capture data on closed animal populations, *Wildlife Monograph*, 62, 135 pp.

WHITE, G. C. & BURNHAM, K. P. (1999) Program MARK: survival estimation from populations of marked animals, *Bird Study*, 46, Supplement, pp. 120-138.