Precision of Population Viability Analysis

STEPHEN P. ELLNER,*#§ JOHN FIEBERG,* DONALD LUDWIG,† AND CHRIS WILCOX‡

*Biomathematics Program, Department of Statistics, North Carolina State University, Raleigh, NC 27695–8203, U.S.A.
†Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, British Columbia
V6T 1Z2, Canada
‡Environmental Studies Department, 339 Natural Sciences 2, University of California, Santa Cruz, CA 95064, U.S.A.

Although population viability analysis (PVA) is widely used in setting conservation policy, there is disagreement about the usefulness of this method. Objections have been raised concerning the precision of predictions in view of the short time series of data available and the sensitivity of estimates of extinction risk to estimated parameters (Hamilton & Moller 1995; Taylor 1995; Groom & Pascual 1998; Ludwig 1999). Beissinger and Westphal (1998) reviewed the use of demographic models for endangered-species management. They pointed out that poor data cause difficulties in parameter estimation, which in turn lead to unreliable estimates of extinction risk. There are additional problems with model validation, especially if all available data have been used to estimate parameters. Beissinger and Westphal (1998) recommend that PVA be used to evaluate relative rather than absolute extinction risk, that projections be made only over short time periods, and that simple models be used rather than complicated ones. Fieberg and Ellner (2000) showed that values of the quasi-extinction probability—the probability of decline to a lower population threshold—for a simple model range between 80% and 5% as the value of the intrinsic growth rate r varies between -0.03 and $+0.02$. Such a range in estimates of r is common for data sets of moderate size. They also show that a precise estimate of extinction probability over a horizon of t years requires between $5t$ and $10t$ years of data, and that similar results hold for age-structured models. In a recent article, Brook et al. (2000) used field data on declining species to test the accuracy and bias of PVA models for predicting extinction risk and concluded that "PVA is a valid and sufficiently accurate tool for categorizing and managing endangered species." We examined the reasons for these differing assessments of the value of PVA.

Brook et al. (2000) considered 21 long-term data sets (11–57 years, mean = 24.9). They used the first half of each set to estimate parameters, with a variety of PVA software packages. They used the second half of each set to test the predictions of each package. The predictions were tested by comparing the actual and predicted numbers of species that declined below a given threshold abundance, which was defined by specifying a target risk level and using the PVA model to identify the corresponding threshold. They applied a significance test to these differences between the predicted and actual number of species falling below the thresholds, and failed to detect any significant differences (their Table 1 & Fig. 1). They applied similar tests to final population sizes with analogous results.

Brook et al. (2000) tested PVA models only on actual field data, whereas other authors used simulated data. Various methods for testing PVA predictions with field data are reviewed by McCarthy and Broome (2000) and McCarthy et al. (2001). As McCarthy et al. (2001) emphasize, a valid test must be based on data that were not used to fit the model. In view of the amount and quality of data necessary for parameterizing a complex population model, field data sets adequate for both parameterizing and testing a model will generally be scarce. In contrast, simulated data allow for the replication necessary to evaluate the precision of model-derived estimates relative to the true extinction risk or population growth rate, which are known exactly for a simulated population. Simulated data may also have shortcomings. Taylor et al. (2000) discuss the merits of using simulated data for model testing.

Modern statistical practice requires that every statistical estimate be accompanied by a measure of its precision if inferences are to be drawn from these estimates (Sokal & Rohlf 1981). This general principle has special force for extinction-risk estimates based on PVA, for which investigations have repeatedly shown lack of precision. For complicated statistical models such as those used in PVAs, there may be no way to derive confidence intervals analytically, but they are readily obtained from

258

simulations. White (2000) and Sæther et al. (2000) show how to use simulations to obtain confidence intervals for complicated models. Alternatively, confidence intervals can be obtained for some models by repeatedly resampling real data (Sokal & Rohlf 1981). Brook et al. (2000) do not estimate confidence intervals for their extinction-risk estimates, leaving the precision of their estimates unclear.

The tests of Brook et al. (2000) are applied to an ensemble of species rather than to individual species. Such a test provides information about the bias in the risk estimates, but it provides little information about their precision because the expected total number of extinctions depends only on the average risk over the ensemble. We illustrate this distinction with two models (Fig. 1) based

on Tables 2 and 3 in the supplementary material of Brook et al. (2000). The first model (Fig. 1a & 1b) is the unstructured density-independent growth model $n(t + 1) = n(t)\exp(r(t))$ (the model used for theoretical analyses by Dennis et al. [1991] and Fieberg & Ellner [2000]), with $r(t)$ a Gaussian(μ, σ^2) random variable with $\mu = -0.044$, $\sigma = 0.3$. The value of σ is a rounded average over species of the values in Brook et al.'s Table 2, and the absolute value of μ (0.044) equals the average of $|\mu|$ over species. The second model (Fig. 1c & 1d) is an age-structured Leslie matrix model with logistic density dependence in neonate survival. The mean and coefficient of variation of each vital rate for this model were derived by taking the average of the corresponding values for each species in Brook et al.'s Table 3, and rounding
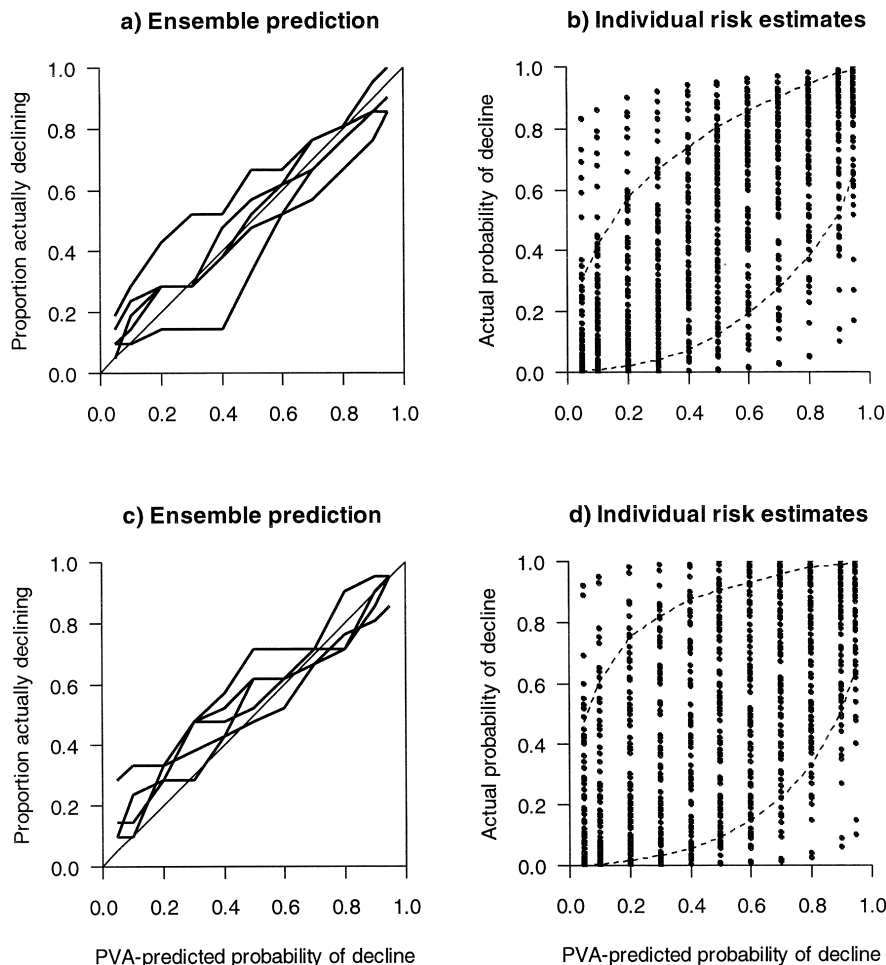

Figure 1. Results from simulated population viability analyses (PVA) using parameters based on Tables 2 and 3 in the supplementary material of Brook et al. (2000). (a) Comparison of observed and predicted total number of extinctions for a set of 21 species, as in Fig. 1 of Brook et al. (2000), based on the unstructured model described in the text. Five replicates are plotted. For each of the 105 trials (21 species × 5 replicates), we simulated a PVA as described in the text. (b) Comparison of actual and estimated extinction risks for each of the 105 trials used in panel (a). Actual extinction risks were calculated by running 25,000 simulations of the model using the true parameter values and recording the fraction of runs crossing below each estimated threshold from (a). Dashed lines show the tenth and ninetieth percentiles of the distributions of estimated extinction risks. (c) Comparison of observed and predicted total number of extinctions, as in panel (a), based on the age-structured, density-dependent model described in the text. All simulations began with a population in stable age distribution for the mean matrix with a population density of 500. (d) Comparison of actual and estimated extinction risks, as in panel (b), for the 105 trials with the age-structured, density-dependent model.

slightly. Our model species had juvenile and adult stages, with first breeding at 2 years of age and a maximum age of 15 years. The mean (coefficient of variation) of vital rates were as follows: adult annual fecundity, 0.6 (40%), juvenile survival, 0.6 (15%), adult survival, 0.75 (20%). Gaussian distributions truncated at 0 were used for random variations in vital rates. We assumed that neonate survival (or, equivalently, adult fecundity) was a function of adult density, with the value given above holding at 500 adults and decreasing linearly to zero at 1000 adults—hence, a maximum adult fecundity of 1.2/year at low adult densities. We assumed that this form of density dependence was known a priori, but the mean and variance of the survival rates and maximum fecundity had to be estimated from data.

For each model we simulated a PVA as follows. We assumed $n = 24$ years of data and generated 12 simulated years of data to simulate the data-collection process (i.e., 12 simulated $r(t)$ values for the unstructured model, 12 values each of adult survival, juvenile survival, and maximum adult fecundity for the age-structured model). We parameterized the models by computing the sample mean and standard deviation of the simulated data for each vital rate. We then simulated the parameterized model 25,000 times to determine a series of quasi-extinction thresholds yielding extinction risks of 5%, 10%, 20%, and so forth, over a 12-year time period. We performed one model run with the true parameter values for each species, and we recorded the total number of species crossing below each threshold.

Figures 1a and 1c are analogous to Fig. 1 of Brook et al., showing good agreement between actual and PVA–estimated total number of extinctions over ensembles of 21 test cases. Figures 1b and 1d compare the actual and estimated risks for each species individually. The spread in individual risk estimates is wide, so these estimates would not be reliable for assessing or comparing individual species. Fig. 1d illustrates that using a more realistic (and therefore more complicated) model only aggravates the problem, even though the amount of data was increased in exact proportion to the larger number of parameters in the more complex model and the density dependence was assumed to be known a priori. These results show that the ensemble-level tests of Brook et al. (2000) were inadequate to assess the precision of PVA risk estimates. The results of Fieberg and Ellner (2000) suggest that PVA will not be precise unless the sample size greatly exceeds the prediction interval.

There is an additional reason for caution in applying the results of Brook et al. (2000): their conclusions were drawn from failure to reject null hypotheses. In such a case, proper inference requires that the size of the Type II error be examined by a power analysis (Peterman 1990; Thompson et al. 2000), but Brook et al. do not provide such an analysis. The bottom line of their Table 1, where errors of a factor of two too high or too low are not statistically significant, suggests that extremely large differences would be

required to reject the null hypothesis for the 5% extinction risk typical of published PVAs. This lack of power is not due to poor choice of methods but is an unavoidable consequence of the scarcity of long-term data sets.

How useful is PVA, in view of its limitations? Thompson et al. (2000) provide an example to be emulated. They use PVA and power analysis to explore the consequences of some management strategies. They base their analysis on a range of assumptions about the rate of population decline, rather than relying on a single estimated rate. They use power analysis to determine the length of data series required to detect the decline. An important feature of this analysis and others presented in the same special section of *Conservation Biology* is careful accounting for uncertainty and its consequences for management.

Similar uses of PVA for comparative purposes take advantage of its ability to summarize diverse data and explore the consequences of alternative actions (Groom & Pascual 1998; Burgman & Possingham 2000). For example, Lindenmayer and Possingham (1996) used PVA to compare timber-management options for conservation of Leadbetter's possum (*Gymnobelideus leadbeateri*) in southeastern Australia and found that the ranking among the options was robust to parameter and model uncertainties. This is quite different from attempting to make quantitative predictions of extinction risk based on small data sets. It is limited, however, to within-species comparisons of relative risk under different management scenarios. When a comparison between species is attempted—for example, to assay which has the greatest need for immediate intervention—the uncertainties in the absolute risk estimates for each species are likely to be too high for such comparisons to be meaningful.

Brook et al. (2000) tested predictions over an average time interval of about 13 years, so their results are relevant to 10- and 20-year time frames used for World Conservation Union listing of critically endangered and endangered species. Such short-term predictions can be important for formulating a management framework. But published PVAs generally have used much longer time intervals, 50–200 years, with 100 years the most common (Beissinger & Westphal 1998; Fieberg & Ellner 2001). As our results indicate, risk estimates for longer time intervals are increasingly imprecise, with most estimates near zero or 1 because the predicted long-term risk is extremely sensitive to the estimated mean growth rate (Dennis et al. 1991; Ludwig 1999; Fieberg & Ellner 2000).

Coulson et al. (2001) raise additional concerns about the conclusions of Brook et al. (2000). They caution that data for most threatened or endangered species will be sparse and of lower quality than the data sets analyzed by Brook et al. Furthermore, they argue that predictions are likely to be accurate only if future mean and variation of vital rates or population growth will be similar to the data used to parameterize the model. Populations that are subject to rare high-recruitment events or cata-

strophic mortalities will therefore provide additional difficulties with regard to model parameterization and reliability. In summary, the results of Brook et al. (2000) are not sufficient to justify PVA as an accurate tool for categorizing individual species, even for short (10- to 20-year) time intervals. Their results provide evidence (subject to concerns about power) that risk and growth rate estimates are unbiased, which implies that PVA could be useful in predicting the total loss rate for a large group of species.

For assessment of individual species, it is essential to account for imprecision in parameter estimates and its consequences for risk assessment. A variety of tools are available. We have already mentioned confidence intervals on the risk of extinction within a given time horizon. An analogous tool is prediction intervals for the time to extinction (Engen & Sæther 2000), but methods to compute these are available only for very simple models. Alternatively, extinction probabilities can be calculated for the range of plausible parameter values by Bayesian methods (Ludwig 1996). Similarly, using frequentist methodology, one can calculate the level of confidence that the true probability of extinction is less than any value (essentially a p value associated with the true probability of extinction). One may then display the range of likely extinction probabilities or weight them by a measure of their plausibility in light of the data. A weighting procedure has the merit of producing a single measure of risk, but this measure is sensitive to various assumptions made in the assessment process. Perhaps a better strategy would be to produce a prediction interval for the population size over the entire time horizon of interest, taking into account uncertainty in parameter estimates (as described by Sæther et al. 2000). This eliminates the subjective choices of a specific time horizon and quasi-extinction threshold for computing an extinction risk. Population viability analysis may then be one useful tool among a variety of decision-making aids, which might include historical and predicted future habitat losses, recent population trends, and genetic considerations. As is often the case with important environmental problems, even the best available science may be unable to provide the level of predictability and accuracy we might wish.

Acknowledgments

Literature Cited

Beissinger, S. R., and M. I. Westphal. 1998. On the use of demographic models of population viability in endangered species management. Journal of Wildlife Management **62:**821–841.

Brook, B. W., J. J. O'Grady, A. P. Chapman, M. A. Burgman, R. Akcakaya, and R. Frankham. 2000. Predictive accuracy of population viability analysis in conservation biology. Nature **404:**385–387.

Burgman, M., and H. P. Possingham, 2000. Population viability analysis for conservation: the good, the bad and the undescribed. Pages 97–112 in A. G. Young and G. M. Clarke, editors. Genetics, demography and viability of fragmented populations. Cambridge University Press, Cambridge, United Kingdom.

Coulson, T., G. M. Mace, E. Hudson, and H. Possingham. 2001. The use and abuse of population viability analysis. Trends in Ecology and Evolution **16:**219–221.

Dennis, B., P. L. Mulholland, and J. M. Scott. 1991. Estimation of growth and extinction parameters for endangered species. Ecological Monographs **61:**115–143.

Engen, S., and B-E. Sæther. 2000. Predicting the time to quasi-extinction for populations far below their carrying capacity. Journal of Theoretical Biology **205:**649–658.

Fieberg, J., and S. P. Ellner. 2000. When is it meaningful to estimate an extinction probability? Ecology **81:**2040–2047.

Fieberg, J., and S. P. Ellner. 2001. Stochastic matrix models for conservation biology: a comparative review of methods. Ecology Letters **4:**244–266.

Groom, M. J., and M. A. Pascual. 1998. The analysis of population persistence: an outlook on the practice of viability analysis. Pages 4–27 in P. L. Fiedler and P. M. Kareiva, editors. Conservation biology. 2nd edition. Chapman and Hall, New York.

Hamilton, S., and H. Moller 1995. Can PVA models using computer packages offer useful conservation advice? Sooty Shearwaters *Puffinus griseus* in New Zealand as a case study. Biological Conservation **73:**107–117.

Lindenmayer, D. B., and H. P. Possingham. 1996. Ranking conservation and timber management options for Leadbeater's possum in southeastern Australia using population viability analysis. Conservation Biology **10:**235–251.

Ludwig, D. 1996. Uncertainty and the assessment of extinction probabilities. Ecological Applications **6:**1067–1076.

Ludwig, D. 1999. Is it meaningful to estimate a probability of extinction? Ecology **80:** 298–310.

McCarthy, M. A., and L. S. Broome. 2000. A method for validating stochastic models of population viability: a case study of the mountain pygmy-possum (*Burramys parvus*). Journal of Animal Ecology **69:**599–607.

McCarthy, M. A., H. P. Possingham, J. R. Day, and A. J. Tyre. 2001. Testing the accuracy of population viability analysis. Conservation Biology **15:**1030–1038.

Peterman, R. M. 1990. Statistical power analysis can improve fisheries research and management. Canadian Journal of Fisheries and Aquatic Sciences **47:**2–15.

Sæther, B-E., S. Engen, R. Lande, P. Arcese, and J. N. M. Smith. 2000. Estimating the time to extinction in an island population of Song Sparrows. Proceedings of the Royal Society of London, Series B **267:**621–626.

Sokal, R., and F. J. Rohlf. 1981. Biometry. 2nd edition. W. H. Freeman, San Francisco.

Taylor, B. L. 1995. The reliability of using population viability analysis for risk classification of species. Conservation Biology **9:**551–558.

Taylor, B. L., S. J. Chivers, S. Sexton, and A. Dizon. 2000. Evaluating dispersal estimates using mtDNA data: comparing analytical and simulation approaches. Conservation Biology **14:**1287–1297.

Thompson, P. M., B. Wilson, K. Grellier, and P. S. Hammond. 2000. Combining power analysis and population viability analysis to compare traditional and precautionary approaches to conservation of coastal cetaceans. Conservation Biology **14:**1253–1263.

White, G. 2000. Population viability analysis: data requirements and essential analyses. Pages 288–331 in L. Boitani and T. K. Fuller, editors. Research techniques in animal ecology: controversies and consequences. Columbia University Press, New York.