

21. Begun, D. R. Knuckle-walking ancestors. *Science* **259**, 294 (1993).
 22. Andrews, P. J. & Martin, L. B. Cladistic relationships of extant and fossil hominoids. *J. Hum. Evol.* **16**, 101–118 (1987).
 23. Begun, D. R. Miocene fossil hominids and the chimp-human clade. *Science* **257**, 1929–1933 (1992).
 24. Ruvolo, M. Molecular phylogeny of the hominoids: inferences from multiple independent DNA sequence data sets. *Mol. Biol. Evol.* **14**, 248–265 (1997).
 25. Dainton, M. & Macho, G. A. Did knuckle walking evolve twice? *J. Hum. Evol.* **36**, 171–194 (1999).
 26. Hunt, K. D. The postural feeding hypothesis: an ecological model for the origin of bipedalism. *S. Afr. J. Sci.* **9**, 77–90 (1996).
 27. Hewes, G. W. Food transport and the origin of hominid bipedalism. *Am. Anthropol.* **63**, 687–710 (1961).
 28. Lovejoy, C. O. The origin of man. *Science* **211**, 341–350 (1981).
 29. Napier, J. R. & Davis, P. The forelimb skeleton and associated remains of *Proconsul africanus*. *Foss. Mamm. Afr.* **16**, 1–70 (1959).
 30. Grine, F. E. & Susman, R. L. Radius of *Paranthropus robustus* from Member 1, Swartkrans Formation, South Africa. *Am. J. Phys. Anthropol.* **84**, 229–248 (1991).

Acknowledgements

We thank L. Aiello, E. Delson, B. Demes, J. Fleagle, F. Grine, W. Jungers, S. Larson, D. Lieberman, O. Pearson, D. Pilbeam, J. Polk, E. Sarmiento, R. Susman and B. Wood for providing valuable comments; F. Grine, L. Gordon, R. Thorington, R. Potts, A. Walker, and C. Ward for providing access to specimens in their care; and The Henry Luce Foundation.

Correspondence and requests for materials should be addressed to B.G.R. (e-mail: brich@gwu.edu)

Predictive accuracy of population viability analysis in conservation biology

Barry W. Brook*†, Julian J. O’Grady*, Andrew P. Chapman*, Mark A. Burgman‡, H. Resit Akçakaya§ & Richard Frankham*

* Key Centre for Biodiversity and Bioresources, Department of Biological Sciences, Macquarie University, New South Wales 2109, Australia

‡ Environmental Science, School of Botany, University of Melbourne, Parkville, Victoria 3052, Australia

§ Applied Biomathematics, 100 North Country Road, Setauket, New York 11733, USA

† Present address: Key Centre for Tropical Wildlife Management, Northern Territory University, Darwin, Northern Territory 0909, Australia

Population viability analysis (PVA) is widely applied in conservation biology to predict extinction risks for threatened species and to compare alternative options for their management^{1–4}. It can also be used as a basis for listing species as endangered under World Conservation Union criteria⁵. However, there is considerable scepticism regarding the predictive accuracy of PVA, mainly because of a lack of validation in real systems^{2,6–8}. Here we conducted a retrospective test of PVA based on 21 long-term ecological studies—the first comprehensive and replicated evaluation of the predictive powers of PVA. Parameters were

estimated from the first half of each data set and the second half was used to evaluate the performance of the model. Contrary to recent criticisms, we found that PVA predictions were surprisingly accurate. The risk of population decline closely matched observed outcomes, there was no significant bias, and population size projections did not differ significantly from reality. Furthermore, the predictions of the five PVA software packages were highly concordant. We conclude that PVA is a valid and sufficiently accurate tool for categorizing and managing endangered species.

PVA is a way to predict the probability of population (or species) extinction, by inputting actual life-history information and projecting it forward using stochastic computer simulation^{1–4}. PVA is commonly used in a comparative way, to evaluate the effectiveness of different management options; because of this it has been argued that PVA predictions do not need to be precise^{2,9,10}. However, there is no clear dichotomy between relative and absolute predictions—as conservation actions entail costs, management decisions are based not only on whether the proposed strategy is sufficient to achieve recovery, but also on whether the likely benefit will justify the expenditure. These considerations require PVA predictions to be quantitatively reliable. Uncertainties surrounding its predictive reliability have led to conclusions drawn from PVA being rejected in the law courts¹¹. It is therefore essential that the predictions of PVA be compared and tested^{12–14}. Here we assess the predictive accuracy of PVA, and determine whether different generic PVA computer packages differ in their predictive capabilities.

Historical data have been used to test and improve the predictions of complex climatic¹⁵, economic¹⁶, geological¹⁷ and ecological¹⁸ models. As PVA models have important stochastic components, the conclusions of past studies based on a single test^{13,19} lack power, and are prone to case-specific peculiarities²⁰. Consequently, a valid test of PVA predictions must incorporate a large number of data sets to obtain representative assessments. We conducted retrospective analyses on 21 wildlife populations—8 avian, 11 mammalian (representing 9 species), 1 reptilian and 1 piscine.

The 21 data sets used were the only long-term studies we identified that presented data of sufficient duration and quality to be suitable for retrospective testing (see Methods). Five of the most commonly applied ‘generic’ PVA packages (GAPPS, INMAT, RAMAS Metapop, RAMAS Stage and VORTEX) were used. These packages are all suitable for generic, single-population risk assessments. They offer the most realistic prospects for improving PVA as they are subject to wide scrutiny, are repeatedly used and are frequently revised and updated. All have been used in the management and conservation of endangered species. The key features and differences between the five PVA packages are given in the Supplementary Information.

To avoid circularity, the total data available for each population was split. The first half was used to develop and parameterize the models; the latter half was reserved for testing the accuracy of the PVA predictions. To ensure that the two time periods were kept completely separate, no information from the second half of the population history was used in formulating or parameterizing the

Table 1 Comparison of actual versus predicted quasi-extinction risks

	Expected number	GAPPS*	INMAT	R META	R STAGE	VORTEX
No. $N(\text{actual}) < N(Q 90\%)$	18.9	17	16	19	17	17
G-test P		0.48	0.07	0.94	0.21	0.21
No. $N(\text{actual}) < N(Q 50\%)$	10.5	11	10	8	11	8
G-test P		0.65	0.83	0.27	0.83	0.27
No. $N(\text{actual}) < N(Q 10\%)$	2.1	3	1	3	4	3
G-test P		0.48	0.38	0.54	0.21	0.54

The number of actual populations that declined below the predicted population size corresponding to quasi-extinction probabilities of 90%, 50% and 10%, for the population viability analysis software packages GAPPS, INMAT, RAMAS Metapop, RAMAS Stage and VORTEX. The sample is based on 21 retrospective studies. None of the PVA packages’ predictions differed significantly from the expected outcome based on goodness of fit (G) tests.

* GAPPS crashes at very large population sizes because of memory limitations, so the fish was not modelled with this package (expected numbers adjusted to 18, 10 and 2).

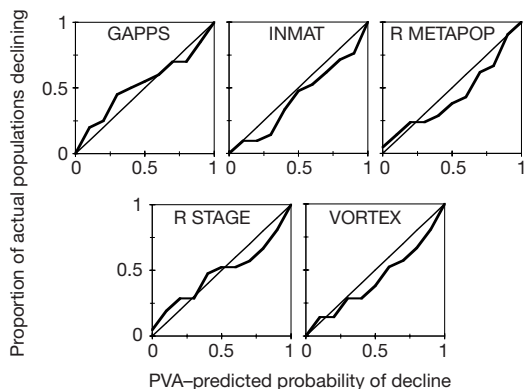


Figure 1 Plot of the PVA-predicted probability of population decline (quasi-extinction risk) versus the actual proportion of the 21 real populations that decline below the corresponding threshold size. These threshold sizes represent different percentage declines in different populations, but are always associated with the same level of risk. For example, half (10.5) of the 21 historical populations should have actually declined below the size assigned a 50% probability by the PVA. For each of the five PVA software packages, a perfect fit with reality lies on the 45° line.

PVA models. All predictive tests were done within species and then pooled across taxa. The accuracy of PVA was assessed by comparing the predicted quasi-extinction risk (the likelihood that the population will decline below a given size in a specified time frame²¹) and population size projections with reality (see Methods).

The results showed a surprisingly close relationship between the PVA predictions and the historical behaviours of the 21 real populations. The quasi-extinction probabilities predicted by the PVAs were not significantly different from the proportion of actual populations that declined below a given threshold (Fig. 1). The risks of decline predicted by the PVAs did not deviate significantly from reality (Table 1), although there was a trend towards slightly pessimistic predictions. Furthermore, there were no overall differences between PVA software packages in their predictions of quasi-extinction risks at 90%, 50% and 10% risk levels (for example, at the 50% quasi-extinction level, Kruskal-Wallis rank test, $H = 1.01$, d.f. = 4, $P = 0.907$).

Actual population sizes consistently fell within the bounds predicted by the stochastic simulations. The projected mean final population sizes were not significantly different from the actual final population sizes (for any of the PVA packages), on the basis of a distribution test combining all 21 studies. Likewise, predicted and actual population growth rates (r) did not differ significantly. There was no consistent over- or underestimation of future population numbers (Fig. 2). A sign test on the predictive bias was not significant for any package, nor when pooled across all packages. There was also no difference in bias across the five PVA packages ($G[\text{Heterogeneity}] = 0.957$, d.f. = 4, $P = 0.916$).

The predictions of the five PVA packages were significantly correlated across the 21 retrospective studies, on the basis of both quasi-extinction probabilities and the fit of projections to future population sizes (Table 2). There were some significant differences

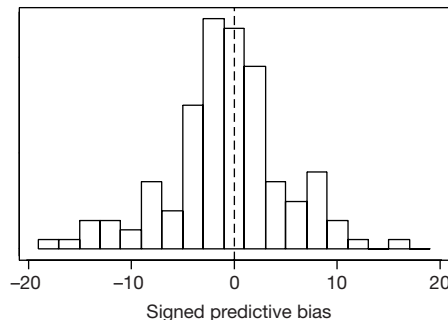


Figure 2 Signed predictive bias in projected population size (compared with actual population numbers), taken across 21 populations for five PVA packages. There is no significant positive or negative bias.

between particular packages for individual species (on the basis of Kolmogorov–Smirnov confidence interval tests²² on quasi-extinction probabilities). However, there was no overall consistency in these differences, and no package gave demonstratively better or worse predictions than any other. RAMAS Metapop and VORTEX gave the best least-squares model fit to actual population size (taken across all 21 species; see Table 3 of Supplementary Information), although a Kruskal-Wallis rank test showed no significant difference among the PVA packages ($H = 7.9$, d.f. = 4, $P = 0.097$).

The short-term predictions of the PVA models were relatively accurate, with a good overall correspondence between simulated and observed outcomes, on the basis of several criteria. The PVA predictions were not overly optimistic, as has been assumed by some¹⁰ (in fact, the general trend was for the PVA models to be slightly pessimistic). Following the precautionary principle, it is preferable to produce circumspect predictions of risk when managing endangered species²³. These findings contradict the widespread view that stochastic population models are poor predictors of a population’s future fate. Nevertheless, if the means and/or variances of vital rates were to change substantially in the future, the absolute accuracy of PVA predictions would be questionable. Furthermore, these results may not apply to plants (which were not included in this study), nor to cases where too few life-history or population data are available to estimate model structure and input parameters reliably¹³.

In some cases, there were significant absolute differences between the predictions of different software packages, due to differences in model structure. For example, a ceiling carrying capacity had to be implemented for the cycling Soay and Boreray sheep populations in INMAT, even though this was known to be inappropriate for these populations²⁴. In these cases, the other packages that allowed over-compensatory density-dependent survival to be modelled gave more realistic predictions than INMAT. For a given species, the best packages are likely to be those that most realistically model its life-history. Unfortunately, it is not always possible to discern the type of model that fits a particular case *a priori*. Despite the best efforts of PVA modellers, it is possible that some events (such as catastrophes) may be overlooked, making it impossible to guarantee

Table 2 Correlations between the predictions of five PVA packages

	Probability of decline*				Model fit to actual N_t †			
	INMAT	R META	R STAGE	VORTEX	INMAT	R META	R STAGE	VORTEX
GAPPS	0.83	0.78	0.86	0.86	0.66	0.94	0.88	0.90
INMAT		0.63	0.87	0.76		0.70	0.82	0.78
R META			0.83	0.87			0.94	0.89
R STAGE				0.88				0.89

* Probability of the maximum observed historical population decline.
 † Scaled fit of model projections to actual population size (N_t).
 All correlations are highly significant ($P < 0.01$).

that incorrect predictions will not be made. It is therefore important to explore alternative model structures and conduct sensitivity analyses, particularly when data are scarce^{2,3,14}. Moreover, information collected from past monitoring should be used to test and refine PVA models, forming a continual feedback process of development and improvement²⁵. PVA is the best tool we have for estimating extinction risk, and the alternatives are subjective, less rigorous, and likely to provide poorer predictions²⁶.

PVA predictions are surprisingly accurate, given adequate data, and should be useful in the conservation contexts in which they are currently applied. There are also high correlations between the predictions of different PVA packages. These results provide strong empirical justification for the use of PVA for categorizing the vulnerability of endangered animal species and evaluating options for their recovery. Furthermore, they validate PVA as a useful research tool for addressing unresolved issues in conservation biology. □

Methods

Data sets

The protocol for choosing examples was independent of their structure, detail and outcome. The criteria used were: (1) a minimum duration of 10 years; (2) data of sufficient quality and detail to build PVA models; (3) a small population size (very large populations are generally not of conservation concern); (4) data concerning endangered species or isolated populations. Thus, the filter was based on the priorities of our colleagues in collecting ecological data. The 21 populations that we were able to use (see Supplementary Information) covered a range of taxa (birds, mammals, reptiles and fish), tropic levels (omnivores, herbivores and carnivores), environmental conditions (low to high environmental variability) and insular and mainland populations, encompassed several different modes of population regulation (exponential growth or decline, a population ceiling, and density dependence), and spanned a range of geographical and climatic zones (Africa, Europe, North America and Oceania). Nevertheless, not all taxa are completely or proportionally represented in our sample, and the focus is on extant species.

PVA model structure

The structure of each PVA model depended on the biology of the given species and the built-in features of each software package (see Supplementary Information). To ensure that the models were as realistic as possible, all relevant aspects of the population's ecology were included (within the capabilities of each PVA package). The exact structure of the PVA models built using different packages often differed for a given species, depending on the available features, limitations and assumptions associated with each program²⁷. This made it possible to test whether some packages were better predictors than others because of their different features.

Parameter estimation

We used the best estimates for each parameter, given the available demographic, environmental and population data. Where the estimated population parameters were given directly in the literature, we checked them independently when possible. In most cases the compilation and analysis of data were done in collaboration with the people who worked on the taxon in question. The protocols used to estimate the PVA model parameters are presented in the Supplementary Information.

PVA runs

Five hundred stochastic simulation runs were performed in each case to ensure statistical reliability.

Testing PVA predictions

The predictive accuracy of PVA was assessed as follows. (1) Quasi-extinction probabilities were used as a surrogate for absolute extinction risk, as only one study population went extinct. If the PVA predictions are correct, the proportion of historical populations that declined below a given threshold should equate with the proportion of simulation trajectories that declined below the same threshold. To test this, the number of actual populations (out of 21) that declined below the threshold population sizes corresponding to predicted quasi-extinction probabilities of 0%, 10%, 20%, ..., 100% were recorded. (2) Mean population size predictions and growth rates were compared with reality using a combined test based on the sum of standard deviates²⁸. The objective was to test for equality over multiple populations, and combine the results of individual paired tests to increase statistical power. A log-transformation was used to normalize population size data. (3) Signed predictive bias (mean error): models with good predictive properties will produce a bias close to zero. Bias was calculated as $\Sigma(\text{Predicted size} - \text{Actual size}) / \text{number of years } (N)$. This was converted to a normal deviate $W = \sqrt{N} \times \text{bias}/s$, where s is the standard deviation of bias values. (4) The 'fit' of predicted versus actual population size over time was used to assess projections of future population numbers, calculated as the sum of squared deviations between observed and actual population sizes $(\Sigma[\text{Predicted} - \text{Actual}]^2 / \text{Actual})$ for each simulation replicate, averaged over 500 repli-

cates. (5) Pearson's product moment correlations were used to determine concordance amongst the predictions of different PVA packages for quasi-extinction risk and population size projections.

Statistical assumptions

This study represents a general 'meta-analysis' problem²⁹, the units of analysis being independent studies rather than individual subjects. The statistical population was composed of all studied biological populations with sufficient data, and was sampled completely. The underlying assumption of all our statistical tests on the combined results was that the 'errors' were independent among the 21 cases²⁷. As the case studies had little overlap in terms of place, time-period or researcher, this is a valid assumption. The quasi-extinction analysis was based on a simple goodness of fit test, carrying no assumption about the distributions of individual cases. For tests comparing different PVA packages, we verified that the assumptions of parametric tests^{13,14} were met.

Received 7 September 1999; accepted 4 February 2000.

1. Boyce, M. S. Population viability analysis. *Annu. Rev. Ecol. Syst.* **23**, 481–506 (1992).
2. Beissinger, S. R. & Westphal, M. I. On the use of demographic models of population viability in endangered species management. *J. Wildl. Mgmt.* **62**, 821–841 (1998).
3. Groom, M. J. & Pascual, M. A. in *Conservation Biology* (eds Fiedler, P. L. & Kareiva, P. M.) 4–27 (Chapman & Hall, New York, 1998).
4. Norton, T. W. (ed.) Special issue: applications of population viability analysis to biodiversity conservation. *Biol. Conserv.* **73**, 91–176 (1995).
5. International Union for the Conservation of Nature. *IUCN Red List Categories* (Gland, Switzerland, 1994).
6. Caughley, G. Directions in conservation biology. *J. Anim. Ecol.* **63**, 215–244 (1994).
7. Taylor, B. L. The reliability of using population viability analysis for risk classification of species. *Conserv. Biol.* **9**, 551–558 (1995).
8. Ludwig, D. Is it meaningful to estimate a probability of extinction? *Ecology* **80**, 298–310 (1999).
9. Possingham, H. P., Lindenmayer, D. B. & Norton, T. W. A framework for the improvement of threatened species based on population viability analysis (PVA). *Pac. Conserv. Biol.* **1**, 38–45 (1993).
10. Ralls, K. & Taylor, B. L. in *The Ecological Basis of Conservation: Heterogeneity, Ecosystems, and Biodiversity* (eds Pickett, S. B. A., Ostfeld, R. S., Shachak, M. & Likens, G. E.) 228–235 (Chapman & Hall, New York, 1997).
11. Talbot, J. *Land and Environment Court of New South Wales: Record of Hearing #10151 of 1994* (Court Report, Sydney, 1994).
12. Soulé, M. E. & Kohm, K. A. *Research Priorities for Conservation Biology* (Island, Washington DC, 1989).
13. Brook, B. W., Lim, L., Harden, R. & Frankham, R. Does population viability analysis software predict the behaviour of real populations? A retrospective study on the Lord Howe Island woodhen *Tricholimnas sylvestris* (Slater). *Biol. Conserv.* **82**, 119–128 (1997).
14. Brook, B. W., Cannon, J. R., Lacy, R. C., Mirande, C. & Frankham, R. A comparison of the population viability analysis packages GAPPs, INMAT, RAMAS and VORTEX for the Whooping crane (*Grus americana*). *Anim. Conserv.* **2**, 23–31 (1999).
15. Wigley, T. M. L. Climate change: a successful prediction? *Nature* **376**, 463–464 (1995).
16. Ramanathan, R. *Introductory Econometrics with Applications* 2nd edn (Dryden, Fort Worth, 1995).
17. Oreskes, N., Shrader-Frechette, K. & Belitz, K. Verification, validation, and the confirmation of numerical models in the earth sciences. *Science* **263**, 641–646 (1994).
18. Wahlberg, N., Moilanen, A. & Hanski, I. Predicting the occurrence of endangered species in fragmented landscapes. *Science* **273**, 1536–1538 (1996).
19. Huggard, D. J. A linear programming model of herbivore foraging: imprecise, yet successful? *Oecologia* **100**, 470–474 (1994).
20. Belovsky, G. E. How good must models and data be in ecology? *Oecologia* **100**, 475–480 (1994).
21. Ginzburg, L. R., Slobodkin, L. B., Johnson, K. & Bindman, A. G. Quasi-extinction probabilities as a measure of impact on population growth. *Risk Anal.* **2**, 171–182 (1982).
22. Sokal, R. R. & Rohlf, F. J. *Biometry* 4th edn (Freeman, New York, 1995).
23. Akçakaya, H. R., Burgman, M. A. & Ginzburg, L. R. *Applied Population Ecology* 2nd edn (Sinauer, Sunderland, Massachusetts, 1999).
24. Grenfell, B. T., Price, O. F., Albon, S. D. & Clutton-Brock, T. H. Overcompensation and population cycles in an ungulate. *Nature* **355**, 823–826 (1992).
25. Burgman, M. A., Ferson, S. & Akçakaya, H. R. *Risk Assessment in Conservation Biology* (Chapman & Hall, London, 1993).
26. Zeckhauser, R. J. & Viscusi, W. K. Risk within reason. *Science* **248**, 559–564 (1990).
27. Mills, L. S. *et al.* Factors leading to different viability predictions for a grizzly bear data set. *Conserv. Biol.* **10**, 863–873 (1996).
28. Reynolds, M. R., Burkhart, H. E. & Daniels, R. F. Procedures for statistical validation of stochastic simulation models. *Forest Sci.* **27**, 349–364 (1981).
29. Arqvist, G. & Wooster, D. Meta-analysis: synthesizing research findings in ecology and evolution. *Trends Ecol. Evol.* **10**, 236–240 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank K. Armitage, M. Boyce, J. Cannon, C. Catterall, T. Clutton-Brock, M. Curry, B. Grenfell, R. Harden, J. Kikkawa, M. Lynch, C. Mirande, R. Peterson, C. Schwartz, P. Smith, J. Vucetich and the Interagency Grizzly Bear Study Team for supplying data, and S. Andelman, J. Ballou, J. Bull, S. Ferson, I. Hanski, R. Harris, J. Kikkawa, R. Lacy, H. McCallum, M. McCarthy, S. Mills, P. Miller, S. Pimm, H. Regan and M. Soulé for comments on the manuscript. This study was funded by Australian Research Council and Macquarie University grants.

Correspondence and requests for materials should be addressed to B.W.B. (e-mail: barry.brook@ntu.edu.au).

