Conservation Biology

Contributed Paper

Testing Decision Rules for Categorizing Species' Extinction Risk to Help Develop Quantitative Listing Criteria for the U.S. Endangered Species Act

TRACEY J. REGAN,*† BARBARA L. TAYLOR,† GRANT G. THOMPSON,‡ JEAN FITTS COCHRANE,§ §§ KATHERINE RALLS,** MICHAEL C. RUNGE,†† AND RICHARD MERRICK‡‡

*The School of Botany, The University of Melbourne, Parkville, Victoria 3010, Australia, email tregan@unimelb.edu.au
†Southwest Fisheries Science Center, National Marine Fisheries Service, 8604 La Jolla Shores Drive, La Jolla, CA 92037, U.S.A.
‡Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, National Marine Fisheries Service, NOAA, 7600 Sand Point Way NE, Seattle, WA 98115-6349, U.S.A.
§U.S. Fish and Wildlife Service, Endangered Species Program, 4401 N. Fairfax Drive, Arlington, VA 22203, U.S.A.
**Smithsonian Conservation Biology Institute, National Zoological Park, Washington, D.C. 20008, U.S.A.
††U.S. Geological Survey, Patuxent Wildlife Research Center, 12100 Beech Forest Road, Laurel, MD 20708, U.S.A.
‡‡Office of the Assistant Administrator, National Marine Fisheries Service, 1315 East West Highway, Silver Spring, MD 20910, U.S.A.

Abstract: *Lack of guidance for interpreting the definitions of endangered and threatened in the U.S. Endangered Species Act (ESA) has resulted in case-by-case decision making leaving the process vulnerable to being considered arbitrary or capricious. Adopting quantitative decision rules would remedy this but requires the agency to specify the relative urgency concerning extinction events over time, cutoff risk values corresponding to different levels of protection, and the importance given to different types of listing errors. We tested the performance of 3 sets of decision rules that use alternative functions for weighting the relative urgency of future extinction events: a threshold rule set, which uses a decision rule of x% probability of extinction over y years; a concave rule set, where the relative importance of future extinction events declines exponentially over time; and a shoulder rule set that uses a sigmoid shape function, where relative importance declines slowly at first and then more rapidly. We obtained decision cutoffs by interviewing several biologists and then emulated the listing process with simulations that covered a range of extinction risks typical of ESA listing decisions. We evaluated performance of the decision rules under different data quantities and qualities on the basis of the relative importance of misclassification errors. Although there was little difference between the performance of alternative decision rules for correct listings, the distribution of misclassifications differed depending on the function used. Misclassifications for the threshold and concave listing criteria resulted in more overprotection errors, particularly as uncertainty increased, whereas errors for the shoulder listing criteria were more symmetrical. We developed and tested the framework for quantitative decision rules for listing species under the U.S. ESA. If policy values can be agreed on, use of this framework would improve the implementation of the ESA by increasing transparency and consistency.*

Keywords: Bayesian analysis, loss functions, performance testing, population viability analysis

Evaluando Reglas de Decisión para Categorizar el Riesgo de Extinción de Especies con el Fin de Desarrollar de Criterios Cuantitativos de Alistamiento en el Acta de Especies en Peligro de los EE. UU.

Resumen: *La falta de orientación para interpretar las definiciones de en peligro y amenazada en el Acta de Especies en Peligro de E.U.A. ha resultado en la toma de decisiones caso por caso, con lo cual el proceso es vulnerable para ser considerado arbitrario o caprichoso. La adopción de reglas de decisión cuantitativas podría remediar esta situación pero requiere la especificación por parte de la agencia de la urgencia relativa*

§§*Current address: 29 Wood Mountain Road, Grand Marais, MN 55604, U.S.A.*
Paper submitted June 6, 2012; revised manuscript accepted November 27, 2012.

Conservation Biology, Volume xx, No. x, 1–11

DOI: 10.1111/cobi.12055

concerniente a eventos de extinción en el tiempo, los valores de corte de riesgo correspondientes a diferentes niveles de protección y la importancia otorgada a diferentes tipos de errores de enlistado. Probamos el funcionamiento de 3 conjuntos de reglas de decisión que usan funciones alternativas para ponderar la urgencia relativa de eventos de extinción futuros: un conjunto de reglas umbral, que utiliza una regla de decisión de x% de probabilidad de extinción en y años; un conjunto de reglas cóncavo, en el que la importancia relativa de los eventos de extinción futuros declina exponencialmente en el tiempo; y un conjunto de reglas que utiliza una función de forma sigmoidea, en donde la importancia relativa declina lentamente al principio y más rápido posteriormente. Obtuvimos valores de corte de decisiones mediante entrevistas con varios biólogos y luego emulamos el proceso de enlistado con simulaciones que cubrieron un rango de riesgos de extinción típico de las decisiones de enlistado del AEP. Evaluamos el funcionamiento de las reglas de decisión bajo cantidades y calidades diferentes de datos con base en la importancia relativa de los errores de clasificación. Aunque hubo poca diferencia entre el funcionamiento de las reglas de decisión alternativas para enlistados correctos, la distribución de errores de clasificación difirió dependiendo de la función utilizada. Los errores de clasificación para los criterios umbral y cóncavo resultaron en errores de sobreprotección, particularmente a medida que incrementó la incertidumbre, mientras que los errores de la función sigmoidea fueron más simétricos. Desarrollamos y probamos un marco de referencia para reglas de decisión cuantitativas para enlistar especies en el Acta de Especies en Peligro de los EE. UU. Si se logran acuerdos sobre los valores para establecer políticas, el uso de este marco de referencia podría mejorar la implementación del AEP al incrementar la transparencia y consistencia.

Palabras Clave: análisis bayesiano, análisis de viabilidad poblacional, pruebas de rendimiento

Introduction

The U.S. Endangered Species Act (ESA) of 1973 is one of the most powerful and influential environmental laws in the United States (Bean 2009). Under the ESA a species can be categorized as endangered (in danger of extinction throughout all or a significant portion of its range), threatened (likely to become endangered in the foreseeable future), or not warranted (16 U.S.C. §§ 1532). The conservation and economic consequences of the law's protections have motivated litigation challenging many listing decisions (Doremus 2006). Such litigation can result in substantial costs to the government agencies that implement the ESA and reduce the amount of funding available to mitigate threats to endangered species or to assess species that may warrant protection (Miller et al. 2002; Wilcove & Master 2005).

The ESA requires that listing decisions be based solely on the best scientific and commercial data available to avoid any social or economic implications influencing the decision (Bean 2009). Science can estimate the probability that a species will become extinct over time, but a listing decision also requires value judgments about the level of extinction risk that warrants protection under the act (Doremus 1997). Currently, there are no policy guidelines for interpreting the vague definitions of *endangered* or *threatened* or on how much extinction risk warrants protection, and agency decision makers make these value judgments on a case-by-case basis. This element of the listing process lacks transparency and consistency and is the reason some consider the process arbitrary or capricious (Robbins 2009).

An alternative to the current approach is to devise quantitative decision rules on which to base the risk of extinction that could be applied to all species considered for listing, but this approach would require being explicit about the value judgments regarding the timing and amount of extinction risk that would trigger listing. Several quantitative thresholds have been suggested for acceptable levels of extinction risk within particular time frames, including endangered if the probability of extinction is greater $\geq$1% in 1000 years (Shaffer 1981), $\geq$5% in 100 years (Shaffer 1983), $\geq$1% in 100 years (Angliss et al. 2001), $\geq$20% within 20 years or 5 generations (IUCN 2001), and $\geq$15% in 100 years (Goodman 2002*a*). However, none of these thresholds have been formally adopted into ESA policy.

The primary goal of the ESA is to protect species from extinction (Goble et al. 2006). It is impractical to list all species; thus, an operational goal of the ESA is to identify those species more likely to go extinct in the near future than species likely to go extinct in the distant future. Decision rules such as "endangered if the probability of extinction is $>x$% in y years" can capture this urgency, but it is unclear whether the sharp time threshold of y years fully represents how society weighs the importance of future extinction events. In practical terms, this type of decision rule implies that all possible extinction events up until the time threshold of y years have the same level of urgency. After this time threshold, any likelihood of extinction is not considered. This could be problematic in identifying borderline cases or identifying species where the effect of a threat has a long time lag.

In other disciplines, such as decision analysis and economics, weighting the relative importance of uncertain future events is commonplace and is known as time preferences, or time discounting (Price 1993; Frederick et al. 2002). Rather than using sharp time thresholds, the

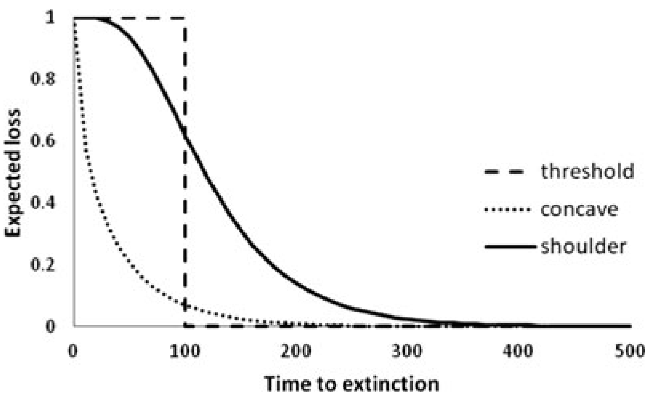

Figure 1. Expected loss of future extinction events as a function of time to extinction (in years) for a threshold function (dashed line, relative importance of future extinction events equal until time threshold, then drop to zero); concave function (dotted line, relative importance of future extinction events declines exponentially over time); and shoulder function (solid line, sigmoid-shape function for which relative importance declines slowly at first and then more rapidly later). Expected loss is a relative measure between zero and one and time to extinction is in years.

importance of future events are weighted according to a discount rate that captures the relative importance of events that occur in the near, medium, and far term. Although any extinction event is a tragedy, time preferences could be applied to decision rules for listing species, could be used to identify those species in more urgent need of protection, and may better capture society's concern about future extinction events than a sharp time threshold.

We explored how future time preferences can be scaled in a rational way to devise decision rules for listing species under the ESA. In particular, we investigated the properties and performance of 3 alternative sets of decision rules devised from different functions for weighting the relative importance of future extinction events: a threshold rule set, in which a decision rule of x% probability of extinction over y years is applied; a concave rule set, in which the relative importance of future extinction events declines exponentially over time; and a shoulder rule set, which assumes a sigmoid shape function and relative importance of future extinction events declines slowly at first and then more rapidly over time (Fig. 1). To define the level of risk associated with categories such as endangered and threatened, we conducted an elicitation exercise with a number of scientists to generate hypothetical values for these levels for testing purposes. Although adoption of decision rules for listing species under the ESA requires government agencies to adopt policy parameters regarding the acceptability of different levels of extinction risk, our objective was not to determine policy values for risk thresholds or specific time preferences, but to determine possible implications and consequences of different ways of estimating the relative importance of future extinction events.

Ultimately the goal of the ESA is to list species that warrant protection and not to list species that do not need protection. Thus, a good decision rule for listing species under the ESA would rank species according to extinction risk and minimize misclassifications. This is difficult because humans cannot predict intuitively the behavior of complex systems with interacting components, and uncertainty and incomplete information hinder one's ability to understand the true status of the system of interest. Quantitative listing criteria involve 3 main interactive components: a function that represents the relative importance of extinction events over time; decision cutoff values that correspond to different levels of protection (i.e., endangered and threatened); and the weight given to different types of errors (e.g., listing a species that does not warrant protection and failing to list a species that requires protection). We investigated these interactive components for each of the alternative sets of decision rules through computer simulations to help policy makers better understand the behavior of different sets of decision rules and how they perform under uncertainty.

Exploring the behavior of alternative decision rules with computer simulations is helpful because it forces transparency in representation of a system. It allows a comprehensive investigation of the effect of different levels of uncertainty inherent in the decision-making process by quantifying their effects on the performance of alternative decision options (Harwood & Stokes 2003). We refer to this method as performance testing. Performance testing has been used to evaluate alternative management strategies in fisheries management (Cooke 1999; Punt & Smith 1999), develop management procedures for marine mammals (Cooke 1995; Taylor et al. 2000), evaluate model uncertainty when classifying species at risk (Taylor 1995), and to test the effect of observation errors on extinction-risk estimates (Meir & Fagan 2000; Taylor et al. 2002).

To evaluate performance of the 3 alternative sets of decisions rules for ESA listing, we simulated the underlying population dynamics for multiple species, assuming perfect biological information, to generate the probability distribution of time to extinction given only environmental variation (i.e., the true probability of extinction). We then simulated various schemes for data collection and estimated the probability of extinction. Applying the 3 listing criteria to both the true and estimated fate of species allowed us to evaluate the alternative decision rules by comparing how well they correctly classified species under uncertainty and by evaluating the consequences of different types of misclassifications (i.e., listing a species that does not warrant protection and failing to list a species that requires protection).

Methods

Alternative Decision Rules

We refer to the function that represents the relative importance of future extinction events as a loss function, represented as the expected loss as a function of time (Fig. 1). A loss function associated with extinction is a value judgment that represents how society's concern about species extinction changes as the projected time of extinction becomes more distant. Estimating a loss function requires formally eliciting value judgments regarding extinction and its timing. For this study, we used an average loss function we derived from an elicitation exercise focused on how biologists judge species endangerment. We refer to this loss function as the shoulder function (Regan et al. 2009; Cochrane et al. 2011). To investigate a more contrasting form of a loss function, we chose a concave-shaped loss function, recommended by the U.S. Congressional Budget Office for discounting environmental and natural resources (see Supporting Information for details of the parameterization on the loss functions). For the threshold rules set, we used a time horizon of 100 years and 150 years for endangered and threatened, respectively. The 3 loss functions (threshold, concave, and shoulder) captured a range of forms that encapsulated various value judgments regarding extinction and revealed alternative behavior for each of the decision rules (Fig. 1). We used the same loss function for all species to ensure equal treatment across species. Use of a different loss function for different species implies the extinction of one species is deemed more or less important than another.

Hypothetical policy values for decision cutoffs were determined through an elicitation exercise with 8 scientists familiar with the ESA listing process (Regan et al. 2009). The objective of the exercise was to identify a set of species on which the scientists agreed with the assigned threat categories (hereafter consensus species). Each participant was asked to categorize 20 species as either endangered, threatened, or not warranted for listing. The species included birds, reptiles, plants, mammals, and fish with different life histories and a nonzero probability of extinction within the next 500 years. Decision cutoffs for each of the alternative listing criteria were derived by ensuring all consensus species were placed in the agreed-upon threat category (see Supporting Information for details of the elicitation process and examples of the information given to participants). The performance-testing method described hereafter does not specifically follow a formal statistical decision analysis (Berger 1985), although we adopted some of the concepts of statistical decision analysis for use in our decision rules, such as the use of loss functions and expected loss.

Performance Testing

We simulated making listing decisions with perfect information (i.e., known functional form for the population model and known values for all parameters thereof) under each of the listing criteria and compared these decisions with listing decisions made with uncertain information (i.e., known functional form for the population model, but uncertain parameter values) (Fig. 2). This process involved creating an assumed reality and then simulating the underlying biological processes of interest: the probability of extinction through time. We simulated the set of consensus species and a set of cases that were configured so the extinction risk would be near the category boundaries (i.e., challenge cases). These challenge cases were constructed to be both plausible and likely to highlight performance differences among the alternative listing criteria. Details of the population-dynamics model are in Supporting Information.

We used a virtual biologist (Fig. 2) who did not know the true values of the model parameters for the species. This biologist collected and analyzed data and made inferences on those parameters. We used 4 scenarios for data collection in the performance testing. Each scenario had data of different quantity and quality so we could determine how the alternative decision rules performed under different levels of uncertainty. The data have 2 types of error, process error (i.e., environmental stochasticity) and observation error (i.e., error due to random sampling strategy). The data-gathering scenarios were data of high quantity and quality (20 years of annual abundance estimates with a coefficient of variation [CV] in observation error of 0.1) (hereafter high and high); few data of high quality (4 abundance estimates over the last 10 years with $CV = 0.1$) (hereafter low and high); data of high quantity and low quality (20 years of annual abundance estimates with $CV = 0.8$) (hereafter high and low); and data of low quantity and quality (4 abundance estimates over the last 10 years with $CV = 0.8$) (hereafter low and low). For each species, a trajectory was generated from the population model that was conditional on the true parameter values and that would result in an ending population size that matched the true population size. Lognormal distributed observation error was then applied to the time-series data. We assumed the distribution from which individual observation errors were drawn had a known coefficient of variation estimated from abundance surveys. The data-generating procedure was repeated 100 times for each of the consensus species and 5000 times for the challenge species for each of the data-generation scenarios.

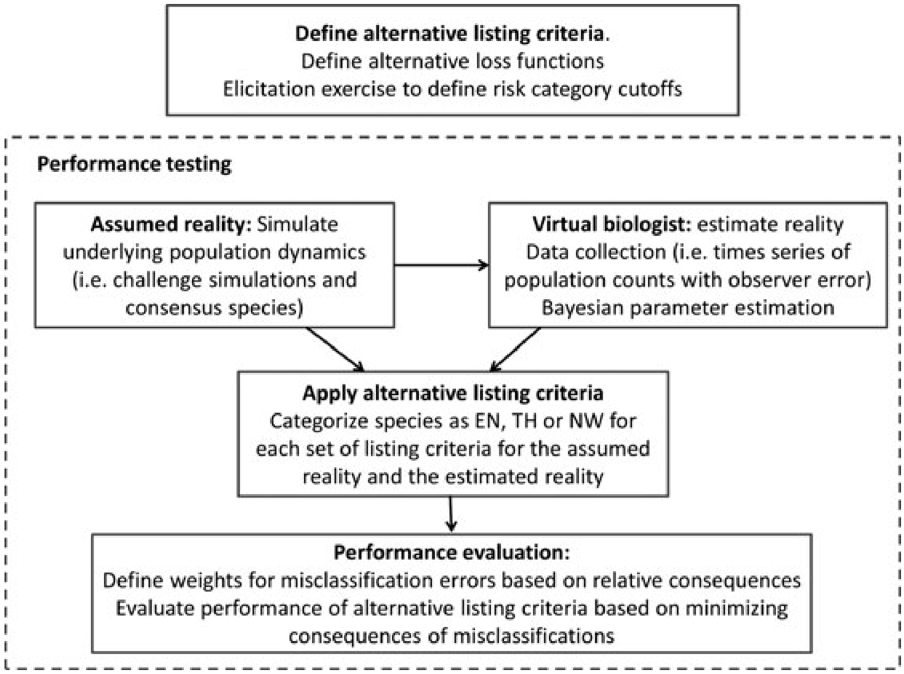

Figure 2. Framework for performance testing of alternative listing criteria, including definitions of alternative listing criteria; performance-testing algorithm (dashed box); the assumed reality of the underlying population dynamics of species; the virtual biologist, who collects and analyzes data and estimates model parameters; application of the alternative listing criteria to categorize species as endangered (EN), threatened (TH), and not warranted (NW); and the performance evaluation that is based on minimizing misclassifications.

Parameter Estimation

The population-dynamics model for each species had 3 parameters: mean growth rate (r), standard deviation of the process error (σ_p), and the current population size (N_0). These were estimated with Bayesian analysis by defining a prior for each of the parameters and a likelihood function. The probability distributions that were defined for the challenge cases were used as prior distributions for estimation. Broad, vague priors, for which the likelihood was non-negligible, were used for the consensus species. The data to estimate the parameters had both process error and observation error. To account for these 2 types of error we derived a likelihood function with a Kalman filter, an approximation of the fully state-spaced models that are typically used for parameter estimation when these 2 types of errors are present (Kalman 1960; Meinhold & Singpurwalla 1983). Supporting Information contains details and derivation of the likelihood function. We used Bayesian inference software (D. Goodman, see http://www.esg.montana.edu/) to estimate each of the model parameters. The program manages Bayesian inference for a small number of parameters based on an algorithm that samples the prior and then weights each sampled set of values of parameters by the likelihood, cumulating histograms, and posterior summaries of the sampled parameter values weighted accordingly.

We used a Bayesian population viability analysis (PVA) and calculated a derived parameter of the time to extinction from the weighted set of sampled values: r and σ_p and N_0. Results were summarized in a probability distribution of time to extinction and incorporated all the parameter uncertainty reflected in the data. The risk of extinction given each loss function was calculated by integrating through time the product of the probability of extinction and the loss. Risks of extinction for the concave and shoulder functions are unitless because the risk is integrated over all time horizons (0–∞) (Supporting Information). We then applied the alternative decision rules to the true and estimated extinction risks to determine the corresponding threat categories. We developed several subroutines in Fortran 90 to automate the entire performance-testing process.

Performance Evaluation

Performance of the alternative decision rules was evaluated by comparing the proportion of correct and incorrect listing decisions, whether the misclassifications were over or under protection errors, and the magnitude of the differences. To determine how value judgments affected the performance of decision rules, we used misclassification costs to weight different errors. These costs were weights that denoted the relative degree of concern with misclassification errors, not a literal or specific monetary cost.

We examined 4 possible weighting systems. In the first we assumed all misclassifications had equal weighting, which means either the costs were equal or ignored. In the second, we assumed a 2-category misclassification (i.e., assessed as not warranted when truly endangered) had twice the cost of a 1-category misclassification (i.e., assessed as not warranted when truly threatened). We referred to this weighting system as symmetrical weights because the weighting is the same whether the misclassification is over- or underprotection for the species. We also examined 2 precautionary weighting systems that penalized misclassification asymmetrically by preferring

Table 1. Hypothetical policy values of extinction risk for classifying species as endangered, threatened, and not warranted for the 3 alternative listing criteria.

Listing category	Alternative listing criteria* Threshold	Shoulder	Concave
Endangered	probability of extinction $\geq$0.30 in 100 years (0.18–0.37)	extinction risk $\geq$0.54 (0.50–0.54)	extinction risk $\geq$0.05 (0.05–0.06)
Threatened	probability of extinction $\geq$0.08 in 150 years (0.001–0.100)	extinction risk $\geq$0.18 (0.02–0.18)	extinction risk $\geq$0.01 (0.001–0.010)
Not warranted	probability of extinction <0.08 in 150 years	extinction risk <0.18	extinction risk <0.01

**Threshold, shoulder, and concave refer to the loss function for each listing criterion. Values in parentheses are ranges of extinction risk values that ensure the consensus species are categorized correctly. Consensus species are the set of species for which there is an agreed-upon threat category among a group of 8 scientists. The risks of extinction for the concave and shoulder functions are unitless because the risk is integrated over all time horizons (0–∞). See Eq. (1) in Supporting Information.*

overprotection to underprotection errors. The first was similar to the symmetrical weight system in that it considered 1-category versus 2-category misclassifications; however, underprotection errors incurred twice the cost of overprotection errors. We called the second precautionary system list versus not list. In this system, endangered and threatened were treated the same and weights were assigned only to misclassification between either of those categories and not warranted. Details of the weighting systems are in Supporting Information.

Results

The elicitation exercise resulted in 13 consensus species, 5 endangered, 5 threatened, and 3 not warranted (Supporting Information). The risks of extinction for these species ranged from 0.00 to 1.00 (Supporting Information). For all loss functions, a range of values allowed the consensus species to be categorized correctly (Table 1). The resulting hypothetical policy values for performance testing were those that maximized the number of identical listings across the 3 alternatives for the challenge simulations but were still within the range that would correctly categorize the consensus species (Table 1).

The performance testing for the consensus species and the challenge scenarios resulted in similar outcomes; thus, results presented here were based on a combined summary. We present results for some individual consensus species to highlight specific performance issues (full results in Regan et al. [2009]). The 3 alternative listing criteria produced roughly the same proportion of correct decisions (Fig. 3). The shoulder rule set had consistently higher proportions of correct decisions relative to the threshold and concave decision rules, but only marginally. When the quantity and quality of data were good (high-and-high data scenario), approximately 70% of species were classified correctly. The percentage of correctly classified species decreased as data availability and precision decreased. When the quality and quantity of data were poor (low-and-low data scenario), 40% of species were classified correctly for each of the alternative listing criteria. These percentages do not represent performance of actual listing cases; rather, they are challenge cases chosen to highlight the differences between the decision rules.

Although the proportions of correct decisions were similar across the 3 alternatives for the 4 data scenarios, the distribution of misclassifications differed depending on the rule set used. Misclassifications under the shoulder rule set tended to be symmetrical across the data scenarios, with roughly similar proportions of over- and underprotection errors, whereas the threshold and concave listing criteria tended to have more overprotection errors than underprotection errors. These trends became more prominent as data availability and precision decreased. For example, under the low-and-low data scenario, the threshold and concave rules had 87% and 92% of the misclassifications as overprotection errors, respectively, whereas 53% of the misclassifications for the shoulder rule set were overprotection errors. Most of the misclassifications were 1-category errors. For the high-and-high data scenario the 2-category errors ranged from 3% to 5% of the misclassifications across the 3 listing criteria, whereas for the low-and-low data scenario, the threshold and concave listing criteria had more 2-category errors (25% and 29%, respectively) than the shoulder rules (8%).

A similar pattern for the alternative listing criteria occurred in simulations for individual consensus species. We present results for 3 consensus species, passerine (endangered), pinniped 2 (threatened), and tortoise (not warranted) (Table 2), that represented the most challenging cases in which extinction risks were closest to the cutoffs between risk categories. The threshold and concave listing criteria were the most protective, which resulted in incorrectly listing the tortoise (not warranted) 49% and 39% of the time, respectively, even under the high-and-high data scenario. Under the low-and-low data scenario, the tortoise was correctly listed as not warranted 6% and 4% of the time for the threshold and concave

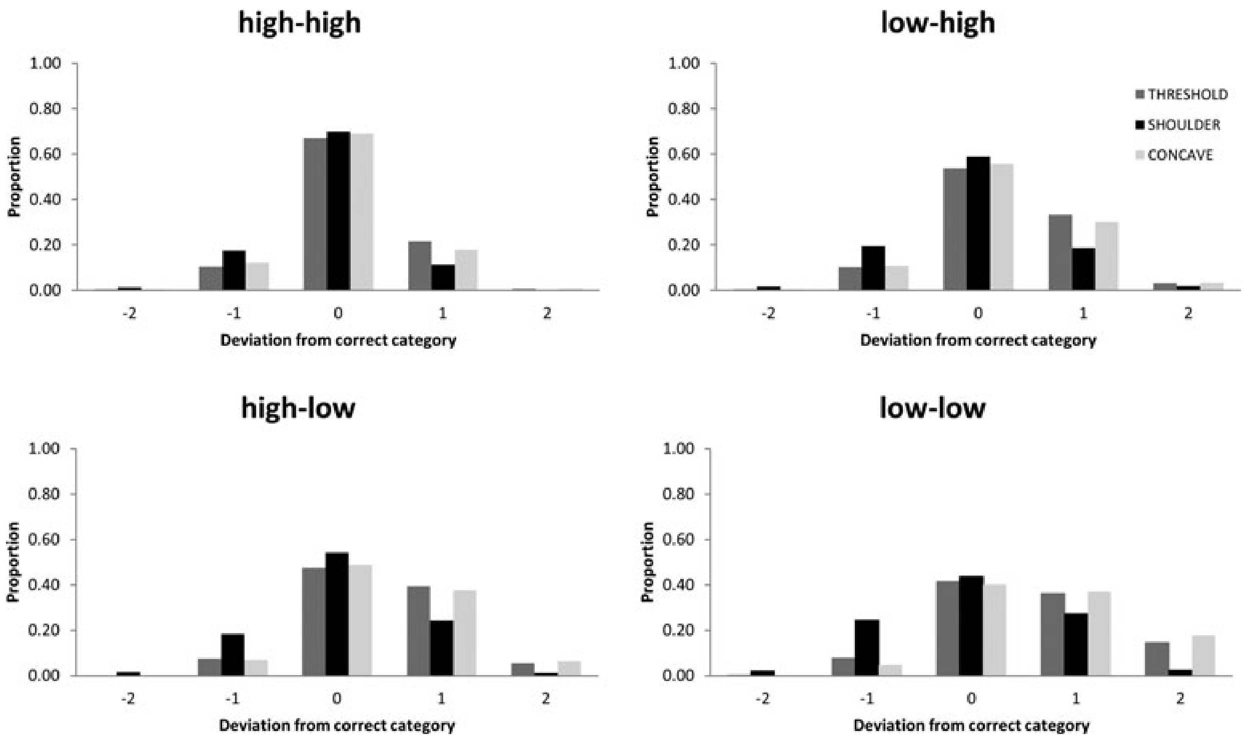

Figure 3. Proportion of correct and incorrect listing decisions on the basis of true and estimated extinction risks for the 3 alternative listing criteria for all species for different data scenarios (high–high, high quantity and quality; low–high, low quantity and high quality; high–low, high quantity and low quality; low–low, low quantity and quality). Zero deviations are correct decisions, positive deviations are overprotection errors, and negative deviations are underprotection errors. Threshold, shoulder, and concave are the alternative listing criteria that are based on the 3 loss functions investigated.

Table 2. Percentage of correct and incorrect listings for 3 consensus species[a] that represent the most challenging cases in which true extinction risks are closest to the decision cutoffs between threat categories[b] for 2 data scenarios.

Scenario species	*Threshold function*			*Shoulder function*			*Concave function*		
	EN	*TH*	*NW*	*EN*	*TH*	*NW*	*EN*	*TH*	*NW*
High–high[c]									
passerine	58[d]	46	6	43[d]	38	19	59[d]	35	6
pinniped 2	13	64[d]	23	4	51[d]	45	13	53[d]	34
tortoise	1	48	51[d]	0	21	79[d]	2	37	61[d]
Low–low									
passerine	48[d]	50	2	18[d]	57	25	66[d]	34	0
pinniped 2	36	62[d]	2	9	57[d]	34	49	50[d]	1
tortoise	21	73	6[d]	3	44	53[d]	27	69	4[d]

[a]*Consensus species are the set of species for which there is an agreed-upon threat category among a group of 8 scientists. Passerine (EN), pinniped 2 (TH) and tortoise (NW).*
[b]*Abbreviations: EN, endangered; TH, threatened; NW, not warranted.*
[c]*High–high, data of high quantity and quality; low–low, data of low quantity and quality.*
[d]*Percentage of correct listings.*

listing criteria, respectively. The threatened pinniped 2 had similar correct listings under the high-and-high and low-and-low data scenarios for the threshold and concave listing criteria, but there was a large shift in the errors from greater underprotection errors for the high-and-high data scenario to larger overprotection errors for the low-and-low data scenario. For the shoulder rule set, the error distribution was similar for both data scenarios. For the concave rule set, the endangered passerine was correctly categorized more often under the low-and-low data scenario than the high-and-high data scenario due to this rule's tendency to be more precautionary under uncertainty.

When the consequences of incorrect decisions were not specified (i.e., implying equal weighting for all misclassifications), the shoulder rule set resulted in the best net performance with minimum relative costs for all data scenarios (Fig. 4). The shoulder rule set also had the lowest relative costs when weights were higher for 2-category errors under the symmetrical weighting. However, when the consequences of misclassifications were weighted in a precautionary manner (i.e., precautionary and list vs. not-list weighting systems) the threshold and concave rules were optimal. This was apparent for all data scenarios. The difference between the relative costs for the concave and threshold rules was minimal for the precautionary and the list versus not-list weighting structures.

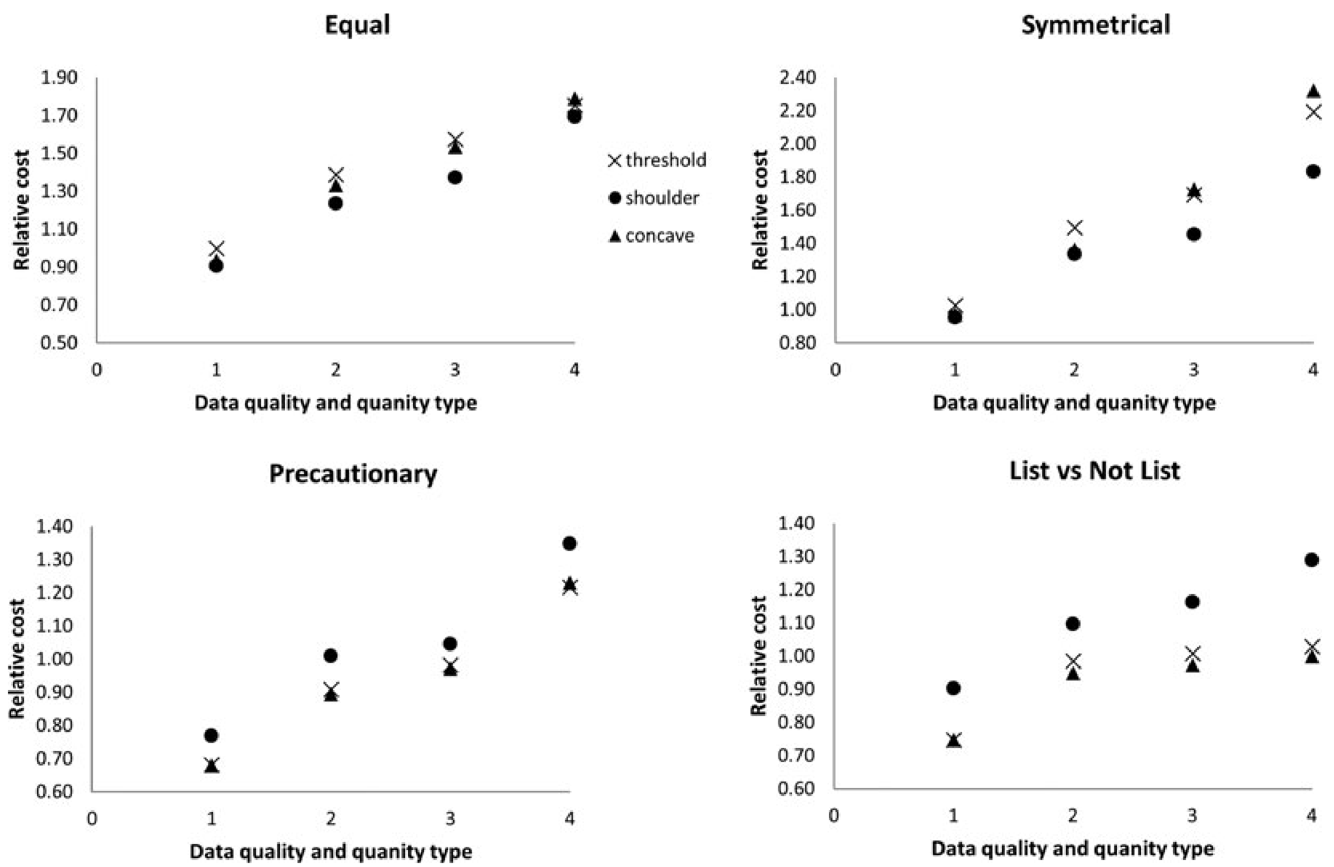

Figure 4. Relative costs of 4 weighting systems of listing misclassification (equal, costs of misclassification are equal or ignored; symmetrical, 2-category misclassification has twice the cost as a 1-category misclassification; precautionary, underprotection errors are twice as costly as overprotection errors; list versus not list, endangered and threatened are treated the same and weights are assigned only to misclassification of those categories and not warranted) for each of 3 alternative listing criteria under 4 alternative data-quality and data-quantity scenarios (1, high quantity and quality; 2, high quantity and low quality; 3, low quantity and high quality; 4, low quantity and quality). Within a weighting system the best performer has the lowest costs.

Discussion

Adopting a quantitative approach to listing species under the ESA that is more objective, consistent, and transparent may reduce controversies and delays arising with individual listing decisions, but such an approach requires advance policy decisions regarding the relative importance of future extinction events through specification of a loss function, explicit values for acceptable levels of extinction risk across species, and how misclassification errors are weighted. Our framework applies hypothetical policy values to an investigation of the performance of alternative sets of quantitative decision rules that can be used to list species under the ESA. The alternative listing criteria performed similarly with respect to the number of correct listings, even as uncertainty increased. This was an unanticipated result and may have been due in part to the careful design of the alternative decision rules, ensuring they were equivalent on the basis of a set of consensus species prior to performance testing. This is an encouraging result because the rate of correct listings was largely not affected by the form of the loss function, at least for the loss functions we examined.

The form of the loss function did have an effect on misclassification. The misclassifications for the shoulder function tended to be equally distributed between over- and underprotection errors even as uncertainty increased. Although this symmetry may be preferable to the threshold or concave functions, for which misclassifications were more skewed toward overprotection errors, it may become less desirable when the consequences of misclassifications are considered. These consequences include not identifying a species that is in need of protection and protecting a species that does not warrant it. The relative weighting of under- and overprotection errors is also a policy choice. Although the actual consequences of over- and underprotection errors associated with the listing process depends largely on the situation, we have provided some relative measures of the consequences so that the implications of a particular policy stance are clear. That is, when a precautionary approach to listing is preferred (i.e., erring on the side of overprotection rather than underprotection), then the threshold and concave rules sets are desirable. The shoulder rule set is more desirable if policy values favor either equal weight for both types of errors or favor underprotecting species.

The error rates for the 3 alternative decision rules were roughly 30% for the high-and-high data scenario and around 60% for the low-and-low data scenario. These error rates do not reflect actual ESA listing decisions because the species that were used to test the alternative decision rules were not actual species that were subject to petitions and status reviews and we constrained the performance to a set of challenge cases that would highlight performance close to the category boundaries. Furthermore, we only tested 4 data quality and quantity scenarios. In reality there would be a continuum of uncertainty in data quality and quantity. The actual error rate in listing decisions is impossible to know because the current listing process is done in a less systematic and transparent way. Consequently, the actual error rate could potentially be higher than we observed. Nevertheless, the magnitude of the errors, particularly for the low-and-low data scenario, may concern decision makers. The poor performance in classifying the low-and-low data scenarios was similar for decision rules that used near-extinction levels of 50 and 250 breeding individuals (Regan et al. 2009). This was somewhat surprising because these decision rules shortened the time horizon to 50 years for endangered and 100 years for threatened, which gives a shorter period for the uncertain growth rates to be projected into the future and potentially result in reaching the near-extinction threshold. Although no set of quantitative decision rules can fully compensate for information gaps, particularly when trends are completely unknown, important contributions of our performance-testing process presented include the insights into the relative magnitude and direction of misclassifications and the provision of a framework to investigate and provide recommendations for minimum amounts of data necessary for listing decisions. We suggest potential alternatives to making listing decisions for data-poor cases, such as prioritizing research funding so as to reduce uncertainty for cases with the highest risk as indicated by threats.

We estimated population parameters with a Kalman filter to account for both process error and observation error. This is not common practice when estimating parameters for PVA models. Instead, the observation error is often subsumed in the process-error term; thus, the amount of variation in environmental noise is overestimated. This will generally result is higher estimates of the probability of extinction and more overprotection errors (Morris & Doak 2002; Drake & Lodge 2004). Although this may also reduce the number of underprotection errors in some cases, we do not advocate ignoring observation error as a means of reducing classification errors. We used Bayesian estimation methods to account for parameter uncertainty so that the full extent of parameter uncertainty was propagated faithfully through to the posterior distribution of the derived parameter (probability of time to extinction). An advantage of this approach is that one can use prior information about the species to help improve precision of parameter estimates (McCarthy & Masters 2005). Although PVA models with posterior distributions are not common in the PVA literature, they have been advocated for classifying species because parameter uncertainty is directly incorporated into the probability distribution of time to extinction. Thus, one does not have to use point estimates and perform sensitivity analyses (Wade 1999; Goodman 2002*b*; Taylor et al. 2002).

We calculated the probability of extinction with respect to time in years rather than generations. Using years makes ESA implementation technically easier, but has the side effect of potentially treating long-lived species with less precaution. Conversely, one could argue that measuring time in units of generations rather than years could mean that long-lived species are valued more highly than short-lived ones, resulting in unequal treatment of species. O'Grady et al. (2008) reported that extinction risk, measured as a minimum viable population size, scaled better to generations than years. We report results using absolute extinction risk here, but results with near-extinction default values of 50 and 250 mature adults are given in Regan et al. (2009), which revealed similar patterns. However, use of a near-extinction value could result in different listing decisions for long-lived species that may remain extant at a few individuals for decades. It could also result in a process with shorter time horizons, which might lead to a better risk ranking of species than use of longer time scales that increase uncertainty in probability of extinction estimates (Feiberg & Ellner 2001).

Adopting quantitative decision rules for the ESA listing process would be a step forward in the conservation of threatened species. Choosing the most appropriate set of decision rules requires consideration of the loss incurred from possible extinction events, the relative consequences of misclassifications, performance under uncertainty, and ease of implementation. Although the use of a continuous loss function, like the shoulder or concave functions, may be a more realistic representation of the actual human perception of the loss incurred from possible extinction events, the threshold function may be easier to explain and implement because this form of decision rule is widely understood and accepted by conservation professionals. Our results provide a clearer understanding of the consequences of adopting one set of decision rules over another and is the first of many steps required to develop a more systematic listing process. We acknowledge that quantitative data to derive extinction risk estimates are not available for many species. The classification process will also need to include criteria of tested proxies that equate with the chosen quantitative decision rules. We anticipate that practical implementation of these rules would require detailed guidelines that would allow agency biologists to efficiently and effectively apply the quantitative definitions (standards) and their proxies across a great variety of species and data types. As a next step, we recommend use of quantitative decision rules in a retrospective analysis of currently listed species followed by the development of appropriate and well-tested proxy criteria that can serve as a substitute for these rules when there are insufficient data to estimate extinction risk directly. Such an analysis would complement and refine the research presented here and illustrate the feasibility of implementing quantitative decision rules under the ESA.

Acknowledgments

T.J.R. was jointly funded by the National Marine Fisheries Service and the U.S. Fish and Wildlife Service through a fellowship with the National Research Council and the Australian Research Council Centre of Excellence for Environmental Decisions. We thank L. Maguire, M. Nammack, T. Ragen, and S. Rumsey for expert advice; and A. Read, P. Boveng, P. Wade, and M. Burgman for their useful comments. We honour the passing of Dr. Dan Goodman who provided software and intellectual stimulation for this project. Dan devoted much of his career to integrating decision theory into conservation biology and his mentoring was very important to several authors of this work.

Supporting Information

Details of the estimation of the loss function (Appendix S1), description of elicitation exercise for determining decision cutoffs (Appendix S2), information for 3 species given to participants to categorize as endangered, threatened, or not warranted (Appendix S3), details of the population modeling used in performance testing (Appendix S4), description of the Kalman filter (Appendix S5), details of the weighting systems used for performance evaluation (Appendix S6) and results of the elicitation exercise (Appendix S7) are available online. The authors are responsible for the content and functionality of these materials. Queries (other than absence of the material) should be directed to the corresponding author.

Literature Cited

Angliss, R. P., G. K. Silber, and R. Merrick. 2001. Report of a workshop on developing recovery criteria for large whale species. Technical memorandum NMFS-F\OPR-21. National Oceanic and Atmospheric Administration, Silver Spring, Maryland.

Bean, M. J. 2009. The Endangered Species Act, science, policy, and politics. Annals of the New York Academy of Sciences **1162:**369–391.

Berger, J. O. 1985. Statistical decision theory and Bayesian analysis. Springer-Verlag, New York.

Cochrane, J. F., L. A. Maguire, G. G. Thompson, and T. J. Regan. 2011. How biologists judge species endangerment. Unpublished report to National Marine Fisheries Service, Protected Species Program, Silver Spring, Maryland; and the U.S. Fish and Wildlife Service. Endangered Species Program, Arlington, Virginia.

Cooke, J. G. 1995. The International Whaling Commission's revised management procedure as an example of a new approach to fishery management. Pages 647–657 in A. S. Blix, L. Walloe, and O. Ulltang, editors. Whales, seals, fish and man. Elsevier Science, Amsterdam.

Cooke, J. G. 1999. Improvement of fishery-management advice through simulation testing of harvest algorithms. ICES Journal of Marine Science **56:**797–810.

Doremus, H. 1997. Listing decision under the Endangered Species Act: why better science isn't always better policy. Washington University Law Quarterly **75:**1029–1153.

Doremus, H. 2006. Lessons learned. Pages 195–207 in D. D. Goble, J. M. Scott, and F. W. Davis, editors. The Endangered Species Act at thirty. Island Press, Washington, D.C.

Drake, J. M., and D. M. Lodge. 2004. Effects of environmental variation on extinction and establishment. Ecology Letters **7:**26–30.

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

Frederick, S., G. Loewenstein, and T. O'Donoghue. 2002. Time discounting and time preference: a critical review. Journal of Economic Literature **40:**351–401.

Goble, D. D., J. M. Scott, and F. W. Davis. 2006. The Endangered Species Act at thirty. Island Press, Washington, D.C.

Goodman, D. 2002*a*. Predictive Bayesian population viability analysis: a logic for listing criteria, delisting criteria and recovery plans. Pages 447–469 in S. R. Beissinger and D. R. McCullough, editors. Population viability analysis. University of Chicago, Chicago.

Goodman, D. 2002*b*. Uncertainty, risk, and decision: the PVA example. American Fisheries Society Symposium **27:**171–196.

Harwood, J., and K. Stokes. 2003. Coping with uncertainty in ecological advice: lessons from fisheries. Trends in Ecology & Evolution **18:**617–622.

IUCN (International Union for Conservation of Nature). 2001. Redlist categories. IUCN, Gland, Switzerland.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. Journal of Basic Engineering **82:**35–45.

McCarthy, M. A., and P. Masters. 2005. Profiting from prior information in Bayesian analyses of ecological data. Journal of Applied Ecology **42:**1012–1019.

Meinhold, R. J., and N. D. Singpurwalla. 1983. Understanding the Kalman filter. The American Statistician **37:**123–127.

Meir, E., and W. F. Fagan. 2000. Will observation error and biases ruin the use of simple extinction models? Conservation Biology **14:**148–154.

Miller, J. K., J. M. Scott, C. R. Miller, and L. P. Waits. 2002. The Endangered Species Act: Dollars and sense? BioScience **52:** 163–168.

Morris, W. F., and D. F. Doak. 2002. Quantitative conservation biology: theory and practice of population viability analysis. Sinauer, Sunderland, Massachusetts.

O'Grady, J. J., D. H. Reed, B. W. Brook, and R. Frankham. 2008. Extinction risk scales better to generations than to years. Animal Conservation **11:**442–451.

Price, C. 1993. Time, discounting and value. Blackwell Publishers, Oxford, United Kingdom.

Punt, A. E., and A. D. M. Smith. 1999. Harvest strategy evaluation for the eastern stock of gemfish (*Rexea solandri*). ICES Journal of Marine Science **56:**860–875.

Regan, T. J., B. L. Taylor, G. G. Thompson, J. F. Cochrane, R. Merrick, M. Nammack, S. Rumsey, K. Ralls, and M. Runge. 2009. Developing a structure for quantitative listing criteria for the U.S. Endangered Species Act using performance testing: phase I report. Technical memorandum NMFS-SWFSC-437. National Oceanic and Atmospheric Administration, La Jolla, California.

Robbins, K. 2009. Strength in numbers: setting quantitative criteria for listing species under the Endangered Species Act. Journal of Environmental Law and Policy **27:**1–37.

Shaffer, M. L. 1981. Minimum population sizes for species conservation. BioScience **31:**131–134.

Shaffer, M. L. 1983. Determining minimum viable populations sizes for the grizzly bear. International Conference for Bear Research and Management **5:**133–139.

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

Taylor, B. L., P. R. Wade, D. P. De Master, and J. Barlow. 2000. Incorporating uncertainty into management models for marine mammals. Conservation Biology **14:**1243–1252.

Taylor, B. L., P. R. Wade, U. Ramakrishnan, M. Gilpin, and H. R. Akcakaya. 2002. Incorporating uncertainty in population viability analyses for the purpose of classifying species by risk. Pages 239–252 in S. R. Beissinger and D. R. McCullough, editors. Population viability analysis. University of Chicago Press, Chicago.

Wade, P. R. 1999. Bayesian population viability analysis. Pages 213–238 in S. R. Beissinger and D. R. McCullough, editors. Population viability analysis. University of Chicago Press, Chicago.

Wilcove, D. S., and L. L. Master. 2005. How many endangered species are there in the United States? Frontiers in Ecology and Evolution **3:**414–420.