Conservation Biology

Review

A Framework for Developing Objective and Measurable Recovery Criteria for Threatened and Endangered Species

GINA K. HIMES BOOR

Ecology Department, Montana State University, P.O. Box 173460, Bozeman, MT 59717, U.S.A., email gkhimesboor@montana.edu

Abstract: *For species listed under the U.S. Endangered Species Act (ESA), the U.S. Fish and Wildlife Service and National Marine Fisheries Service are tasked with writing recovery plans that include "objective, measurable criteria" that define when a species is no longer at risk of extinction, but neither the act itself nor agency guidelines provide an explicit definition of objective, measurable criteria. Past reviews of recovery plans, including one published in 2012, show that many criteria lack quantitative metrics with clear biological rationale and are not meeting the measureable and objective mandate. I reviewed how objective, measureable criteria have been defined implicitly and explicitly in peer-reviewed literature, the ESA, other U.S. statutes, and legal decisions. Based on a synthesis of these sources, I propose the following 6 standards be used as minimum requirements for objective, measurable criteria: contain a quantitative threshold with calculable units, stipulate a timeframe over which they must be met, explicitly define the spatial extent or population to which they apply, specify a sampling procedure that includes sample size, specify a statistical significance level, and include justification by providing scientific evidence that the criteria define a species whose extinction risk has been reduced to the desired level. To meet these 6 standards, I suggest that recovery plans be explicitly guided by and organized around a population viability modeling framework even if data or agency resources are too limited to complete a viability model. When data and resources are available, recovery criteria can be developed from the population viability model results, but when data and resources are insufficient for model implementation, extinction risk thresholds can be used as criteria. A recovery-planning approach centered on viability modeling will also yield appropriately focused data-acquisition and monitoring plans and will facilitate a seamless transition from recovery planning to delisting.*

Keywords: Endangered Species Act, environmental laws, environmental policy, extinction risk assessment, population modeling, population viability analysis, recovery plans, threatened species

Un Marco de Referencia para Desarrollar Criterios de Recuperación Objetivos y Medibles para Especies Amenazadas y en Peligro

Resumen: *Para las especies enlistadas bajo el Acta de Especies en Peligro de los EUA, el Servicio Estadunidense de Pesca y Vida Silvestre y el Servicio Nacional de Pesquerías Marinas tienen la labor de establecer planes de recuperación que incluyan "criterios objetivos y medibles" que definan cuando una especie ya no está en riesgo de extinción, pero ni el acta ni las guías de trabajo de las agencias proporcionan una definición explícita de criterios objetivos y medibles. Resúmenes anteriores de planes de recuperación, incluyendo uno que se publicó en 2012, muestran que muchos criterios carecen de métricas cuantitativas con un razonamiento biológico claro y no están cumpliendo el mandato medible y objetivo. Revisé cómo los criterios objetivos y medibles han sido definidos implícita y explícitamente en literatura revisada por pares, el Acta de Especies en Peligro, otros estatutos de los EUA y decisiones legales. Basado en una síntesis de estas fuentes, propongo que los siguientes seis estándares se usen como requerimientos mínimos para criterios objetivos y medibles: que contengan un umbral cuantitativo con unidades calculables, que estipulen un marco de tiempo dentro del cual deben cumplirse, que explícitamente definan la extensión espacial o la población a la que aplican, que especifiquen*

Paper submitted February 29, 2012; revised manuscript accepted May 18, 2013.

DOI: 10.1111/cobi.12155

un procedimiento de muestreo que incluya tamaño de muestra, que especifiquen un nivel de significancia estadística y que incluyan una justificación proporcionando evidencia científica de que los criterios definen una especie cuyo riesgo de extinción se ha reducido al nivel deseado. Para que cumplan estos seis estándares, sugiero que los planes de recuperación sean guiados explícitamente por y organizados alrededor de un marco de trabajo de modelación de viabilidad poblacional aunque los datos o los recursos de las agencias sean muy limitados para completar el modelo de viabilidad. Cuando los datos y los recursos estén disponibles, los criterios de recuperación pueden desarrollarse a partir de los resultados del modelo de viabilidad poblacional, pero cuando los datos y los recursos son insuficientes para implementar un modelo, se pueden usar los umbrales de riesgo de extinción como criterios. Un método de planificación de recuperación centrado en el modelo de viabilidad también proporcionará datos y planes de monitoreo enfocados apropiadamente y facilitará una transición sencilla de la planificación de la recuperación al retiro de las especies de la lista.

Palabras Clave: Acta de Especies en Peligro, análisis de viabilidad poblacional, estudio de riesgo de extinción, especies amenazadas, leyes ambientales, modelación de poblaciones, planes de recuperación, políticas ambientales

Introduction

When a species is listed as threatened or endangered under the U.S. Endangered Species Act (ESA), the act requires that a recovery plan "for the conservation and survival" of the species be developed. Following amendments to the ESA in 1988, recovery plans are required to include "objective, measurable criteria which, when met, would result in a determination . . . that the species be removed from the list." Despite the pivotal role of recovery criteria in defining recovery for a species and indicating when delisting should be initiated, the ESA provides no clear guidance concerning the substance or structure of criteria beyond the objective, measurable language. Neither does the ESA define *endangered* or *recovered* in unequivocal terms that could dictate criteria content (Vucetich et al. 2006; Robbins 2009).

Scientific reviews of recovery plans and recovery criteria conducted in the third decade of the ESA found them effective in promoting and guiding recovery in some respects but lacking in others, and reviewers made the following recommendations for improvement: make recovery criteria more quantitative and consistent (Tear et al. 1993; Gerber & Hatch 2002), create clearer and more consistent links between criteria and a species' biological status and threats (Tear et al. 1993; Schemske et al. 1994; Gerber & Hatch 2002), increase the use of quantitative models for assessing extinction risk (e.g., population viability analysis [PVA]), and more thoroughly integrate such models into the recovery-planning process (National Research Council 1995; Morris et al. 2002). Some of these reviews (e.g., Gerber & Hatch 2002; Morris et al. 2002) were conducted collaboratively with the U.S. Fish and Wildlife Service (USFWS) and led directly to changes in the Recovery Plan Guidance Document (NMFS & USFWS 2010). The Guidance Document is an ESA-mandated how-to manual for the recovery-planning process written by and for the federal agencies responsible for implementing the ESA (USFWS and National Marine Fisheries Service [NMFS]; hereafter, the Services). The updated Guidance Document explicitly discussed the recommendations and suggested that recovery plans should "provide clearer and more consistent linkage between the biology of the species and the recovery criteria" should "outline and *justify* a strategy to achieve recovery" (emphasis added), and should be "specific," "technically feasible," "grounded in good science," and "time-referenced."

Despite the direction provided in the Guidance Document and previous recommendations, a recent assessment of recovery criteria and delisting thresholds showed that recovery plans and recovery criteria, although improved in some ways, have not improved with respect to many of the critical deficiencies previously identified. For example, 90% of species with potential for delisting (as defined by the Services) have at least one quantitative demographic recovery criterion listed in their recovery plan (an increase from past reviews), but justification for those numbers through quantification of decline in abundance, habitat, or range is rare, and "probabilistic assessments of persistence over time are nearly nonexistent" (Neel et al. 2012). The majority of recovery plans (93%) require higher total abundance than existed at recovery plan writing, but the percentage of species that would rank as secure (>10 populations) under International Union for Conservation of Nature (IUCN) standards at delisting has decreased in recent plans compared with older plans (Neel et al. 2012). Recent plans are also just as likely as older plans to allow delisting to occur with the same or fewer populations than existed at the time of listing or recovery plan writing (Neel et al. 2012).

Given these patterns, it bears asking why recovery criteria have not improved more over time. This question is complex and involves many factors, including sociopolitical ones, but one factor may be the lack of explicit, detailed protocols for developing objective, measurable criteria. The qualitative language of the ESA and lack of specificity in the Guidance Document has long been blamed for inefficiency and ineffectiveness in implementing the law (Easter-Pilcher 1996; Shelden et al. 2001; Vucetich et al. 2006). Although previous reviews and the updated Guidance Document provide more directed guidance than the ESA, they too lack clear and explicit guidelines on how to develop truly objective, measurable recovery criteria with clear biological justification. I sought to address this absence of an explicit set of best practices for recovery criteria and recovery plans. After synthesizing previous recommendations, the content of the ESA and Guidance Document, and other relevant statutes, policies, and court decisions, I devised a concrete working definition of *objective, measurable criteria* on the basis of 6 unequivocal standards. I also formulated a specific approach to the recovery-planning process that I suggest will result in comprehensive recovery plans that are quantitative and thoroughly justified, complete with recovery criteria that meet the 6 proposed standards. I designed the approach to be feasible under the different circumstances faced by writers of recovery plans, including those with time, budget, and informational constraints.

Definition of Objective and Measurable

The Oxford English Dictionary defines *objective* as "not influenced by personal feelings or opinions in considering and representing facts" (OED Online 2012). *Measurable* is synonymous with *quantifiable*, *determinate*, and *estimable* (i.e., that which can be determined without ambiguity). These definitions suggest that an objective, measurable criterion contains an unambiguous target that leaves no room for interpretation as to whether it has been achieved. In other words, "they minimize interpretation and judgment so that, at least in principle, anyone who applies the same criteria to the same data will get the same results" (Sagoff 1987), and those results will provide a definitive yes or no answer as to whether the criterion has been met.

Objective and *measureable* should, however, be defined not only from dictionary definitions, but within the broader context of the ESA and the policies, documents, and scholarly work that pertain to the statute and accepted scientific practices. For example, in multiple rulings regarding ESA implementation, courts have faulted the Services for their lack of specificity and for providing inadequate justification for decisions and actions (e.g., Middle Rio Grande Conservancy District v. Norton 2002; National Wildlife Federation v. NMFS 2011). Additionally, the Administrative Procedures Act, the Information Quality Act, and the Policy on Information Standards under the ESA require that federal decisions be objective, justified, and transparent (i.e., not be "arbitrary and capricious") (Murphy & Weiland 2011) and that justification be documented (APA 1946; USFWS 1994; IQA 2001). These sources in addition to previously published recommendations and the Guidance Document emphasize quantification, justification, clarity, consistency, specificity, and biological relevance (e.g., Schemske et al. 1994; Easter-Pilcher 1996; Tear et al. 2005).

To meet the measureable mandate as defined by all of these sources, I suggest that a recovery criterion must be quantitative and descriptively and statistically complete. To be objective it must be explicitly justified in the context of a species' ecology, and that justification must be documented and transparent. Specifically, I suggest that the following 6 standards define an objective and measurable criterion: (1) contains a quantitative threshold with specified quantifiable units, (2) stipulates a time frame over which it must be met, (3) explicitly defines the spatial extent or population to which it applies, (4) specifies a statistical significance level, (5) specifies a sampling procedure including sample size, and (6) is justified by providing scientific evidence that the threshold defines a species with the desired risk of extinction.

Quantification

The quantitative standard has 3 components: a numeric threshold, specified units of measure, and units technically measurable under standard scientific protocol. The value of using numeric metrics in recovery criteria and other natural resource management situations has long been recognized (Easter-Pilcher 1996; Tear et al. 2005; Robbins 2009). Although most recent plans include at least one numeric demographic criterion, other types of criteria pertaining to threat abatement or habitat conservation, for example, are rarely quantitative (Neel et al. 2012). Although demographic criteria should be the primary numerical criteria defining when the actual population is no longer at risk of extinction, management- or threat-focused criteria should also be written quantitatively. For example, a requirement to establish a cooperative agreement between 2 (or more) specific entities qualifies as quantitative in that it specifies that one agreement be established between a specific number and set of entities. In contrast, a criterion indicating, for example, that cooperation should be established with affected state agencies is qualitative and provides no definitive threshold to judge whether it has been met. To ensure measurability, clarity, and transparency, and to eliminate ambiguity and subjectivity, an actual numeric value is necessary in all criteria.

A quantitative threshold is useless, however, if units are not specified or if the units specified are not defined. For example, if a criterion requires a population to have 5 viable subpopulations, the units are stated (viable subpopulations); but unless *subpopulation* and *viable* are defined in explicit and numerical terms (e.g., the population occurring above 10,000 feet in the Green Mountains has $<5\%$ probability of extinction within 100 years), the quantitative nature of the criterion is lost. Similarly, if a criterion requires a specific amount of habitat, but *habitat* is not defined in explicit terms (e.g., what are the necessary components and in what quantity), determining whether that criterion has been met will necessarily be based on subjective judgment (Carroll et al. 2010). The same standard applies to management-related criteria. For the cooperative-agreement example above, the criterion should explicitly state what the agreement is to address and how it will be enforced.

Finally, if the specified units are not technically quantifiable or logistically possible to measure, the criterion is rendered unachievable. For example, the requirement that a population attain a specified size is a commonly used criterion. However, for most populations it will be impossible to know with certainty when a population reaches a specific level; rather, it is only possible to know when the estimated population size reaches the threshold. In most cases, each member of a population cannot be counted individually, so the size of a population is estimated through sampling and statistical estimation. Such approaches inherently involve uncertainty; thus, knowing when the true population size has reached the specified threshold will be equivocal. Although this may sound overly pedantic, unless the uncertainty in population sampling and population size estimation is considered and explicitly acknowledged, the criterion loses both scientific credibility and true attainability. If, however, a criterion states that the estimated population size or the field counts, for example, reach a specified threshold, the units are stated and technically measurable. Again, management- or threat-based criteria must meet this same standard. A criterion requiring a cooperative agreement, for example, must specify how the agreement is to be finalized and officially adopted to avoid uncertainty about completion of the criteria. Thus, to meet the quantitative standard of measurability, a criterion (regardless of content or focus) must include a numerical value defining a threshold, must have defined units, and those units must be realistically measurable or estimable under standard scientific practices.

Temporal and Spatial Specificity

To ensure the quantitative metric can be met without ambiguity or alternate interpretation, temporal and spatial references must be included (Mattson & Craighead 1994). A temporal reference indicates over how many years, seasons, or generations a metric must be met to satisfy the criterion. For example, the estimated population size or specific threat may be required to meet or exceed a specified threshold for 10 years before the criterion is met. Without a temporal requirement, the criterion is open to interpretation, and it will be unclear whether delisting should be initiated as soon as the threshold is crossed or whether that level must be sustained over some period. Threat levels and population parameters (e.g., size, growth rate, range) vary over time as a result of intrinsic and extrinsic factors; thus, a population or threat that meets a threshold at a single point in time is usually insufficient evidence of threat neutralization or population viability. The actual length of time specified in a criterion will be a function of the species' biology and policy determinations based on the level of extinction risk acceptable for the particular decision (downlisting or delisting). Non-demographic criteria should also contain temporal specificity. For example, a criteria requiring the development of a management plan should specify how long the plan should be in place prior to delisting and how frequently it should be updated once in place.

Spatial specificity means stipulating either the geographic area or specific populations or subpopulations to which the criterion applies. As noted above, if a criterion fails to indicate to which population or to what portion of the species' range it applies, it is neither technically measurable nor unequivocal and thus does not meet the objective and measurable standard. For example, habitat protection criteria must describe not only the extent of protected habitat, but also the spatial distribution of protected habitat to adequately protect necessary habitat components (e.g., protected patches should be within 20 km of one another to allow for dispersal [Carroll et al. 2010]). Spatial specificity is also necessary because a species' abundance can be increasing at a broad scale (e.g., over its entire range) and declining or remaining static at smaller scales (e.g., among localized populations). Threats and habitat decline can be similarly discordant at different scales. Recognizing the importance of scale and specifying the scale at which recovery will be measured and particular criteria will be applied are necessary to provide a clear unequivocal metric that will be less subject to alternate interpretations. As with temporal specification, the spatial object of the criterion should not be an arbitrary decision; rather, it should be based on the species biology and on policy decisions regarding acceptable losses and relevant scales.

Statistical and Sampling Specificity

If the quantitative criterion involves data collection and parameter estimation, as most demographic and habitat criteria do (e.g., population size, quantity of habitat occupied), it must include statistical specifications, including statistical significance level, sampling procedure, and sample size or some equivalent measure of statistical resolution. Inclusion of a significance level (e.g., 95% confidence that the estimated value meets the specified threshold) or a statement of probability (in a Bayesian context) is an acknowledgement that, as discussed above, the true value of a population metric cannot be known with certainty but only estimated, and that an estimate will inherently contain uncertainty. Such specification is critical to safeguard against falsely concluding that a species has met the criteria when it actually has not (type II error) (McGarvey 2007). For example, a point estimate of population growth rate may exceed the specified threshold, but considerable uncertainty may surround that estimate such that the confidence limits or credible interval spans zero, indicating the population may be declining rather than growing at the requisite rate (Taylor et al. 2002; Martin et al. 2007). Including a significance level or statement of probability sets a maximum tolerance for making an incorrect decision.

Specifying a sampling procedure, or referencing documentation that describes a sampling protocol to be followed, is also necessary. Such specification will require forethought about what kind of sampling protocols are reasonable given the species' biology and life history as well as agency budget and time constraints (Legg & Nagy 2006; Martin et al. 2007; Lindenmayer & Likens 2010). The sampling protocols meeting these criteria will in turn dictate what kind of population metrics are likely to result from such sampling and therefore what kind of quantitative metrics and error tolerances should be included in the recovery criteria. An illustrative example of the importance of considering biologically reasonable sampling protocols when identifying demographic recovery criteria can be seen in the case of the West Virginia northern flying squirrel (*Glaucomys sabrinus fuscus*). The population metrics measured for this species over many years of monitoring (representing the most feasible metric to monitor) did not match the population metrics required under the recovery criteria (Friends of Blackwater v. Salazar 2012).

The sampling procedure specified should also take detectability into account and should adequately spread sampling over time and space to ensure that heterogeneity is measured (Martin et al. 2007). As part of this, the sample size required to make such a determination is essential and should be specified in the criterion. Sample size is an indication of statistical power and, as such, is as important to full statistical specification as the statistical significance level (Ellison 1996; Legg & Nagy 2006). Statistical power, to use a population abundance trend example, is the ability of an analysis to detect a trend in abundance (positive or negative) if one exists. Thus, including within each criterion the sampling procedure and sample size necessary to adequately measure the metric of interest will ensure that the appropriate type, quality, and quantity of data will be collected for estimating whether the parameters of interest have met the specified criteria thresholds.

Scientific Justification

Of the 6 standards, the requirement for formal scientific justification is arguably the most important. If a criterion is quantitative and descriptively and statistically complete (thus measurable) but no justification is provided for the numeric value stipulated, the criterion may not define a truly recovered state by objective biological standards. Federal courts have indicated that agencies must articulate "a rational connection between the facts found and the choice made" *(*Baltimore Gas & Elec. v. Natural Resource Defense Council 1983) and, more specifically, that criteria must address the threats identified as the source of the species decline (Fund for Animals v. Babbitt 1995). These requirements for justification are not only legal requirements, but they serve the intent of the ESA; species with clear links between recovery plan criteria and their biology are more likely to have improving status than those with no link between their biology and criteria (Gerber & Hatch 2002).

Justification of criteria solely through the use of logical argument, however, is insufficient to meet this justification requirement. Although the logic behind some recovery criteria may appear self-evident and not in need of formal justification, many of the cause-and-effect arguments to justify recovery thresholds do not stand up under closer scrutiny (Taylor et al. 2002). For example, a population-size or growth-rate criterion might be set at some arbitrary level higher than the current depressed value based on the logic that a larger or increasing population size is sufficient evidence to demonstrate that threats have been addressed. Although such a line of reasoning may seem logical, both empirical and theoretical evidence suggests that large population sizes and short-term positive growth trajectories are neither assurance against extinction nor guarantors of a species' future success (Goodman 1987; Mangel & Tier 1994). Similarly, if a population's decline appears clearly linked to a specific threat, it cannot be assumed that removal of that threat will inevitably lead to species recovery (Peterman 1977; Suding et al. 2004; Metzger et al. 2010). The Cook Inlet beluga whale (*Delphinapterus leucas*) population is a prime example. Overharvest appeared to be the obvious reason for the species' decline, but a moratorium on harvest did not result in the expected population rebound (Goodman 2009). Although straightforward logic may provide guidance regarding recovery actions, it does not provide adequate scientific justification for criteria—the metrics used to measure whether recovery has actually occurred.

As an alternative to qualitative reasoning and logical argument, use of formalized demographic modeling can lead to biologically based recovery criteria and provide the scientific rationale to justify the defined criteria thresholds. PVA, an approach that uses quantitative methods to provide a probabilistic estimate of a population's viability (Table 1), is the most widely used and studied type of model available for these purposes (e.g., Beissinger & McCullough 2002; Morris & Doak 2002). Its utility in ESA implementation and recovery criteria development has been widely recognized in both agency documents (Angliss et al. 2002; DeMaster et al. 2004; U.S. Marine Mammal Commission 2008) and peer-reviewed literature (Carroll et al. 1996; Morris et al. 2002; Murphy & Weiland 2011).

Table 1. Definition of population viability analysis (PVA) and descriptions of PVA-related techniques.

Concept	*Description*	*Examples and background reading*
Population viability analysis (PVA)	Population viability analysis (PVA) comprises a broad set of analytical and modeling approaches used to assess a species' risk of falling below some specified threshold (extinction, quasi extinction, or other management goal). Such approaches can integrate a species' life history, demography, and sometimes genetics, with environmental variability, measurement uncertainty, and assumptions and uncertainties about future management, threats, and ecological conditions to project the future status of a species. The output of a PVA usually includes a probabilistic statement about a species' risk within a set time horizon, e.g., species X has 8.6% probability of falling below 250 individuals within 100 years.	Beissinger and McCullough (2002), Morris and Doak (2002), Beissinger et al. (2006)
Types of PVA and PVA-related techniques		
Count-based PVA	A probabilistic estimate of extinction risk based only on population count data over a number of years (not necessarily consecutive). From the count data, the population growth rate mean and variance are estimated and the probability of extinction calculated or simulated under the assumption of static mean growth and variance. This is the simplest type of PVA with the lowest data requirements.	Dennis et al. (1991), Shelden et al. (2001), NMFS (2008)
Demographic PVA, stage-structured PVA, demographically structured PVA	A probabilistic estimate of extinction risk based on survival, growth, and reproduction data for each life stage of a population (e.g., pre-breeders, seedlings, adults, senescent individuals). This type of PVA requires significantly more data than count-based PVA, but can yield more information relevant to managers, such as which vital rates are unnaturally low and most limiting to population growth or stability.	Holmes (2001), Bakker et al. (2009)
Metapopulation PVA	A PVA that incorporates multiple populations (of the same species) and dispersal rates among them to project a risk of extinction for the individual populations and the metapopulation as a whole. In addition to the data necessary to conduct PVAs on the individual populations, information about dispersal rates is also required for this type of PVA.	Beissinger and McCullough (2002), Schultz and Hammond (2003), USFWS (2010)
Sensitivity analysis	Any number of analytical approaches that estimate how sensitive one variable in a demographic PVA model is to changes in another variable. The sensitivity is usually measured in a population's growth rate or risk of extinction. Sensitivity analyses can be used to assess the likely effectiveness of different management strategies.	Mills and Lindberg (2002), Morris and Doak (2002)
Bayesian PVA	An approach that applies Bayesian statistical techniques to the estimation of PVA input parameters and extinction risk. Specifically, a Bayesian PVA incorporates parameter uncertainty into the model through the use of probability distributions rather than point estimates for the various input parameters (e.g., survival rate, frequency of catastrophic events). The result of a Bayesian PVA is a posterior probability of extinction that incorporates both parameter uncertainty and environmental stochasticity and can include model structure uncertainty (e.g., the level of density dependence).	Goodman (2002*a*), Wade (2002), NMFS (2008)
Quantitative threats analysis, risk assessment, and effects analysis	Represent a variety of analytical approaches to identifying or quantifying how threats will affect a species. Some threats analyses are a form of sensitivity analyses, wherein threat levels within a model are systematically varied to assess their effect on demographic parameters and extinction risk. A similar approach formally compares output from multiple PVAs under varying future threat or management scenarios. Risk assessment is a formal process of gathering information on specific threats to inform decision-making and to develop risk-management strategies to minimize those threats.	USFWS (1992), Runge et al. (2007), Murphy and Weiland (2011)
Population viability management	An approach that combines adaptive management and population viability assessment into a single comprehensive framework. It uses iterative PVA modeling to assess extinction risk and management efficacy, which then can be used to update species monitoring approaches and management actions.	Bakker and Doak (2009), Ralls et al. (2002), Burgman (2006)

Population viability analyses vary from simple count-based models to highly complex metapopulation analyses. The type of model and the degree of complexity needed depend on the species of interest and the available data. In their various forms, PVA models can provide a structured method for exploring how threats might have affected a species and how the species might respond to reductions in those threats in the future (Beissinger & McCullough 2002; Burgman 2006). They can also be used to assess what population parameters (e.g., survival) are most crucial in maintaining or increasing population size and to suggest different combinations of population size and other parameter levels that define a population that is no longer readily susceptible to extinction (Mills & Lindberg 2002; Morris & Doak 2002). When integrated with adaptive management approaches PVA can also yield directed management and monitoring recommendations focused on achieving specified population targets (Ralls et al. 2002; Burgman 2006; Bakker and Doak 2009). Given the uncertainty inherent in the estimation of population parameters and in modeling both past and future scenarios, a good viability model will also incorporate this uncertainty and yield quantitative estimates of the uncertainty as output (Goodman 2002*a*; Burgman 2006; Bakker et al. 2009).

Much of the data and assumptions used in a PVA would likely be gathered as part of the typical recovery-planning process, but the biologically based mathematical modeling involved in PVA provides a formalized way to synthesize the information and distill the results into a common unit of measurement (i.e., extinction risk). It is this formalized synthesis that renders PVA more scientifically robust and transparent than cause-and-effect logic, historical patterns, educated guesses, expert opinion, or any other approaches used to set recovery thresholds (Burgman 2006). Through the formalized modeling process, a PVA can point to appropriate criteria thresholds that meet management limitations and goals, provide robust justification for those thresholds in the context of the species biology and its threats, make explicit any assumptions that are used in the model, suggest management and monitoring schemes to meet specified goals, and describe the degree of certainty (or uncertainty) that specified thresholds truly define a recovered state. It may not be possible to incorporate every threat or aspect of management into a PVA, but the majority of recovery criteria, whether focused on demography, threats, or management actions, can and should be linked to the species biology through a formalized modeling process. Criteria developed from such an approach will be scientifically justified, biologically based, and transparent. PVA also allows for new data and changes in assumptions to be readily incorporated into the synthesis and can serve as a unifying framework beyond recovery planning, including the 5-year review process and the threat assessments necessary for delisting.

An objective, measurable criterion that meets the letter and intent of the ESA, relevant statutes, policy statements, previous recommendations, and current scientific standards is one that is quantitative, temporally, spatially, and statistically specific, specifies a sampling protocol, and is justified on the basis of current population viability modeling methods that incorporate demographics, limiting factors, threats, future management actions, and uncertainty. Adherence to a subset of these recommendations will result in improved recovery criteria, but I suggest that unless all 6 standards are applied, recovery criteria will continue to fall short of the ESA's objective, measurable mandate.

Implementing the Standards

Although the 6 standards are scientifically and legally justified, the reality of limited time and resources makes it imperative to also address how the Services can feasibly meet these standards. In this section, I outline an approach to recovery planning that I believe can make achievement of all 6 standards realistically attainable even when data and resources are limited. The approach uses PVA as an organizing framework, but in cases of limited data and resources, it does not require actual development of a PVA model.

PVA-Framed Recovery Planning with Model Implementation

When expertise and data are available to conduct a PVA as part of the recovery-planning process, the development of the PVA model should be done simultaneously and in conjunction with the information-gathering and writing phases of the recovery plan. The PVA can be used to direct and focus the entire recovery-planning process, an approach that will yield a cohesive recovery plan centered on the factors most important for recovering a species and for developing recovery criteria that define a species with an explicit level of extinction risk. All the components necessary for developing a scientifically credible PVA are the same elements of a comprehensive recovery plan (Ralls et al. 2002; NMFS & USFWS 2010):

- assemble available data about a species (e.g., current status, biology, population dynamics, habitat requirements, spatial structure, etc.),
- document what data are not yet available and need to be gathered,
- identify the types and level of current and historical threats affecting the species (i.e., threats assessment),
- specify the level of various population, habitat, management, and threat parameters that define the desired level of extinction risk or population stability (recovery criteria), and

- outline the recovery strategy and the management actions necessary to reduce threats and improve the status of the species to reach the recovery criteria thresholds.

Approaching these steps with the goal of not only writing a recovery plan, but developing a PVA model will improve efficiency of the process and result in recovery plans and models of higher quality.

For example, in the threats assessment step, both the writing of the recovery plan and the development of a PVA should include an accounting of the data and uncertainty related to the mechanisms and intensities of threats and the projected effect of those threats in the future. Using the requirements of a PVA to focus information-gathering will necessitate defining quantitative metrics rather than the qualitative assessments that might otherwise be deemed sufficient for writing the recovery plan (e.g., Neel et al. 2012). In turn, the thorough accounting of threats for the recovery plan may clarify the structure and scope of the PVA model. Melding the 2 processes of writing the recovery plan and conducting a PVA will yield a more focused and quantitative recovery plan and a more comprehensive viability analysis, ensuring that synthesis of the available data and justification of recovery criteria will be an integral part of the recovery-planning process. In addition, the natural results of the melded processes should be targeted management actions (which are required under the ESA), appropriately directed data-acquisition and monitoring plans (Bakker & Doak 2009), and a more seamless transition from recovery planning to Section 7 consultation (Murphy & Weiland 2011), 5-year reviews, and the threat assessment necessary for delisting.

PVA-Framed Recovery Planning without Model Implementation

In cases when the necessary data or resources are not available to fully complete a PVA as part of the recovery-planning process, the structure and data requirements of a viability analysis can still be used as the organizing theme around which a recovery plan is focused. Although such an approach does not require implementation of a model, it will require the participation of a person knowledgeable in viability-analysis modeling to help guide the process. Using this approach, each step of the recovery-planning process, as described above, will focus on gathering the specific data that can be used in a viability model in the future and on identifying the information that is unavailable but necessary for a future model. Identification and quantification of missing data can be used to develop specific management actions for collection of appropriate data in the future or can be incorporated as quantified uncertainty in the future model. In this approach, the recovery-planning process itself serves as the early stage of PVA model development and informs the structure and scope of the future model. Combining the recovery planning and PVA development processes in this way reduces redundancy in effort, results in a more coherent and quantitative plan, informs monitoring needs and management actions (Bakker & Doak 2009), and can lead to development of a PVA for the delisting threat-assessment process.

When a PVA is not conducted as part of the recovery-planning process, however, the quantitative rationale for defining specific recovery criteria will be lacking. To address this, extinction risk thresholds (e.g., $<1\%$ risk of extinction in 100 years) can be used as recovery criteria. Extinction-risk criteria are by definition quantitative and descriptively and statistically specified, thus meeting the quantification and temporal, spatial, and statistical specificity standards outlined above. In addition, if the PVA structure is used to direct the recovery-planning process as recommended, the recovery plan should yield management actions in the form of data-collection protocols (including sample size and sample design) to conduct a future PVA and assess extinction risk. These data-collection protocols should be included with the extinction risk criteria to satisfy the sampling protocol standard. Because the level of extinction risk that defines a recovered (or threatened or endangered) species is a policy question rather than a scientific question, extinction-risk thresholds do not need a scientific rationale, only an explicit policy decision (Goodman 2002*b*; Robbins 2009). Nonetheless, use of an extinction-risk threshold as a criterion presumptively leads to implementation of a PVA in the future, which will ultimately provide the biological rationale for the demographic metrics deemed necessary to achieve the extinction risk specified in the criterion. If threats are included in the PVA, as recommended above, it can be used directly in the threat-assessment process necessary for delisting. In this way, recovery criteria adhering to all 6 proposed standards can be developed even when data and resource limitations preclude implementation of a PVA as part of the recovery-planning process.

Discussion

Although the standards and protocol outlined above are unlikely to raise objections on the basis of biological or scientific questions, several arguments could be made on the basis of more practical grounds that these standards are either unreasonable or unwarranted. Some argue, for example, that PVAs require too much data or provide answers that are too imprecise; therefore, their use as tools for assessing and managing threatened and endangered species is impractical or unconstructive (e.g., Coulson et al. 2001). The ESA requires the development of recovery criteria regardless of the quantity or quality of the existing information, and the uncertainty of a species' future prospects are present whether a PVA is used or not. As others have demonstrated, PVA is not perfect, but it is the best tool currently available and is valuable even when data are limited (Beissinger & McCullough 2002; Brook et al. 2002; Burgman 2006). In addition, the explicit quantification of uncertainty is one of its primary benefits rather than one of its drawbacks (Goodman 2002*a*; Taylor et al. 2002; Murphy & Weiland 2011).

In a recent court ruling concerning the delisting of the West Virginia Northern Flying Squirrel (Friends of Blackwater v. Salazar 2012) a circuit court of appeals ruled that recovery criteria are not legally binding and that recovery plans are not contractual documents. Some might argue that this ruling alleviates the need for the specificity and quantitative rigor specified in the 6 proposed standards. Regardless of whether criteria themselves are legally binding, many recovery plans are failing to meet either the letter or intent of the ESA in conserving and recovering species such that they will be self-sustaining and no longer in need of special management under the law (Neel et al. 2012). The recovery-planning approach and 6 standards I advocate here will yield comprehensive and quantitatively focused recovery plans and scientifically justified recovery criteria that define truly recovered species and thus will increase the effectiveness and efficiency of the ESA and the delisting process.

Some might argue that too much specificity in recovery criteria will impede future management flexibility. Although my approach requires explicit and transparent documentation of criteria and management and policy intentions, maintaining policy- and management-related flexibility is no more or less achievable under the framework outlined here than under any less well-defined approach. Provisions in the ESA allow recovery plans to be modified as new information becomes available, and the court decision in the flying squirrel case allows for further agency flexibility (Friends of Blackwater v. Salazar 2012). Use of PVA can significantly streamline incorporation of new data and new policy directives; thus, it is easier to change criteria or management direction while maintaining transparency and scientific justification for those decisions when PVA is used. When a PVA cannot be developed, use of extinction-risk thresholds as recovery criteria provides maximal flexibility in future management direction while maintaining scientific rigor, adhering to the 6 criteria standards, and making policy decisions explicit and transparent.

Recovery planning and development of recovery criteria are difficult endeavors, made more difficult in the all-too-common situation when data are limited and budgets are tight. My recommendations are made in the spirit of providing a framework that might be useful to the Services even under these difficult limitations, and are not meant to trivialize the task of recovery planning or the constraints under which the Services operate in implementing the ESA. Strictly adhering to the approach recommended here may, in some cases, make recovery-plan development more time-consuming initially, but I argue that, in the long run, it will aid the Services and threatened and endangered species by integrating recovery planning, species monitoring, status review, and the threat-assessment delisting process into a unified framework defined by the application of PVA. The framework is also designed to increase transparency, thus potentially decreasing legal challenges related to recovery planning and delisting. However, it is unlikely that the Services could meet the rigorous standards set out herein without assistance. If academics and other scientists with expertise in conservation biology and PVA development are willing to contribute to the process, significant progress could be made toward making recovery criteria fully justified and truly objective and measurable.

Acknowledgments

This research was partially funded by the U.S. Army Corps of Engineers and the U.S. Marine Mammal Commission. I especially thank M. Porter and T. Ragen for their support of this project. I am also indebted to D. Doak, W. Morris, V. Bakker, and 3 anonymous reviewers, whose critiques and recommendations significantly improved this paper. Finally, I dedicate this paper to D. Goodman, whose intellect and mentorship significantly influenced this work and so much of my professional development. Your brilliant mind, sage advice, and unflagging drive to improve conservation science and decision-making will be missed.

Literature Cited

Angliss, R. P., G. K. Silber, and R. Merrick. 2002. Report of a workshop on developing recovery criteria for large whale species. Technical memorandum NMFSF/OPR-21. National Oceanic and Atmospheric Administration, Silver Spring, Maryland.

APA (Administrative Procedure Act). 1946. U.S. code. Volume 5. Sections 551–706.

Bakker, V. J., and D. F. Doak. 2009. Population viability management: ecological standards to guide adaptive management for rare species. Frontiers in Ecology and the Environment **7:**158–165.

Bakker, V. J., D. F. Doak, G. W. Roemer, D. K. Garcelon, T. J. Coonan, S. A. Morrison, C. Lynch, K. Ralls, and R. Shaw. 2009. Incorporating ecological drivers and uncertainty into a demographic population viability analysis for the island fox. Ecological Monographs **79:**77–108.

Beissinger, S. R., and D. R. McCullough, editors. 2002. Population viability analysis. University of Chicago Press, Chicago.

Beissinger, S. R., J. R. Walters, D. G. Catanzaro, K. G. Smith, J. B. Dunning Jr., S. M. Haig, B. R. Noon, and B. M. Stith. 2006. Modeling approaches in avian conservation and the role of field biologists. Ornithological Monographs **59:**1–56.

Brook, B. W., M. A. Burgman, H. R. Akcakaya, J. J. O'Grady, and R. Frankham. 2002. Critiques of PVA ask the wrong questions: throwing the heuristic baby out with the numerical bath water. Conservation Biology **16:**262–263.

Burgman, M. 2006. The logic of good decisions: learning from population viability analysis. Society for Conservation Biology Newsletter **3:**17–18.

Carroll, R., C. Augspurger, A. Dobson, J. Franklin, G. Orians, W. Reid, R. Tracy, D. Wilcove, and J. Wilson. 1996. Strengthening the use of science in achieving the goals of the Endangered Species Act: an assessment by the Ecological Society of America. Ecological Applications **6:**1–11.

Carroll, C., J. A. Vucetich, M. P. Nelson, D. J. Rohlf, and M. K. Phillips. 2010. Geography and recovery under the U.S. Endangered Species Act. Conservation Biology **24:**395–403.

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

DeMaster, D., R. Angliss, J. Cochrane, P. Mace, R. Merrick, M. Miller, S. Rumsey, B. Taylor, G. Thompson, and R. Waples. 2004. Recommendations to NOAA Fisheries: ESA listing criteria by the quantitative working group. Technical memorandum NMFS-F/SPO-67. National Oceanic and Atmospheric Administration, Silver Spring, Maryland.

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

Easter-Pilcher, A. 1996. Implementing the Endangered Species Act. BioScience **46:**355–363.

Ellison, A. M. 1996. An introduction to Bayesian inference for ecological research and environmental decision-making. Ecological Applications **6:**1046–1046.

Gerber, L. R., and L. T. Hatch. 2002. Are we recovering? An evaluation of recovery criteria under the U.S. Endangered Species Act. Ecological Applications **12:**668–673.

Goodman, D. 1987. The demography of chance extinction. Pages 11–34 in M. E. Soule, editor. Viable populations for conservation. Cambridge University Press, Cambridge, United Kingdom.

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. McCollough, editors. Population Viability Analysis. University of Chicago Press, Chicago.

Goodman, D. 2002*b*. Uncertainty, risk, and decision: the PVA example. American Fisheries Society Symposium **24:**171–196.

Goodman, D. 2009. The future of fisheries science: merging stock assessment with risk assessment, for better fisheries management. Pages 537–566 in R. J. Beamish and B. J. Rothschild, editors. The future of fisheries science in North America. Springer Science + Business Media, New York.

Holmes, E. E. 2001. Estimating risks in declining populations with poor data. Proceedings of the National Academy of Science **98:**5072–5077.

IQA (Information Quality Act). 2001. Public law 106–554. Section 515.

Legg, C. J. and L. Nagy. 2006. Why most conservation monitoring is, but need not be, a waste of time. Journal of Environmental Management **78:**194–199.

Lindenmayer, D. B., and G. E. Likens. 2010. The science and application of ecological monitoring. Biological Conservation **143:**1317–1328.

Mangel, M., and C. Tier. 1994. Four facts every conservation biologist should know about persistence. Ecology **75:**607–614.

Martin, J., W. M. Kitchens, and J. E. Hines. 2007. Importance of well-designed monitoring programs for the conservation of endangered species: case study of the snail kite. Conservation Biology **21:**472–481.

Mattson, D. J., and J. J. Craighead. 1994. The Yellowstone grizzly bear recovery program: uncertain information, uncertain policy. Pages 101–129 in T. W. Clark, R. P. Reading, and A. L. Clarke, editors. Endangered species recovery: finding the lessons, improving the process. Island Press, Washington, D.C.

McGarvey, D. J. 2007. Merging precaution with sound science under the Endangered Species Act. BioScience **57:**65–70.

Metzger, K. L., A. R. E. Sinclair, R. Hilborn, J. Grant, C. Hopcraft, and S. A. R. Mduma. 2010. Evaluationg the protection of wildlife in parks: the case of African buffalo in Serengeti. Biodiversity Conservation **19:**3431–3444.

Mills, L. S. and M. S. Lindberg. 2002. Sensitivity analysis to evaluate the consequences of conservation actions. Pages 338–366 in S. R. Beissinger and D. R. McCullough, editors. Population viability analysis. University of Chicago Press, Chicago.

Morris, W. F., P. L. Bloch, B. R. Hudgens, L. C. Moyle, and J. R. Stinchcombe. 2002. Population viability analysis in endangered species recovery plans: past use and future improvements. Ecological Applications **12:**708–712.

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

Murphy, D. D. and P. S. Weiland. 2011. The route to best science in implementation of the Endangered Species Act's consultation mandate: the benefits of structured effects analysis. Environmental Management **47:**161–172.

National Research Council. 1995. Science and the Endangered Species Act. National Academy Press, Washington, D.C.

Neel, M. C., A. K. Leidner, A. Haines, D. D. Goble, and J. M. Scott. 2012. By the numbers: How is recovery defined by the US Endangered Species Act? BioScience **62:**646–657.

NMFS (National Marine Fisheries Service). 2008. Recovery plan for the Steller sea lion (*Eumetopias jubatus*) Revision. National Marine Fisheries Service, Silver Spring, Maryland.

NMFS (National Marine Fisheries Service) and USFWS (U.S. Fish and Wildlife Service). 2010. Interim endangered and threatened species recovery planning guidance. Version 1.3. National Oceanic and Atmospheric Administration, Silver Spring, Maryland. Available from http://www.nmfs.noaa.gov/pr/pdfs/recovery/guidance.pdf (accessed June 28, 2013).

OED Online (Oxford English Dictionary Online). 2012. Oxford University Press, New York.

Peterman, R. M. 1977. A simple mechanism that causes collapsing stability regions in exploited salmonid populations. Journal of the Fisheries Research Board of Canada **34:**1130–1142.

Ralls, K., S. R. Beissinger, and J. F. Cochrane. 2002. Guidelines for using population viability analysis in endangered-species management. Pages 521–550 in S. R. Beissinger and D. R. McCollough, editors. Population viability analysis. University of Chicago Press, Chicago.

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

Runge, M. C., C. A. Sanders-Reed, C. A. Langtimm, and C. J. Fonnesbeck. 2007. A quantitative threats analysis for the Florida manatee (*Trichechus manatus latirostris*). Open-File Report 2007-1086. U.S. Geological Survey, Reston, Virginia.

Sagoff, M. 1987. Where Ickes went right or reason and rationality in environmental law. Ecology Law Quarterly **14:**265–323.

Schemske, D. W., B. C. Husband, M. H. Ruckelshaus, C. Goodwillie, I. M. Parker, and J. G. Bishop. 1994. Evaluating approaches to the conservation of rare and endangered plants. Ecology **75:**584–606.

Schultz, C. B. and P. C. Hammond. 2003. Using population viability analysis to develop recovery criteria for endangered insects: case study of the Fender's blue butterfly. Conservation Biology **17:**1372–1385.

Shelden, K. E. W., D. P. DeMaster, D. J. Rugh, and A. M. Olson. 2001. Developing classification criteria under the U.S. Endangered Species Act: bowhead whales as a case study. Conservation Biology **15:**1300–1307.

Suding, K. N., K. L. Gross, and G. R. Houseman. 2004. Alternative states and positive feedbacks in restoration ecology. Trends in Ecology & Evolution **19:**46–53.

Taylor, B. L., P. R. Wade, U. Ramakrishnan, M. Gilpin, and H. R. Akçakaya. 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. McCollough, editors. Population viability analysis. University of Chicago Press, Chicago.

Tear, T.H., et al. 2005. Recurrent problem of setting measurable objectives in conservation. BioScience **55:**835–849.

Tear, T. H., J. M. Scott, P. H. Hayward, and B. Griffith. 1993. Status and prospects for success of the Endangered Species Act: a look at recovery plans. Science **262:**976–977.

USFWS (U.S. Fish and Wildlife Service). 1994. Interagency policy on information standads under the ESA. Federal Register **59:**34271.

USFWS (U.S. Fish and Wildlife Service). 1992. Cui-ui (*Chasmistes cujus*) recovery plan. 2nd revision. USFWS, Portland, Oregon.

USFWS (U.S. Fish and Wildlife Service). 2010. Southwest Alaska distinct population segment of the northern sea otter (*Enhyra lutris kenyoni*)—draft recovery plan. USFWS, Region 7, Anchorage, Alaska.

U.S. Marine Mammal Commission. 2008. The biological viability of the most endangered marine mammals and the cost-effectiveness of protection programs. Marine Mammal Commission, Bethesda, Maryland.

Vucetich, J. A., M. P. Nelson, and M. K. Phillips. 2006. The normative dimension and legal meaning of endangered and recovery in the U.S. Endangered Species Act. Conservation Biology **20:**1383–1390.

Wade, P. R. 2002. Bayesian population viability analysis. Pages 213–238 in S. R. Beissinger and D. R. McCollough, editors. Population viability analysis. University of Chicago Press, Chicago.