Recommendations to NOAA Fisheries: ESA Listing Criteria by the Quantitative Working Group
10 June 2004

D. DeMaster (chair), R. Angliss, J. Cochrane, P. Mace, R. Merrick,
M. Miller, S. Rumsey, B. Taylor, G. Thompson, and R. Waples

U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Marine Fisheries Service

NOAA Technical Memorandum NMFS-F/SPO-67
December 2004

Recommendations to NOAA Fisheries: ESA Listing Criteria by the Quantitative Working Group 10 June 2004

D. DeMaster (chair), R. Angliss, J. Cochrane, P. Mace, R. Merrick, M. Miller, S. Rumsey, B. Taylor, G. Thompson, and R. Waples

NOAA Technical Memorandum NMFS-F/SPO-67
December 2004

U.S. Department of Commerce
Donald L. Evans, Secretary

National Oceanic and Atmospheric Administration
Vice Admiral Conrad C. Lautenbacher, Jr., USN (Ret.)
Under Secretary for Oceans and Atmosphere

National Marine Fisheries Service
William T. Hogarth, Assistant Administrator for Fisheries

Suggested citation:

DeMaster, D. (chair), 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, 10 June 2004. U.S. Dep. Commerce, NOAA Tech. Memo. NMFS-F/SPO-67, 85 p.

A copy of this report may be obtained from:

Alaska Fisheries Science Center, F/AKC
NMFS, NOAA
7600 Sand Point Way NE
Seattle, WA 98115

Or online at:

http://spo.nmfs.noaa.gov/tm/

EXECUTIVE SUMMARY

The Endangered Species Act of 1973 (ESA) employs a two-category system: listing species either as endangered (in danger of extinction throughout all or a significant portion of its range) or threatened (likely to become endangered in the foreseeable future). Absence of Congressional guidance on how to interpret the terms used in the statutory definitions of these categories has left the task of defining them to the U.S. Fish and Wildlife Service and NOAA Fisheries (National Marine Fisheries Service). To date, neither of these agencies has developed uniform guidelines for listing, reclassifying, or delisting species. The lack of uniform guidelines for listing decisions has led to inconsistencies and inequities in the listing process. NOAA Fisheries responded to this problem by establishing a Steering Committee and a Quantitative Working Group (QWG) to work toward developing quantitative procedures that will make listing decisions "more transparent, consistent, and scientifically and legally defensible." The Steering Committee, in turn, has provided the QWG with a set of Guiding Principles which state that these procedures should possess characteristics such as applicability, implementability, transparency, flexibility, and equitability. The QWG offers the present report as a possible roadmap by which NOAA Fisheries could eventually develop uniform guidelines for listing, reclassifying, or delisting species.

Briefly, the QWG proposes the following process: (1) *overarching definitions* for both endangered and threatened should be adopted, (2) values of any *policy parameters* associated with the overarching definitions (e.g., the level of extinction risk corresponding to "endangered") should be specified, (3) *decision metrics* that can be used as proxies for (1) and (2) in data-poor cases should be developed for an appropriate range of taxonomic groups or life history types, and (4) all of the above should be done in the context of *performance testing* (use of simulations to evaluate how well an alternative performs relative to the objective). These four steps are explained further below.

The QWG has developed three alternative overarching definitions for "endangered" (EN). The **Probability of Extinction Threshold** definition states that a species is EN if its probability of extinction within a specified time horizon exceeds some cutoff percentage. The **Depensatory Threshold** definition states that a species is EN if its abundance, area of distribution, or other relevant metric falls below the level at which depensatory (Allee) effects are likely to predominate or population processes are largely unknown. The **Comprehensive Threshold** definition is similar to the definition for the Probability of Extinction Threshold, except that instead of looking at a single time horizon, the likelihood of extinction at each point in time is weighted appropriately to arrive at a comprehensive measure of risk.

All of the alternative overarching definitions of EN are associated with policy parameters, so called because their values should be set as a matter of public policy. In the Probability of Extinction Threshold definition, the policy parameters are the time horizon and the cutoff probability. In the Comprehensive Threshold definition, the policy parameters are the cutoff probability and any parameters needed to specify the weighting function. In the Depensatory Threshold definition, the Agency's policy regarding the level of threat of extinction associated with an EN listing, while not explicit, is integral to selecting the value for the depensation threshold. For ease of reference, the combination of a particular overarching definition together with a set of specific values for any associated policy parameters is called a *listing criterion*. For example, the Probability of Extinction Threshold definition with a time horizon of 100 years and a cutoff probability of 0.05 constitutes one possible listing criterion for EN.

Only one listing criterion for "threatened" (TH) is presented in this report. This is because most of the members of the QWG believe that the statutory language regarding TH "in the foreseeable future" is less ambiguous than the statutory language regarding EN. Some members of the QWG believe that Congress intended species classified as EN to be "on the brink" of extinction, while species classified as TH were not in danger of extinction in the foreseeable future. Most of the QWG members recommend that NOAA Fisheries consider a species to be appropriately classified as TH whenever its probability of becoming EN within 20 years exceeds 0.5. However, some members of the QWG would have preferred a definition of TH that had a lower probability of becoming endangered and a longer time period over which a species could become endangered. Pending the receipt of further guidance from the Steering Committee and from the USFWS, and pending the results of the performance analysis of the approaches described in this report, this recommendation should be considered preliminary. It was also noted that the proposed definition of TH is inconsistent with the definition of TH currently used by the FWS, as well as the definition of TH used by NOAA Fisheries in classifying salmon populations under the ESA.

Once a listing criterion (for either EN or TH) has been adopted, it can be used directly to make listing decisions, provided that sufficient data and other resources are available. Often, however, sufficient data or other resources will not be available, in which case other decision metrics must be used. A vast array of such metrics has been proposed in the past, including specified levels of absolute abundance, specified rates of decline in abundance, specified fractions of historical habitat loss, etc. Such decision metrics can be used either singly or in combination to approximate a particular listing criterion in data-poor cases, and several detailed examples are suggested in this report. Any of the suggested decision metrics can be applied to any of the overarching definitions in principle, although the appropriate values of those metrics may depend on the definition to which they are applied.

The QWG recommends that final decisions regarding listing criteria and decision metrics be made in the context of performance testing. The purpose of performance testing is to evaluate how alternative listing criteria and decision metrics perform relative to one or more management objectives. This is accomplished by simulating the performance of the alternatives and using a set of performance measures to translate the simulation output and the objective into a common currency. The QWG recommends that performance testing be conducted in two phases. The first phase would focus on the listing criterion for EN and TH. Here, the purpose would be to evaluate alternative overarching definitions and alternative values of any associated policy parameters. At the conclusion of the first phase, a listing criterion for EN and a listing criterion for TH should be adopted, at least provisionally. The second phase would focus on the decision metrics. Here, the purpose would be to determine which decision metrics serve as the best proxies for the EN and TH listing criteria. In the event that the listing criterion for EN proves too difficult to approximate by any particular decision metric, the first phase could be repeated with a new set of alternatives. In both phases, performance testing is likely to be an iterative process, as intermediate results will likely lead to new alternatives to test.

Because the recommended performance testing is likely to take at least 2 years to complete, the QWG recommends that an interim protocol for NOAA Fisheries be developed that 1) separates risk analysis from risk management (where risk analysis refers to the analyses which result in estimates and describes the individual species' projected status, and where risk management refers to the standards for how much and how soon species must be at risk of extinction and how certain we want to be of those estimates for the species to qualify for ESA protection) and 2) incorporates the use of structured expert opinion for risk analysis following a specified protocol. The QWG further recommends that the approach currently developed by the Northwest Region for the classification of Pacific salmon be used as a starting point for the development of an interim protocol.

Finally, the QWG recommends the establishment of a committee authorized to make final choices of listing criteria and decision metrics, and the establishment of a working group to enhance the use of structured decision-making in the listing process.

GLOSSARY

AFS	American Fisheries Society
BRT	Biological Review Team
CITES	Convention on the International Trade of Endangered Species
DPS	Distinct Population Segment
EN	Endangered
ESA	Endangered Species Act
ESU	Evolutionary Significant Unit
FACA	Federal Advisory Committee Act
FEMAT	Forest Ecosystem Management Assessment Team
IUCN	International Union for Conservation of Nature
NL	Not listed
NOAA Fisheries	National Marine Fisheries Service, National Oceanic and Atmospheric Administration
PVA	Population Viability Analysis
QWG	Quantitative Working Group
TH	Threatened
USFWS	U.S. Fish and Wildlife Service
VSP	Viable Salmonid Population

TABLE OF CONTENTS

EXECUTIVE SUMMARY III

GLOSSARY VII

TABLE OF CONTENTS IX

INTRODUCTION 1

Terms of Reference 1
Structure of the Report 2

TREATMENT OF "ENDANGERED" STATUS 3

Definitions of "Endangered" 3
Basic Concepts 3
Description of the Alternative Overarching Definitions 5
Evaluation of the Alternative Overarching Definitions 7
Decision Metrics That Could be Used to Approximate the Listing Criteria 9
Probability of Extinction Threshold Definition 9
Depensatory Threshold Definition 11
Treatment of Uncertainty 17

TREATMENT OF "THREATENED" STATUS 17

PERFORMANCE TESTING 18

Performance Testing of Alternative Listing Criteria 20
Performance Testing of Alternative Decision Metrics 21

RECOMMENDATIONS 23

Adopt an Interim Policy 23
Conduct Performance Testing of Listing Criteria and Decision Metrics 23
Choose Final Listing Criteria and Decision Metrics 23
Enhance Use of Structured Decision Making 23

ACKNOWLEDGEMENTS 24

CITATIONS 25

APPENDICES 33

APPENDIX 1: LESSONS LEARNED FROM THE EXPERIENCE OF THE U.S. FISH AND WILDLIFE SERVICE 35

Advantages of the Case-by-Case Approach 36
Disadvantages of the Case-by-Case Approach 36

APPENDIX 2: REVIEW OF EXISTING APPROACHES TO DEFINING "ENDANGERED" 39

Population Viability Analysis 39
IUCN, CITES and AFS Rule-Based Approaches 40
IUCN approach 40
CITES approach 41

AFS approach 41
Recent developments 42
Point-scoring, Hybrid Approaches, and the Heritage Ranking Process 44
Description of point-scoring and hybrid methods 44
Examples of point-scoring methods 44
Strengths and weaknesses of point-scoring methods 45
Hybrid rule-based and point-scoring: the Heritage ranking method 45
Structured Expert Opinion 47
Introduction to use of expert opinion 47
Description of structured expert opinion 47
Examples of the use of structured expert opinion 49
Strengths and weaknesses of structured expert opinion 53
Research and development needs for structured expert opinion 53

APPENDIX 3: ALTERNATIVE DEFINITIONS OF "EXTINCTION RISK" **61**
Species Dynamics 61
Probability Functions 61
Weighting Functions 62
Description of Alternative Approaches 63
Traditional Approach 63
Threshold Approach 64
Comprehensive Threshold Approach 64
Evaluation of Alternative Approaches 64
Well established in the conservation biology literature 65
Directly related to extinction 65
Capable of giving some weight to all possible outcomes 65
Free from assumptions regarding depensatory dynamics 65
Free from explicit policy decisions regarding an "endangered" risk level 65
Free from explicit policy decisions regarding other parameter values 66

APPENDIX 4: PROBABILITY OF EXTINCTION THRESHOLD EXAMPLE – MODEL DETAILS **73**
Pinnipeds 73
Cetaceans 74
Results 74

APPENDIX 5: BASELINE EXTINCTION RATES **81**

APPENDIX 6: USE OF PERFORMANCE TESTING TO ADOPT A SET OF DECISION METRICS **83**

INTRODUCTION

The Endangered Species Act of 1973 (ESA) employs a two-category system: listing species either as endangered (in danger of extinction throughout all or a significant portion of its range) or threatened (likely to become endangered in the foreseeable future). Absence of Congressional guidance on how to interpret the terms used in the statutory definitions of these categories has left the task of defining them to the U.S. Fish and Wildlife Service (USFWS) and NOAA Fisheries (National Marine Fisheries Service), which are the Federal agencies responsible for listing and delisting species under the ESA. To date, neither of these agencies has developed uniform guidelines for listing, reclassifying, or delisting species. The lack of uniform guidelines for listing decisions has led to inconsistencies and inequities in the listing process. For example, Appendix 1 reviews the lessons learned from USFWS's experience in implementing the Act. NOAA Fisheries responded to this problem by establishing a Steering Committee and a Quantitative Working Group (QWG) to work toward developing quantitative procedures that will make listing decisions "more transparent, consistent, and scientifically and legally defensible". The QWG offers the present report as a possible roadmap by which NOAA Fisheries could eventually develop uniform guidelines for listing, reclassifying, or delisting species.

Terms of Reference

The Steering Committee established the following guiding principles for the QWG.

Listing criteria should include the following:

- Be applicable to all species, subspecies, and distinct population segments (DPS), first to NMFS species and then to all species (including species managed by the USFWS)
- Be implementable by the agency and transparent to the public
- Be written in such a way as to provide flexibility for unique circumstances or biology
- Ensure different taxa receive equivalent levels of protection
- Cover data-rich and data-poor conditions
- Include definitions for terms used
- Address adding, downlisting, and uplisting species under the ESA.

The QWG recognizes that any set of explicit criteria used by NOAA Fisheries and the USFWS to make ESA listing determinations could also be applied in reverse as criteria for downlisting or delisting. However, some members of the QWG believe that a greater degree of precaution should be exercised for downlistings and delistings to adequately minimize the probability of inappropriately removing the protections offered to listed species under the ESA. Others believe that a species on its way "up" should not necessarily require greater protection than a species on its way "down". The QWG therefore recommends that NOAA Fisheries and the USFWS consider the issue of delisting and downlisting criteria separately from the listing criteria issue addressed in this report.

Structure of the Report

The primary purpose of this report is to advance the development of quantitative methods for use in making ESA listing decisions. In brief, we propose the following process: (1) *overarching definitions* for both endangered and threatened should be adopted, (2) values of any *policy parameters* associated with the overarching definitions (e.g., the level of extinction risk corresponding to "endangered") should be specified, (3) *decision metrics* that can be used as proxies for (1) and (2) in data-poor cases should be developed for an appropriate range of taxonomic groups or life history types, and (4) all of the above should be done in the context of *performance testing* (use of output from simulations to evaluate how well an alternative performs relative to the objective).

We begin by introducing concepts and terms important to understanding the recommendations contained in the remainder of the report. This leads to descriptions of three alternative overarching definitions for "endangered" (EN), including a discussion of any policy parameters associated with each. An overarching definition together with values of any associated policy parameters constitutes a *listing criterion*. The descriptions of the overarching definitions are followed by a brief discussion of the strengths and weaknesses of each definition.

Once a listing criterion is adopted, it can be used directly to make listing decisions, provided that sufficient data and other resources are available. Often, however, this will not be the case, in which case other decision metrics must be used. A vast array of such metrics has been proposed in the past (Appendix 2), and several possible combinations are described later as examples.

We have made a preliminary recommendation for a specific listing criterion for TH, which is conditional on the listing criterion for EN, as the statutory language requires. The QWG was not able to agree unanimously on this definition. A final decision on a listing criterion for TH should only be made after the performance testing phase of this project has been completed.

We recommend that final decisions regarding listing criteria and decision metrics be made in the context of performance testing. We further recommend that this be done in two main phases. The first phase involves performance testing of alternative listing criteria, and we present specific steps for one possible approach to this task. The second phase involves performance testing of alternative decision metrics, and we present specific steps for one possible approach to this task.

We conclude with a summary of the QWG's recommendations, including a timeline, an interim protocol for use until such time as the performance testing can be completed, the establishment of a committee authorized to make final choices of listing criteria and decision metrics, and the establishment of a working group to enhance the use of structured decision making in the ESA listing process.

TREATMENT OF "ENDANGERED" STATUS

Definitions of "Endangered"

Basic Concepts

The extinction process can be complex. Although some species may go extinct quickly because a major portion of their habitat has been eliminated or because of overexploitation by humans (e.g., passenger pigeons), others may drift slowly to a low level of abundance where they persist for a long period of time before going extinct. Some factors tend to *increase* population growth rates at low levels of abundance, a phenomenon known as "compensation." For example, population growth rates can increase at low levels of abundance because there are more food resources per individual. Likewise, population growth rates can increase at low levels of abundance because disease transmissibility can be reduced if animals are not in close proximity or are not stressed due to food limitations. When compensatory factors prevail and anthropogenic forcing is eliminated, extinction tends to be avoided. Other factors, however, tend to *decrease* population growth rates at low levels of abundance, a phenomenon known as "depensation." For example, it can be more difficult for individuals to find mates at low levels of abundance. The gene pool tends to be smaller at low levels of abundance, which can result in a loss of average fitness. Also, at low levels of abundance, a species is likely to be composed of one or only a few populations, making the species more vulnerable to catastrophic events such as floods or droughts. When depensatory factors prevail, even with the elimination of anthropogenic factors, the species tends toward extinction. The abundance level below which depensatory factors prevail is called the *depensatory threshold* (in cases where there is no abundance level below which depensatory factors prevail, the depensatory threshold is zero).

Among the things that complicate attempts to model the process of extinction is random natural variability. While it might be possible to compute the most likely time for extinction to occur, this estimate will almost always be highly uncertain. In other words, extinction could occur at any time, but some extinction times are more probable than others. This idea is conveniently expressed by a *probability function*. A probability function shows the relative likelihood of alternative states or outcomes. Figure 1 shows how the probability of extinction varies with time for two hypothetical Species A and B. Figure 1 shows that an extinction time of about 25 years is more likely than any other value for Species A, while an extinction time of about 20 years is more likely than any other for Species B. The area under any given portion of a probability function represents the probability associated with the corresponding interval on the horizontal axis. In the special case where an interval starts at the lower end of the horizontal axis, the probability associated with the interval is called the *cumulative probability* (or, in context, just "probability"). Figure 2 shows the cumulative probability that extinction will occur before any given point along the time axis for species A and B. Despite the differences in the extinction function, there is a 5% chance that Species A will go extinct within 20 years and a 5% chance that Species B will go extinct within 20 years.

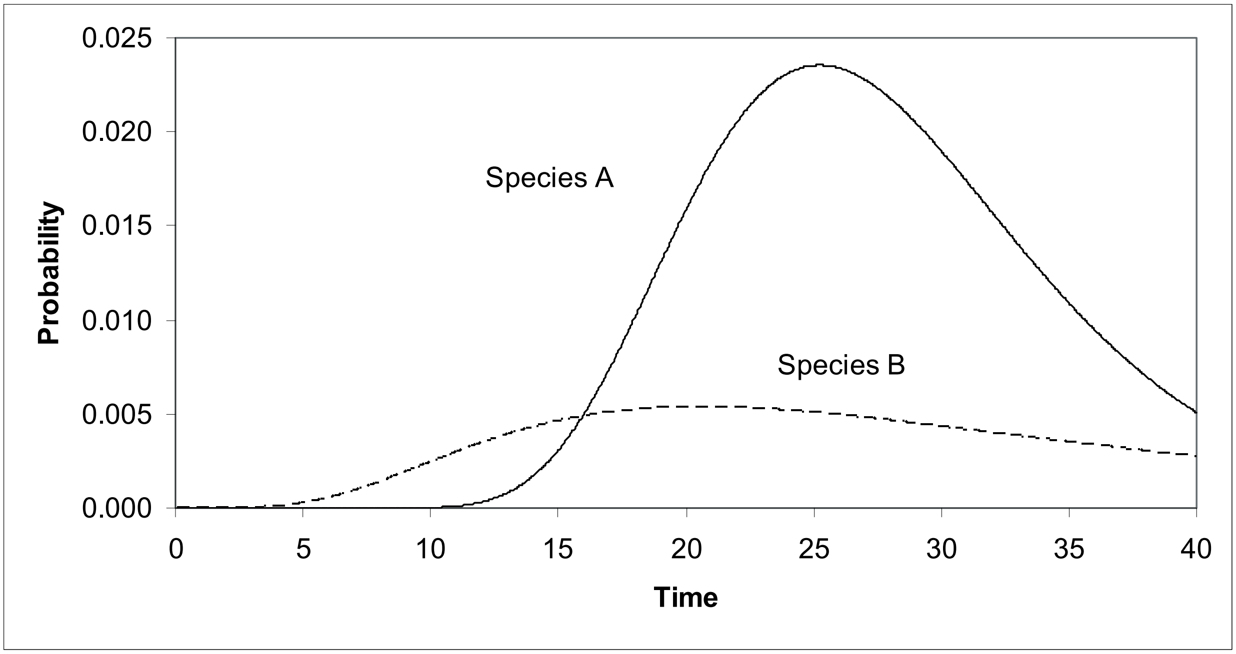


Figure 1. Extinction time probability functions for species A (solid) and species B (dashed).

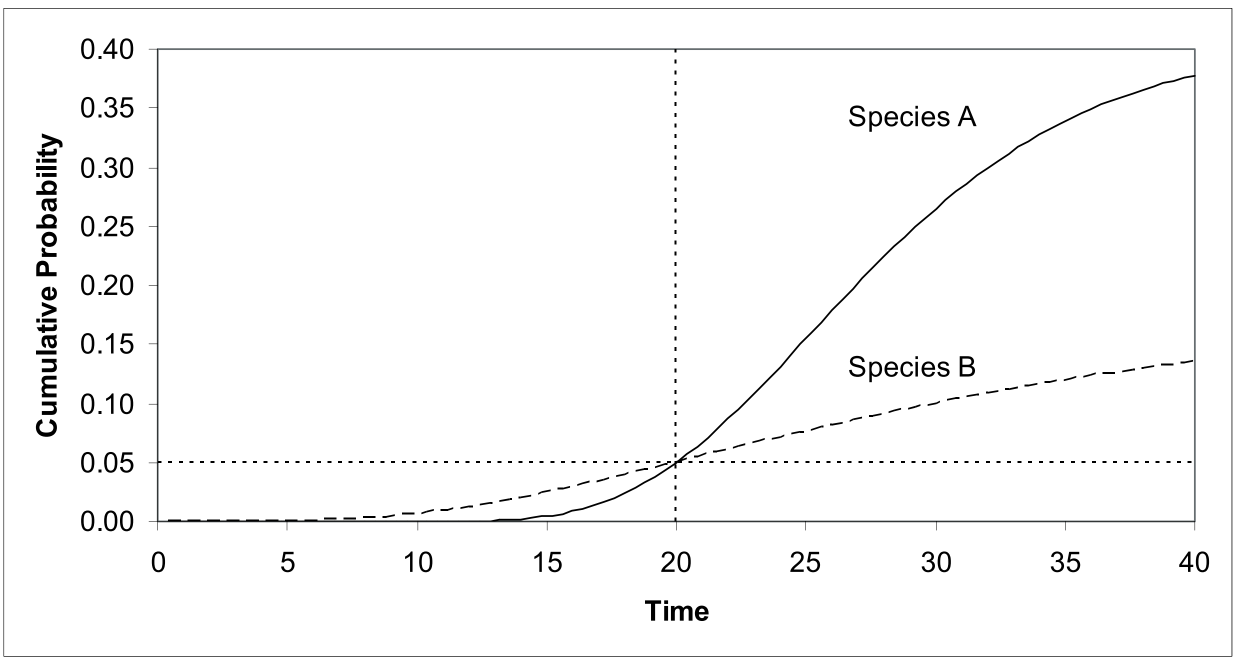


Figure 2. Cumulative probability of extinction as a function of time for species A (solid) and species B (dashed).

Extinction time probability functions are estimated using population viability analysis (PVA). Population viability analysis is a mathematical modeling approach that uses data on the species in question along with prior knowledge gained from similar species to project the species into the future together with all the uncertainties it faces. For example, suppose that a species has recently been reduced to a single small population and suppose further that this population is subject to occasional droughts. Because of the species' current status, suppose that any one of these occasional droughts is of a magnitude sufficient to result in the species' immediate extinction. Finally, suppose that these droughts occur randomly and that the average "waiting time" until the next drought is 30 years. In this hypothetical scenario, we know that extinction will almost certainly occur within the next 100 years, but the exact date of extinction is unknown because the droughts occur randomly. To quantify the extinction probability over time, we can project the population forward in time, each year drawing randomly from the appropriate statistical distribution to determine whether a drought occurs in that year, until the species becomes extinct. If this simulation is repeated thousands of times, we get a distribution of extinctions over time that can be used, for example, to compute the cumulative probability of extinction within any chosen time horizon.

The probability of extinction is often correlated with a number of variables. One is simply abundance itself. Other variables include trends in abundance, current abundance relative to historical abundance, proportion of historical habitat remaining, and the level of habitat fragmentation. Sometimes particular values of these variables are used as proxies for a particular level of extinction probability. When used in this manner, the appropriate values may depend on a number of modifying factors, such as: whether the cause of a decline is known, whether the cause of a decline is reversible, and whether the species has only a single remaining population within a small range.

Description of the Alternative Overarching Definitions

The QWG proposes three alternative overarching definitions for EN:

- The Probability of Extinction Threshold definition: A species is endangered if the probability of extinction within Y years is greater than X in Y years or Z generations, whichever is longer up to a maximum time period of 100 years (X, Y and Z are policy parameters whose values need to be specified). The use of 100 years as the maximum time period in the definition of EN reflects two conflicting requirements in implementing this definition. The first has to do with the assumption that conditions will remain constant such that the projections will provide reliable results. Ideally, projections would never have to go beyond 10-25 years. The second has to do with the time necessary for the dynamics of long-lived species to equilibrate after a significant perturbation. Time periods of hundreds of years are necessary for long-lived species (e.g., over 100 years) or species which mature only after 20 years or more of age. The choice of 100 years was therefore a compromise on the part of the QWG. The robustness of this definition (and the other two definitions) to this parameter value should be investigated in the performance testing phase and adjusted as necessary.

- The Depensatory Threshold definition: A species is endangered if it declines to a taxon-specific threshold below which the demographic and genetic behavior of a population becomes highly uncertain and depensatory effects that heighten extinction risk are likely.
- The Comprehensive Threshold definition: A species is endangered if the extinction risk, computed by decision-theoretic methods, is greater than X (X is a policy parameter whose value needs to be specified; in addition, computation of extinction risk involves a weighting function governed by one or more policy parameters whose values need to be specified. The weighting function allows for the integration of 1) the risk of extinction and 2) the loss to society associated with extinction into a single measure by which an ESA listing determination can be made. See Appendix 3 for more details).

Note that the Probability of Extinction Threshold definition and the Comprehensive Threshold definition are each associated with explicit policy parameters. These are called policy parameters because their values are set as a matter of policy. The Depensation Threshold definition also requires policymakers to provide guidance to scientists implementing this approach, but here the input is implicit, as no specific policy related parameters are required to implement this definition. Scientists may be able to describe the consequences of alternative values for such parameters and they may be able to estimate which values best reflect societal preferences, but non-policy making scientists should not be put in the position of choosing which values to use. For ease of reference, the combination of a particular overarching definition together with a set of specific values for any associated policy parameters will be called a "listing criterion." For example, the Probability of Extinction Threshold definition with X = 0.05 and Y = 100 years constitutes one possible listing criterion.

An explanation of the logic for the three definitions follows.

Two main premises underlie the Probability of Extinction Threshold definition. First, the statutory definition of "endangered" as "in danger of extinction" implies that decisions to list or not to list species as EN should involve some measure of extinction risk, which in turn should have something to do with the probability of extinction. Second, the fact that most people are uncomfortable dealing with formal probabilistic concepts suggests that any use of probability in listing decisions should be kept as simple as possible. Taken together, these premises suggest that extinction risk should be equated with the cumulative probability of extinction at a single point in time.

Two main premises also underlie the Depensatory Threshold definition. First, the *quantitative* dynamics of species at very low abundance levels are extremely difficult to estimate. Second, the *qualitative* dynamics of species at very low abundance levels are often associated with the phenomenon of depensation, where the central tendency of the species' abundance trajectory is toward extinction. Taken together, these two premises suggest that there is little hope of obtaining precise estimates of extinction probability when species are at very low levels of abundance, but there is a very good chance that the extinction probability is high enough to warrant an EN listing.

As with the other overarching definitions, two main premises also underlie the Comprehensive Threshold definition. The first is identical to the first premise underlying the Probability of Extinction Threshold definition. The second is that optimal statistical decisions are obtained only when "risk" involves both the *probability* of each adverse outcome (such as extinction at each future point in time) and the *value* of each adverse outcome. This is accomplished by use of a function that assigns a weight to each possible outcome (i.e., extinction at each future point in time), thereby providing a comprehensive measure of extinction risk.

In the interest of making some of the technical concepts referenced in this section more accessible to graphically oriented readers, Appendix 3 presents another way of approaching the problem. Instead of looking at alternative overarching definitions of "endangered," Appendix 3 describes three alternative definitions of "extinction risk." However, these definitions do not necessarily map into the definitions of endangered described in this section. The definitions of extinction risk described in Appendix 3 represent possible, though not always necessary, counterparts to the definitions of "endangered" described in this section.

Evaluation of the Alternative Overarching Definitions

Characteristics desired under the guiding principles provided by the Steering Committee include applicability, implementability, transparency, flexibility, and equitability. Each of the alternative overarching definitions exhibits all of these characteristics when applied to an ESA listing determination for a "data rich" species. Further, each of the alternative overarching definitions is expected to be broadly applicable across species and a wide range of data availability levels. Finally, each of the definitions is expected to be fully implementable provided that the recommendations contained in this report are followed. The other three characteristics desired under the Guiding Principles may require more explanation, so these are addressed individually below.

Transparency -- Here the QWG uses the term "transparency" to refer to the degree to which a given approach can be understood by the general public. Unfortunately, the lack of policy guidance in the ESA makes it very difficult for either the USFWS or NOAA Fisheries to develop criteria for classification of a species at risk under the ESA that has irrefutable rationale. It is anticipated that whatever approach is selected for classifying a species, considerable effort to explain the approach to the general public will be required. However, at a minimum, knowledge of the protocol used by NOAA Fisheries or USFWS should be available to the general public.

Flexibility -- Because they attempt to consider all aspects of species dynamics, the Probability of Extinction Threshold definition and the Comprehensive Threshold definition are likely to be more flexible than the Depensation Threshold definition. That is, complex features of the dynamics of a species can be incorporated into the model used to evaluate risk of extinction. The Depensatory Threshold definition incorporates relevant aspects of species dynamics in the form of factors that might modify the location of the threshold, but this is conducted in an indirect manner.

Equitability -- If "equitability" is interpreted to be synonymous with "applicability," then all three overarching definitions are fully equitable. However, if "equitability" is interpreted to mean something like, "the definition results in identical listing decisions for species with identical estimates of risk of extinction," then differences between the three overarching definitions emerge. Either the Probability of Extinction Threshold definition or the Comprehensive Threshold definition can be fully equitable in estimating the risk of extinction, depending on the definition of extinction risk (see Appendix 3). That is, the use of either definition for EN would result in species with equal risk of extinction having consistent protection under the ESA (based on their listing status). The Depensatory Threshold definition, in contrast, does not relate directly to extinction risk and so must be inequitable to some extent; however, to the extent that a population that is experiencing depensation is reasonably likely to go extinct, the associated bias should be insignificant. As noted above, performance testing is necessary to ensure that the equitability standard is achieved.

Several characteristics not mentioned in the Guiding Principles are also useful to consider. These are discussed in the paragraphs below.

Well established in the conservation biology literature -- The Probability of Extinction Threshold definition and the Depensatory Threshold definition are familiar concepts in the conservation biology literature, whereas the Comprehensive Threshold definition is not. Similarly, the Probability of Extinction Threshold definition and the Depensatory Threshold definition are reminiscent of hypothesis testing, and so will be easily communicated to practitioners of that particular methodology. On the other hand, the Comprehensive Threshold definition is firmly rooted in decision theory, and so will be easily communicated to practitioners of that particular methodology.

Free from assumptions regarding depensatory dynamics -- The Depensatory Threshold definition is the only definition that does not require modeling of species dynamics at abundances below the depensatory threshold, although it does require estimating the location of the depensatory threshold. The Probability of Extinction Threshold definition and the Comprehensive Threshold definition likewise require estimating the location of the depensatory threshold, at least implicitly, but they also require estimating how the species will respond in the event that it falls below the depensatory threshold.

Capable of reflecting societal attitudes toward extinction -- This characteristic is in some sense the converse of the preceding characteristic. The association of one or more policy parameters with an overarching definition confers some ability to reflect societal attitudes (at least as reflected through our governance system) toward extinction. At one extreme, the weighting function used by the Comprehensive Threshold definition can take whatever shape is needed to reflect society's attitudes toward how much risk of extinction for a given species should be tolerated before conservation measures mandated by the ESA are initiated. The Probability of Extinction Threshold definition is at least somewhat capable of reflecting such attitudes because it involves two policy parameters, but it necessarily regards some range of future outcomes as equally important and the remaining future outcomes as completely unimportant, which limits the definition's capability. The Depensatory Threshold definition involves implicit policy parameters and therefore it will likely be more difficult to reflect directly societal attitudes toward balancing the cost of conservation measures with the risk of extinction. However, the degree to which precaution is incorporated in the method used to select the depensation threshold could also be a vehicle for incorporating societal values regarding how much risk of extinction is to be tolerated prior to listing a species as EN.

Decision Metrics That Could be Used to Approximate the Listing Criteria

Given sufficient data and other resources, any listing criterion (i.e., an overarching definition together with specified values of any associated policy parameters) is sufficient to make a decision as to whether a petitioned species should be listed as EN (here species is used as in the context of the ESA). For example, if the listing criterion consists of the Probability of Extinction Threshold definition with $X = 0.05$ and $Y = 100$ years and there are sufficient data and other resources to conduct a reliable PVA, the output of that PVA can be plugged into the listing criterion directly to determine whether an EN listing is warranted. Likewise, if the listing criterion consists of the Depensatory Threshold definition and there are sufficient data and other resources to estimate the current status of the species with respect to a relevant depensatory threshold (which could, for example, be expressed as a percentage decline relative to historical levels), these estimates could be used directly in using the listing criterion to determine whether an EN listing is warranted.

Often, however, data or other resources will be insufficient to apply a listing criterion directly, in which case some sort of proxy must be used. Examples include some specified level of absolute abundance or a specific risk of extinction. A vast array of such decision metrics could be developed and used either singly or in combination to implement any listing criterion in data-poor cases. It is anticipated that such decision metrics will need to be developed on the basis of individual taxonomic groups or life history types (i.e., "one size fits all" is unlikely to work). The following sections below describe in detail how several alternative decision metrics could be used to implement the Probability of Extinction Threshold definition and the Depensatory Threshold definition in cases where data or other resources are insufficient to apply the respective listing criterion directly. In the interest of brevity, a comparable discussion is not provided for the Comprehensive Threshold definition, because any of the decision metrics described for the other two overarching definitions could also be used as decision metrics under the Comprehensive Threshold definition.

It should be emphasized that the particular values used for decision metrics in the next two subsections are intended primarily to serve as examples. The QWG recommends that actual values of decision metrics be chosen on the basis of performance testing.

Probability of Extinction Threshold Definition

An example of how proxy criteria could be used in the absence of having all of the data needed to implement the Probability of Extinction Threshold definition would be the use of abundance or trends in abundance, alone or in combination, to indicate risk. These measures do not account for all risk factors but do strongly indicate general levels of risk. By tying the proxy criteria to the overarching definitions through proxy PVAs, the goal of equal treatment for all species can be maintained.

Appendix 4 describes an example of how decision metrics can be developed for the Probability of Extinction Threshold definition. This example uses the following listing criterion: the species is endangered if the probability of extinction is greater than or equal to 1% in 100 years. As noted above, the choice of specific values for policy parameters is only intended to serve as an example, and should not bias the selection process which is to be based on performance testing.

The example described in Appendix 4 also demonstrates another technique to enhance use of PVA-based criteria: the use of an extinction threshold. The extinction threshold is used to mitigate modeling difficulties that are peculiar to the population dynamics of very small populations. When populations become very small, growth rates are affected negatively by the combination of genetic, behavioral and demographic problems. There are almost never species-specific data that can be used to estimate parameters for PVA models of such small populations. There are two approaches to treating the dynamics of very small populations in a PVA model: the default approach and the extinction threshold approach. The default approach uses default parameters taken from other species in PVA models and then projects the species forward in time until it reaches extinction (i.e., defined in most PVA models as the point at which there is no possibility of further reproduction). The extinction threshold approach defines "extinction" as the point at which extinction becomes so likely that drastic conservation measures, such as captive breeding, would be needed to preserve the species. Without such intervention, the likelihood of extinction in a short period of time is very high and thus the time between reaching the threshold and reaching actual extinction would typically be short. When populations reach this extinction threshold, biologists expect substantially different behavior from the population as a series of factors interact to further increase extinction risk. The objective of redefining "extinct" as "reaching an extinction threshold" is to put a high value on avoiding population levels where negative (and unpredictable) feedback loops might come into play. It is at this level that extinction risk becomes so high that recovery efforts may be unable to save the species. The other benefit of this redefinition is that few, if any, data exist for the dynamics of most species at extremely low abundance. Most biologists are very uncomfortable creating models for a particular species without any data. Further, without supporting data it is likely that the model outcome will depend substantially on the assumptions the modeler is willing to make and thus would likely not be repeatable among modelers.

In the Appendix 4 example, a set of proxies are developed by creating two PVAs for two life history types (in this example, for marine mammal life histories). We follow the procedure developed for implementing the Marine Mammal Protection Act where the choice of numbers to be used in calculating the allowable catch was made on the basis of some simplified cases. In a similar fashion, we develop PVAs for the most common marine mammal life histories: pinnipeds and cetaceans. We then use those PVAs as default risk assessments to choose the decision metrics that correspond to the listing criterion. This approach is similar to one recommended by the QWG.

The following are examples of decision metrics, where performance tests seek to quantify the unspecified variables (labeled n_1, n_2, and n_3 and X and Y):

A. Population size estimated to number fewer than n_1 mature individuals and either:

1. An observed, estimated, inferred or suspected population size reduction of $\geq n_2\%$ over the last 15 years, where the causes of the reduction are clearly reversible and understood and no longer in existence, based on (and specifying) any of the following:
   (a) direct observation
   (b) an index of abundance appropriate to the taxon
   (c) trend estimated by an appropriate growth rate minus an estimated human-caused mortality rate (such as a bycatch rate)
2. An observed, estimated, inferred, projected or suspected population size reduction of $\geq n_2\%$ - 10 % over any 15-year period, where the time period includes both the past and the future, and where the reduction or its causes may not have ceased OR may not be understood OR may not be reversible, based on (and specifying) any of (a) to (e) under A1.

B. Population size estimated to number fewer than n_3 mature individuals.

C. Quantitative analysis showing the probability of extinction in the wild is at least X% within Y years.

Depensatory Threshold Definition

The following proposal is based on the IUCN, CITES and AFS rule-based approaches, along with recommendations from Mace et al. (2002) and Appendix 2 of this report, and is meant to be applied to all taxonomic groups, with appropriate customizing for each group. As discussed in Appendix 2, this proposal is based on the following common observations about life history strategies of marine fish: (i) the empirical record shows that species with high productivity are much less likely to become extinct than species with low productivity (although local extinctions of small populations may occur, it is difficult to think of many high-productivity species that have become extinct in recent times), and (ii) high productivity, high variability and high absolute numbers are usually positively correlated.

The proposed approach considers four different decision metrics that could be used to implement the Depensatory Threshold definition. The decision metrics are as follows: a marked historical extent of decline, small population size, restricted area of distribution, and quantitative population dynamics analysis indicating a high risk of extinction in the foreseeable future. For each decision metric, numeric thresholds that would trigger urgent concern are specified. Modifying factors that would exacerbate or mitigate the degree of concern are also identified. However, all numbers used to quantify decision metrics are simply placeholders at present and further research is needed to improve upon them. As noted previously, the results of the performance testing would be needed to actually demonstrate a reliable link between a given metric and the definition of EN under the ESA.

The order of the proposed decision metrics does not reflect any preference for the use of one metric over another. This will depend on the data and knowledge of population processes available for the species under consideration.

Metric A. Marked Historical Extent of Decline [†]

The historical extent of decline is the total estimated or inferred percentage reduction from an appropriate baseline[*]. Metrics potentially relevant for measuring or indexing the historical extent of decline include:

- numbers (of individual organisms in a population or subpopulation)
- biomass (total weight of a population or subpopulation)
- area inhabited (area of distribution)
- migratory range (for highly migratory species)
- percentage coverage, or other index of population density (for sessile species)
- relative spawning per recruit
- numbers or biomass of new recruits (recruitment)

Based on Mace et al. (2002), the threshold for "listing" would be reached when the extent of decline, expressed as a percentage decline relative to an appropriate baseline, reaches a specific value. The recommended values from the Mace et al. (2002) report are presented here as an example of how this method could be implemented. As was the case for the Probability of Extinction Definition, the actual values to be used would have to be determined based on the results of the performance testing. In this case, the threshold for "listing" would be reached when the extent of decline reaches 1%-20% [§] of the baseline, depending on the productivity of the species (see Fig. 3). The lower end of the range (i.e., a population estimated or inferred to be near 1% of baseline levels, meaning that it has declined by 99%) is applicable for species with very high productivity, as indicated by factors such as high intrinsic rates of natural increase, short generation times, early maturity, high fecundity, rapid growth rates and high natural mortality. The upper end of the range (i.e., a population estimated or inferred to be near 20% of baseline levels, meaning that it has declined by 80%) is applicable for species with very low productivity, as indicated by factors such as low intrinsic rates of natural increase, long generation times, late maturity, low fecundity, slow growth rates and low natural mortality. For many taxonomic groups, it will likely be appropriate to select a narrower range within the 1-20% guideline. For example, for commercially exploited marine fish and invertebrates, a range of 1-5% may be appropriate for many of the more highly productive species, 5-10% for species with medium productivity and 10-15% for species with low productivity. The 15-20% range would apply only to species with very low productivity and would rarely be invoked for commercially exploited marine taxa. For some taxa with relatively low productivity, such as large terrestrial vertebrates, a range of 10-20% may be more appropriate. Even though the extremes of 1% and 20% will be applicable to only a relatively small number of species, some species may fall outside of these extremes.

Fig. 3. From Mace et al. (2002: p. 15). The table summarizes recommendations for thresholds used to assign species to CITES Appendices. The percentages in the right three columns represent the expected decline relative to the status of the population in a given row after an annual decline over 10 years as specified in the bracketed percentages.

Current population as percent unexploited	**Productivity**		
	Low	Medium	High
100%	70% (11.3%)	80% (14.9%)	95% (25.9%)
90%	67% (10.4%)	78% (14.0%)	94% (25.1%)
80%	63% (9.3%)	75% (12.9%)	94% (24.2%)
70%	57% (8.1%)	71% (11.8%)	93% (23.2%)
60%	50% (6.7%)	67% (10.4%)	92% (22.0%)
50%	40% (5.0%)	60% (8.8%)	90% (20.6%)
40%	25% (2.8%)	50% (6.7%)	88% (18.8%)
30%	0%	33% (4.0%)	83% (16.4%)
20%	0%	0%	75% (12.9%)
15%	0%	0%	67% (10.4%)
10%	0%	0%	50% (6.7%)
5%	0%	0%	0%

Metric B. Small Population Size †

A general guideline for a small wild population is a range of 100 to 100,000 individuals, depending on the productivity of the species. The upper end of this range is applicable for species with high productivity as indicated by factors such as high intrinsic rates of natural increase, short generation times, early maturity, high fecundity, rapid growth rates and high natural mortality. The lower end of the range is applicable for species with low productivity, as indicated by factors such as low intrinsic rates of natural increase, long generation times, late maturity, low fecundity, slow growth rates and low natural mortality. The reason that high-productivity species should have higher threshold absolute numbers is because high productivity is usually associated with high variability. Therefore, if everything else is equal, high productivity species are likely to have a higher probability of extinction than low productivity species, simply because of random natural variability (Appendix 4). For many taxonomic groups, it will likely be appropriate to select a narrower range within the 100-100,000 guideline. Even though the extremes of 100 and 100,000 will be applicable to only a relatively small number of species, some species may fall outside of these extremes.

Metric C. Restricted Area of Distribution †

A general guideline for a restricted area of distribution is 10,000 km^2 (based on IUCN and CITES numbers). At present, thresholds for area of distribution cannot be related to population productivity or any other specific life history attribute, except that populations with high productivity, which will usually also have high variability, will likely require larger threshold areas of distribution. In general, the historical extent of decline of the area of distribution or absolute numbers may be more useful indicators of extinction risk than the absolute area of distribution alone. Alternatively, the absolute area of distribution could be considered in conjunction with Metrics A and B.

Metric D. Quantitative Population Dynamics Analysis Indicating a High Risk of Extinction in the Foreseeable Future

A general guideline for an unacceptably high risk of extinction within the foreseeable future might be, for example, a 5% probability within the next 20 years. Quantitative projections of future population size will often provide a better basis for decision-making than rule-based approaches, but the reverse can also be true.

---

*The data used to estimate or infer a baseline for the historical extent of decline should extend as far back into the past as possible. However, different baselines may be appropriate in different situations. Depending on the species under consideration, the baseline may relate to some point in history, or to a *reasonable* or *potential* baseline given alterations to the environment that have affected current carrying capacity. Use of *reasonable* or *potential* baselines reflects, respectively, the reality that habitat changes have occurred in the past, and the possibility that such changes may be wholly or partially reversible. However, if the potential baseline is very small due to dramatic reductions in the carrying capacity of the habitat over time, it then becomes necessary to ask whether the current carrying capacity is adequate to ensure survival of the species. The time frame over which to examine the historical extent of decline should be as long as possible to enable a meaningful baseline to be chosen.

§ This range is based on a survey of the population dynamics literature for commercially-exploited species (Mace et al. 2002), with additional considerations for species with very high or very low productivity.

†Modifying factors (Table 1) may increase or decrease the threshold that triggers concern about extinction risk for this criterion.

Table 1: Modifying factors to apply to the numeric guidelines for Metrics A-D (from Mace et al. 2002).

(i) Vulnerability factors that would increase concern

- Life history characteristics (e.g., low fecundity, slow growth rates, high age at first maturity, long generation time)
- Selectivity of removals
- Distorted age, size or stage structure of a population
- Social structure, including sex ratio
- Low population density (especially for sessile or semi-sessile species)
- Specialized niche requirements (e.g., diet and habitat)
- Species associations such as symbiosis and other forms of co-dependency
- Strong aggregating behavior (e.g., schooling)
- Extensive migrations
- Secondary ecosystem-based effects
- Uncertainty
- Fragmentation
- Reduced genetic diversity
- Severe habitat loss
- Degree of endemism
- Existence of disease
- Existence of invasive species
- Existence of rapid environmental change (e.g., unfavorable climate regime shifts)

(ii) Mitigating factors that would decrease concern

- Life history characteristics (e.g., high fecundity, rapid growth rates, low age at first maturity, short generation time)
- Existence of natural refugia
- Adaptations to small population size
- Selectivity of removals

Relationship between the Depensatory Threshold decision metrics -- Ideally, the decision metrics should map to the listing criterion. Achieving such a mapping is one of the two main purposes of performance testing. However, the paucity of empirical observations at low population sizes may make this difficult except for a few specific cases. The difficulty of developing meaningful generic relationships between the decision metrics will also be exacerbated by the large diversity of combinations of life history characteristics exhibited by the world's flora and fauna. Thus, it may be necessary to develop different relationships for different taxa. An example of the type of relationship that might roughly apply for a range of animal taxa is shown for percent decline and population size in Figure 4. Here it is assumed that high productivity is usually associated with high natural population size (i.e., high productivity populations not impacted by humans will usually be large), and therefore a greater ability to withstand a given percentage decline in size, as compared to populations with low productivity. However, high productivity is usually also associated with high variability, meaning that, for a given population level, the stochastic risk of extinction may be higher. Thus, while populations with high productivity may be able to withstand larger percentage declines in size, they may also need larger absolute numbers to ensure survival.

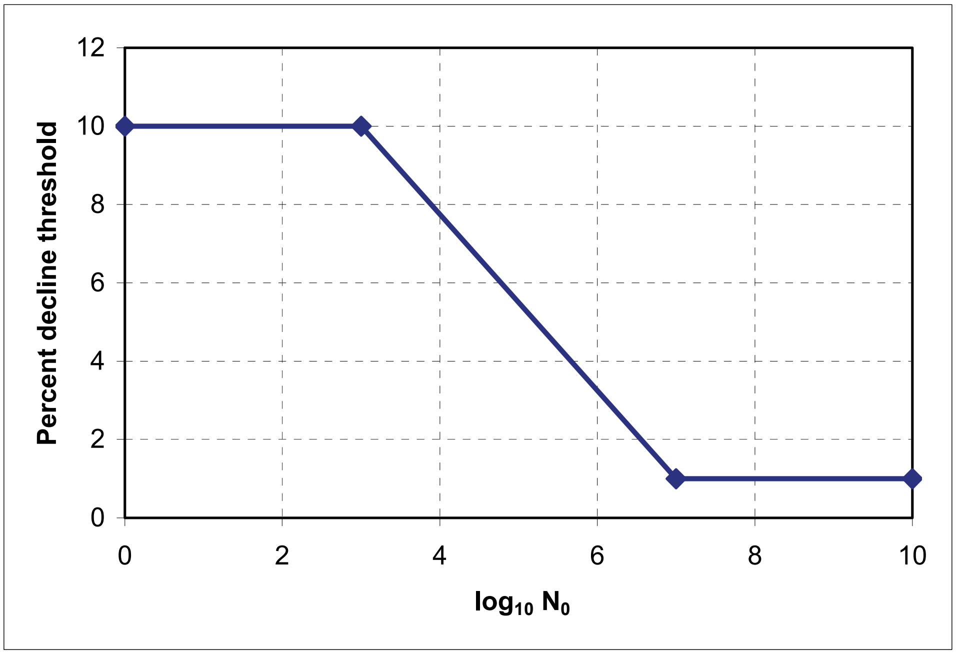


Figure 4. A possible relationship between percent decline thresholds and population size. N_0 is the expected population size for a population unimpacted, or not substantially impacted, by human influence. Thresholds for "small population size" can be calculated by applying the percent decline threshold to the unimpacted population size.

It may also be possible to develop a relationship between percent decline and area of distribution, and between population size and area of distribution, although it is likely that there are even fewer data to quantify such relationships.

Research and analysis needs prior to performance testing of the Depensatory Threshold definition -- Taxonomic working groups need to determine appropriate alternative values of Metrics A-D so that these can be subjected to performance testing. The first step would be to conduct a comprehensive survey of the scientific literature to determine the historical extents of decline, population sizes and areas of distribution that have preceded extinctions of populations or species. If sufficient data exist, a meta-analysis of those data could be performed with the results being used to quantify the depensatory thresholds. However, for most taxa, there is unlikely to be a sufficient number of reliable estimates to enable reliable specification of the appropriate ranges. Therefore, "informed scientific judgment" will be needed to develop analogies between similar species or similar life history characteristics in order to specify default ranges for the metrics. Such specifications could be updated as new information becomes available.

The fact that different decision metrics may give disparate results will need to be dealt with. The QWG does not advocate the IUCN approach of simply choosing the most pessimistic result. When data exist for more than one decision metric, the decision metrics must be considered in conjunction with one another. Again, "informed scientific judgment" could be used to determine which categorization best reflects the status of the species, or to develop methods for combining information on different decision metrics. Such methods could include points-based systems that weight different criteria and structured expert opinion frameworks (Appendix 2). Regardless of the overarching definition adopted, or the extent to which decision metrics are quantified, a structured expert opinion framework may always be needed to capture and incorporate all relevant considerations related to listing species under the ESA.

Treatment of Uncertainty

Evaluating whether a species satisfies a decision metric will often involve estimates derived from statistical sampling. For example, determining whether a species satisfies a particular decision metric may involve estimates of total abundance, the proportion mature, and the growth rate, each of which will typically be somewhat uncertain. These uncertainties can easily combine to result in a confidence interval that spans the decision metric. The choice of how to incorporate uncertainty involves balancing preferences for different types of errors: under- versus over-protection errors. This "choice" must be made by policymakers and not by the agency's scientists. Once these preferences are defined, the decision metrics can be "tuned" to achieve the desired error balance. Careful treatment of uncertainty is mandatory if species with differing levels of data quality are to be treated equally.

TREATMENT OF "THREATENED" STATUS

The statutory definition of "threatened" is "likely to become endangered in the foreseeable future." The QWG agreed that this definition could be fairly translated as "the probability of becoming endangered within XX years exceeds YY%." Thus, the listing criterion for TH derives directly from whatever listing criterion is adopted for EN. As with EN, the listing criterion for TH can be applied directly if sufficient data and other resources are available, or any number of other decision metrics can be developed for use as proxies in cases where data or other resources are insufficient to apply the listing criterion directly.

Some members of the QWG believe that Congress intended species classified as EN to be "on the brink" of extinction, while species classified as TH were not in danger of extinction in the foreseeable future. At this time, most of the QWG members recommend that NOAA Fisheries consider a species to be appropriately classified as TH whenever its probability of becoming EN within 20 years exceeds 0.5. However, some members of the QWG would have preferred a definition of TH that had a lower probability of becoming endangered and a longer time period over which a species could becoming endangered. Pending the receipt of further guidance from the Steering Committee and from the USFWS, and pending the results of the performance analysis of the approaches described in this report, this recommendation should be considered preliminary. It was also noted that the proposed definition of TH is inconsistent with the definition of TH currently used by the USFWS, as well as the definition of TH used by NOAA Fisheries in classifying Pacific salmon populations under the ESA.

PERFORMANCE TESTING

The QWG recommends that performance testing be conducted in two phases. In the first phase, performance testing would be used to evaluate alternative listing criteria (overarching definitions together with values of any associated policy parameters). On the basis of this test, a single listing criterion would then be chosen, at least provisionally. In the second phase, performance testing would be used to evaluate alternative decision metrics given the chosen listing criterion.

The purpose of performance testing is to evaluate how alternatives perform relative to one or more management objectives. This is accomplished by simulating the performance of the alternatives and using a set of performance measures to translate the simulation output and the objective into a common currency. In the first proposed phase, for example, alternative listing criteria would be evaluated. Here, the objective might be something fairly qualitative, such as, “to meet society’s expectations and tolerance for risk of extinction under the assumption that the protections afforded by the ESA would only be realized following the listing of a species.” This objective could be translated into more quantifiable terms by means of a survey or other preference elicitation exercise. Of course any definition of EN under the ESA must also account for society’s expectations regarding the cost of the associated conservation measures required under a typical ESA listing, as well as other expectations (e.g., following an ESA listing, the risk of extinction would be significantly diminished, the relative ratio of EN to TH species, etc.). Such a performance measure might take the form of a curve showing the expected number of future extinctions as a function of time. Such a curve would necessarily fall somewhere between two limits: One limit would correspond to the case where all species were classified as EN and the other limit would correspond to the case where no species were classified as EN. Historical extinction rates (Appendix 5) can give some guidance as to the location of these limits. Somewhere between the limits is the curve that society would expect to result if the criterion was crafted appropriately. An example is shown in Figure 5. Another performance measure might be the cost to society for achieving a given standard (e.g., 20% in 20 years) in reducing the expected number of extinctions under various alternative definitions of EN. It is critical that a suite of performance measures be developed and agreed upon prior to the completion of this phase of developing quantitative listing criteria.

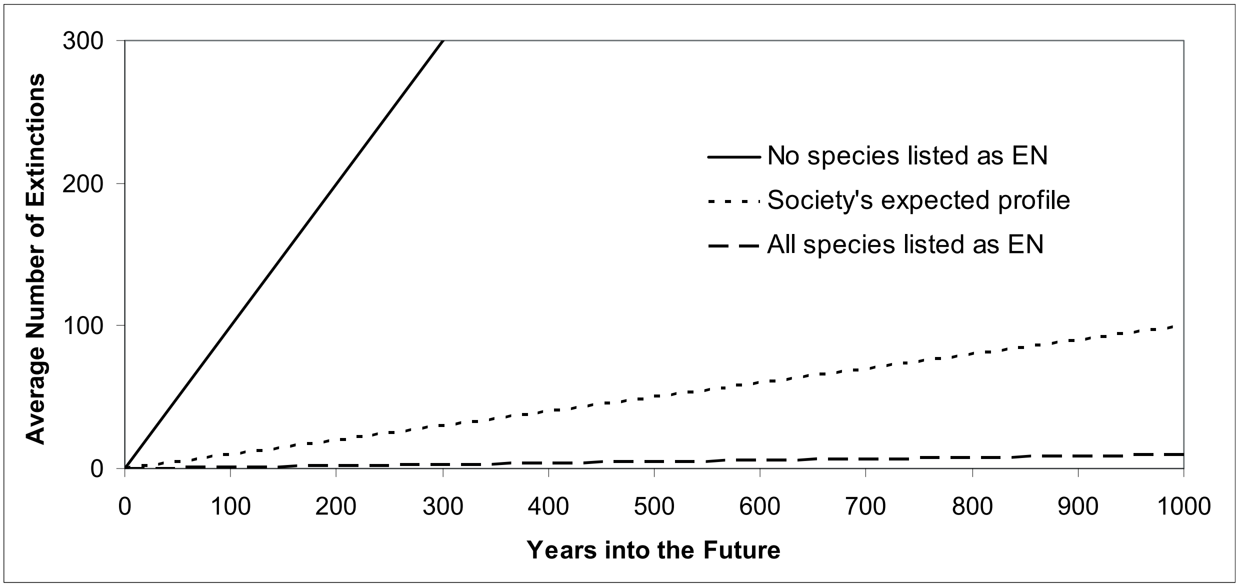


Figure 5. Hypothetical extinction profiles. Upper (solid) line: the number of extinctions over time if *no* species were listed as EN; middle (dotted) line: the number of extinctions over time society would expect and tolerate if the listing criterion was crafted appropriately; lower (dashed) line: the number of extinctions over time if *all* species were listed as EN.

Given a performance measure, or set of performance measures, that translate the objective and the simulation output into a common currency, simulation results can be used to compare the performance of the alternatives against the objective. If one or more of the alternatives comes close to achieving the objective, it would be possible to choose a preferred alternative at this point. If none of the alternatives comes close to achieving the objective, it may be appropriate to "go back to the drawing board" and come up with additional alternatives.

The QWG recommends that at least one modeler be dedicated to the task of carrying out the performance testing recommended in this section. The QWG estimates that this testing should take a single modeler 2 years to complete. Of course, some amount of time can be saved by using multiple modelers, but the testing is likely to take at least a year regardless of the number of modelers assigned to the task. The QWG also recommends that this work be guided by a steering committee that includes modelers and species specialists with knowledge of the disparate life history traits likely to be encountered in classifying species as EN or TH. Example life history types might include a long-lived organism with complex social behavior, a plant with a seed bank, an organism with a growth potential dependent on density (e.g., coral), an organism with complex spatial dynamics (meta-population), and an organism that experiences natural large fluctuations in abundance. Best results will be achieved through an iterative process, at least between the steering committee and the modeler(s), but most likely involving consultation with a broader group including agency managers and the public at one or more points in the process.

The next two subsections are written as though EN will be the only listing category under consideration. However, this is merely to make the presentation more readable. When the actual performance testing occurs, the QWG recommends that consideration of TH status be incorporated in a manner consistent with the rest of the performance test. The listing criterion for TH preliminarily recommended by the QWG should serve as a reasonable starting point.

Performance Testing of Alternative Listing Criteria

The exact sequence of steps used in performance testing can vary widely depending on the task at hand and the available resources. As an example, the steps in the first phase of performance testing recommended by the QWG might be structured as follows:

1) Determine the "extinction profile" (average number of future extinctions as a function of time) that represents what society expects and would tolerate from an appropriately crafted listing criterion.

2) Create alternative listing criteria (overarching definitions and values of any associated policy parameters), for example:

- A) Probability of Extinction Threshold definition and
  - a) 1% chance of extinction in 40 years
  - b) 1% chance of extinction in 80 years
  - c) 2% chance of extinction in 40 years
  - d) 2% chance of extinction in 80 years
- B) Depensatory Threshold definition (no policy parameters required)
- C) Comprehensive Threshold definition and
  - a) 1% risk of extinction with 1% discount rate
  - b) 1% risk of extinction with 2% discount rate
  - c) 2% risk of extinction with 1% discount rate
  - d) 2% risk of extinction with 2% discount rate

3) Simulate an expected extinction profile for each alternative listing criterion.

4) Compare the extinction profiles obtained in Step 3 to the extinction profile obtained in Step 1.

5) If one or more of the extinction profiles obtained in Step 3 is close to the extinction profile obtained in Step 1, choose the listing criterion corresponding to the extinction profile that gives the best fit. If none of the extinction profiles obtained in Step 3 is close to the extinction profile obtained in Step 1, return to Step 2.

It should be noted that in Step 2, the QWG recommends that the time horizon for the Probability of Extinction Threshold definition not exceed the following: min(100 years, max(Y years, Z generations)). That is, the QWG recommends that an additional policy parameter Z representing some number of species generations be incorporated into the Probability of Extinction Threshold definition. For any given species, the greater of Y years or Z generations would be adopted as a provisional time horizon. If this provisional time horizon is less than 100 years, it would constitute the time horizon; otherwise, the time horizon would be set at 100 years.

This recommendation is meant to improve the performance of the Probability of Extinction definition by explicitly allowing for differences in life history among taxa.

Performance Testing of Alternative Decision Metrics

As an example, the steps in the second phase of performance testing recommended by the QWG might be structured as follows, with the understanding that the entire process would be repeated for each taxonomic group or life history type:

1) Simulate a set of hypothetical populations such that the expected extinction time distribution and expected depensation threshold are obtained for each.

2) Determine the appropriate listing decision for each population based on the listing criterion.

3) Simulate hypothetical data for each population.

4) Create alternative decision metrics, for example:
- A) the listing criterion itself
- B) a 90% decline in numerical abundance
- C) a 95% decline in numerical abundance
- D) a population size of 10,000 individuals
- E) a population size of 20,000 individuals

5) Create potentially useful combinations of the decision metrics (e.g., (B) and (D) are a potentially useful combination, but (B) and (C) are not).

6) Determine an estimated listing decision for each population and combination of decision metrics based on the data.

7) Compare the sets of estimated listing decisions with the set of true listing decisions.

8) If one or more of the sets of estimated listing decisions obtained in Step 6 is close to the set of appropriate listing decisions obtained in Step 2, choose the combination of decision metrics that gives the best fit. If none of the sets of estimated listing decisions obtained in Step 6 is close to the set of appropriate listing decisions obtained in Step 2, return to Step 4. (If, after repeated iterations of this phase of performance testing, it is impossible to achieve a good fit, it may be necessary to return to the first phase of performance testing and look at a new set of alternative listing criteria.)

Step 8 in the above procedure may need further elaboration, because determining the "best fit" in this exercise involves several variables. For any given species, there are three possible listing decisions: EN, TH, and "do not list" (NL). Thus, errors in listing decisions have both a *direction* (either over- or under-protecting the species) and a *magnitude* (e.g., failure to list a truly endangered species is not the same as failure to list a truly threatened species). Further, society may weight over- and under-protection errors differently. For example, society may view a decision to list a truly threatened species as EN more positively than a failure to list the same species at all. An example of how decision theory can be used to determine the “best fit” between true and estimated listing decisions is given in Appendix 6.

It may also be desirable to augment the above procedure to consider the effects of alternative policies relating to continued monitoring and reevaluation of NL species. For example, if the status of each NL species were routinely reevaluated at 5-year intervals, initial errors in listing decisions might be weighted differently than they would be in a once-and-for-all listing procedure.

RECOMMENDATIONS

Adopt an Interim Policy

The QWG agreed that there is a need for an interim approach to species listings pending completion of the performance testing detailed in the previous section. The recommended approach is to institute a form of structured decision-making (Appendix 2) for all listing decisions. In addition, the QWG recommends that an interim protocol for NOAA Fisheries be developed that separates risk analysis from risk management. The QWG further recommends that the approach currently developed in the Northwest Region for ESA classification and Distinct Population Segment (DPS) designation of Pacific salmon, Puget Sound marine fish, and southern resident killer whales be used as a starting point for the development of an interim protocol.

In this approach, Biological Review Teams (BRTs) are formed to evaluate the risk of extinction faced by a biological entity (species, population, DPS or Evolutionary Significant Unit). The process is highly structured and is documented in detail so that it is as transparent as possible to reviewers of the results. Although the BRTs develop scientific conclusions about the risk of extinction faced by an entity, NOAA Fisheries Regional Offices make the final listing decisions based on both the scientific input from the BRTs and taking into consideration the likely effects of conservation measures that are proposed or in place (see Appendix 2 for more details). Various aspects of the Pacific salmon listing model may need to be modified to accommodate variation in kinds and amounts of available data, and types and magnitudes of threats. In addition, effort should be made to make the NOAA Fisheries process consistent or at least comparable to the process currently in use by the USFWS.

Conduct Performance Testing of Listing Criteria and Decision Metrics

The two phases of performance testing recommended in the previous section should be conducted. The exact form of such tests can be modified, but the examples previously described should be given serious consideration.

Choose Final Listing Criteria and Decision Metrics

The QWG recommends that a broad-based panel be established and empowered to evaluate the results of the performance tests in light of the Guiding Principles, and make official choices of overarching definitions, values of any associated policy parameters, and decision metrics for both EN and TH.

Enhance Use of Structured Decision Making

The QWG believes that, even with agreed-upon listing criteria and decision metrics, some type of structured decision-making can and should still play an essential role in listing decisions. One compelling reason for this recommendation is that there will inevitably be cases where data are so sparse that no quantitative decision metric can be used. In such cases, it may be necessary to draw upon expert opinion to find suitable proxies using the few existing data and analogies with related species. Thus, we recommend that a working group be formed immediately to develop structured decision making methods for use in the context of the interim policy (Section 5.1) and to develop methods for incorporating the eventual listing criteria and decision metrics into a structured decision making framework.

ACKNOWLEDGEMENTS

The QWG was assisted in its efforts by many people including Susan Pultz, Peter Beck, Gary Duker, and Marta Nammack. Nonetheless, the recommendations in this report are solely those of the QWG and people acknowledged above may or may not agree with the recommendations herein.

CITATIONS

Abell, R. 2002. Conservation biology for the biodiversity crisis: a freshwater follow-up. Cons. Biol. 16:1435-1437.

Ayyub, B.M. 2001. Elicitation of Expert Opinions for Uncertainty and Risks. CRC Press, Boca Raton, FL. 302pp.

Beissinger, S.R., and D.R. McCollough. 2002. Population Viability Analysis. Univ. Chicago Press, Chicago, IL. 593pp.

Beissinger, S.R., J.M. Reed, J.M. Wunderle Jr., S.K. Robinson, and D.M. Finch. 2000. Report of the AOU Conservation Committee on the Partners in Flight Species Prioritization Plan. The Auk 117(2):549-561.

Bruton, M.N. 1995. Have fish had their chips? The dilemma of threatened fishes. Environ. Biol. Fishes 43:1-27.

Carter, M.F., W.C. Hunter, D.N. Pashley, and K.V. Rosenberg. 2000. Setting conservation priorities for landbirds in the United States: the Partners in Flight approach. The Auk 117(2):541-548.

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. Ecol. Appl. 6(1):1-11.

CITES (Convention on the Trade of Endangered Species). 1994. Resolution Conf. 9.24. http://www.cites.org/eng/resols/9/index.shtml

Cleaves, D.A. 1994. Assessing uncertainty in expert judgments about natural resources. General Technical Report SO-1 10. U.S.D.A. Forest Service Southern Forest Experiment Station, New Orleans, LA. July 1994.

Clemen, R.T., and T. Reilly. 2001. Making Hard Decisions with DecisionTools®. Duxbury, Pacific Grove, CA. 733pp.

Cochrane, J.F. 2002. Appendix B. Expert interview process—methods and summary, p.54-60. *In* Cochrane, J.F. and P. Delphey, Dakota Skipper *Hesperia dacotae* (Skinner) (Lepidoptera: Hesperiidae) status assessment and conservation guidelines: Iowa, Minnesota, North Dakota, South Dakota, and Manitoba. Unpublished report, U.S. Fish and Wildlife Service, Twin Cities Field Office, Bloomington, MN. 80pp. Available online at: http://midwest.fws.gov/Endangered/insects/dask-status.pdf

Cooke, R.M. 1991. Experts in uncertainty: Opinion and subjective probability in science. Oxford Univ. Press, New York, NY. 330pp.

Diamond, J.M. 1988. Red books or green lists? Nature 332: 304-305.

Easter-Pilcher, A. 1996. Implementing the Endangered Species Act: assessing the listing of species as endangered or threatened. BioScience 46(5):354-363.

Eisner, T., J. Lubchenco, E.O. Wilson, D.S. Wilcove, and M.J. Bean. 1995. Building a scientifically sound policy for protecting endangered species. Science 268(5228):1231-1232.

FAO (Food and Agriculture Organization of the UN). 2001a. A background analysis and framework for evaluating the status of commercially-exploited aquatic species in a CITES context. Second Technical Consultation on the Suitability of CITES Criteria for Listing Commercially-Exploited Aquatic Species. FI:SLC2/2001/2. 19p.

FAO (Food and Agriculture Organization of the UN) 2001b. Report of the Second Technical Consultation on the Suitability of CITES Criteria for Listing Commercially-Exploited Aquatic Species. FAO Fisheries Report No. 667 [English, French and Spanish]. 87p.

FAO (Food and Agriculture Organization of the UN) 2002. Report of the Second Technical Consultation on the Suitability of the CITES Criteria for Listing Commercially-Exploited Aquatic Species. FAO Fisheries Report No. 667. 87p.

Forest Ecosystem Management Team (FEMAT). 1993. Forest ecosystem management: an ecological, economic, and social assessment. U.S.D.A. Forest Service and five other federal agencies. Portland, OR. [Northwest Forest Plan]

General Accounting Office. 1989. Endangered species: spotted owl petition beset by problems. Report Resources, Community and Economic Development Division 89-79. General Accounting Office, Washington, DC.

Goodman, D. 2002. Predictive Bayesian population viability analysis: a logic for listing criteria, delisting criteria, and recovery plans, . p. 447-469. *In* S.R. Beissinger and D.R. McCollough, (eds.), Population Viability Analysis. University of Chicago Press.

Goodwin, P., and G. Wright. 1999. Decision Analysis for Management Judgment. 2nd Ed. John Wiley & Sons, New York, NY. 468pp.

Harrison, I.J. and M.L.J. Stiassny. 1999. The quiet crisis. A preliminary listing of the freshwater fishes of the world that are extinct or “missing in action”, p. 271-331. *In* R. MacPhee (ed.), Extinctions in Near Time. Kluwer Academic/Plenum Publishers, New York, NY.

Heywood, V.H. (ed.). 1995. Global Biodiversity Assessment. Cambridge University Press, New York, NY.

Hutchings, J.A. 2001. Conservation biology of marine fishes: perceptions and caveats regarding assignment of extinction risk. Can. J. Fish. Aquat. Sci. 58:108-121.

IUCN (The World Conservation Union). 2001. IUCN Red List Categories: Version 3.1. Prepared by the IUCN Species Survival Commission. IUCN, Gland, Switzerland and Cambridge, UK.

Lee, D.C., and B.E. Rieman. 1997. Population viability assessment of salmonids using probabilistic networks. North Am. J. Fish. Manage. 17:1144-1157.

Lehmkuhl, J.F., M.G. Raphael, R.S. Holthausen, J.R. Hickenbottom, R.H. Nancy, and J.S. Shelly. 1997. Historical and current status of terrestrial species and the effects of proposed alternatives. Chapter 4, *in* Quigley, T.M., A.M. Lee, and S.J. Arbelbide. Evaluation of EIS alternatives by the Science Integration Team. Volume II. General Technical Report PNW-GTR-406. U.S.D.A. Forest Service, Pacific Northwest Research Station, Portland, OR. [Interior Columbia Basin Ecosystem Management Project]

Lunney, D., A. Curtin, D. Ayers, H.G. Cogger, and C.R. Dickman. 1996. An ecological approach to identifying the endangered fauna of New South Wales. Pac. Cons. Biol. 2:212-231.

Mace, P.M., A.W. Bruckner, N.K. Daves, J.D. Field, J.R. Hunter, N.E. Kohler, R.G. Kope, S.S. Lieberman, M. W.Miller, J.W. Orr, R.S. Otto, T.D. Smith, N.B. Thompson, J. Lyke and A.G. Blundell. 2002. NMFS / Interagency Working Group Evaluation of CITES Criteria and Guidelines. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-F/SPO-58. 70pp.

Maguire, L. and J.F. Cochrane. 2001. Expert Opinion. Chapter 8, p. 100-118. *In* Andelman, S.J., S. Beissinger, J.F. Cochrane, L. Gerber, P. Gomez-Priego, C. Groves, J. Haufler, R. Holthausen, D. Lee, L. Maguire, B. Noon, K. Ralls, and H. Regan. Scientific standards for conducting viability assessments under the National Forest Management Act: report and recommendations of the NCEAS working group. National Center for Ecological Analysis and Synthesis, Santa Barbara, CA. 160pp. Available online at: http://www.nceas.ucsb.edu/nceas- web/projects/3960/NCEASviabilityfinal report1201.pdf .

Marcot, B.G., R.S. Holthausen, M.G. Raphael, M. Rowland, and M. Wisdom. 2001. Using Bayesian belief networks to evaluate fish and wildlife population viability under land management alternatives from an environmental impact statement. For. Ecol. Manage. 153(1-3):29-42.

Master, L.L. 1991. Assessing threats and setting priorities for conservation. Cons. Biol. 5(4):559-563.

Master, L.L., B.A. Stein, L.S. Kutner, and G.A. Hammerson. 2000. Vanishing Assets: conservation status of U.S. species, p. 93-118 *In* Stein, B.A., L.S. Kutner, and J.S. Adams (eds.), Precious Heritage: the status of biodiversity in the United States. Oxford University Press, NY. 399pp.

Master, L.L., L.E. Morse, A.S. Weakley, G.A. Hammerson, and D. Faber-Langendoen. 2002. Heritage conservation status assessment factors. June 2002 Draft. NatureServe, Arlington, VA.

McCarthy, M.A., S.J. Andelman, and H.P. Possingham. 2003. Reliability of relative predictions in population viability analysis. Cons. Biol. 17(4):982-989.

McElhany, P., M.H. Rucklelshaus, M.J. Ford, T.C. Wainwright, and E.P. Bjorkstedt. 2000. Viable salmonids populations and the recovery of evolutionarily significant units. U. S. Dep. Commer. NOAA Tech. Memo. NMFS-NWFSC-42, 174pp.

Mighton, S.R., R. Meier, E. Padley, T. Doyle, M. Shedd, and J. Gallagher. 2000. Population viability assessment in forest plan revision: National Forests of Wisconsin and Minnesota. Unpublished report, U.S.D.A. Forest Service, Region 9, Milwaukee, WI. May 22, 2000. 154pp.

Millsap, B.A., J.A. Gore, D.E. Runde, and S.I. Cerulean. 1990. Setting priorities for the conservation of fish and wildlife species in Florida. Wild. Monogr. 111:1-57.

Musick, J.A. 1999. Criteria to define extinction risk in marine fishes. Fisheries 24(12):6-12.

Musick, J.A., M.M. Harbin, S.A. Berkeley, G.H. Burgess, A.M. Eklund, L. Findley, R.G. Gilmore, J.T. Golden, D.S. Ha, G.R. Huntsman, J.C. McGovern, S.J. Parker, S.G. Poss, E. Sala, T.W. Schmidt, G.R. Sedberry, H. Weeks and S.G. Wright. 2000. Marine, estuarine, and diadromous fish stocks at risk of extinction in North America (exclusive of Pacific salmonids). Fisheries 25(11):6-30.

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

NatureServe. 2002. Element occurrence data standard. February 6, 2002, draft. Available online at http://whiteoak.natureserve.org/eodraft/index.htm .

Nelson, J.S. 1994. Fishes of the World, 3rd ed. Wiley, New York, NY. 600pp.

Primack, R.B. 1998. Essentials of Conservation Biology, 2nd ed. Sinauer Associates, Inc., Sunderland, MA. 660pp.

Ralls, K., and B. L. Taylor. 1996. How viable is population viability analysis? p. 228-225. *In* S.T.A. Pickett, R.S. Ostfeld, M. Shachak, and G.E. Likens, eds. Enhancing the ecological basis of conservation: heterogeneity, ecosystem function, and biodiversity. Proceedings of the Sixth Cary Conference. Chapman and Hall, New York, NY.

Ralls, K., S. R. Beissinger, and J. F. Cochrane. 2002. Guidelines for using population viability analysis in endangered-species management, p. 521-550. *In* S. R. Beissinger and D. R. McCollough, (eds.), Population Viability Analysis. University of Chicago Press.

Raphael, M. G., M. J. Wisdom, M. M. Rowland, R. S. Holthausen, B. C. Wales, B. G. Marcot, and T. D. Rich. 2001. Status and trends of habitats of terrestrial vertebrates in relation to land management in the interior Columbia River Basin. Forest Ecology and Management 153(1-3):63-87.

Raup, D.M. 1992. Extinction: Bad Genes or Bad Luck? W.W. Norton, New York, NY. 224pp.

Reagan, T.J., L.L. Master, and G.A. Hammerson. 2004. Capturing expert knowledge for threatened species assessments: a case study using NatureServe conservation status ranks. Acta Oecologia 26(2):95-107.

Reid, W.V., and K.R. Miller. 1989. Keeping Options Alive: The Scientific Basis for Conserving Biodiversity. World Resources Institute, Washington, DC.

Roberts, C.M., and J.P. Hawkins. 1999. Extinction risk at sea. Trends in Ecology and Evolution 14:241-246.

Rohlf, D.J. 1991. Six biological reasons why the Endangered Species Act doesn't work and what to do about it. Cons. Biol. 5(3):181-200.

Schenck, T, C. Chaney, T. Doyle, and M. Shedd. 2002. Expert panels for species viability evaluation for preliminary draft EIS alternatives: National Forests in Wisconsin and Minnesota. Draft process, February 26, 2002. Unpublished report, U.S.D.A. Forest Service, Duluth, MN. 18pp. Available online at: http://www.fs.fed.us/r9/chippewa/plan/revision/analyses/sve_second_panel.pdf .

Shaw, C.G. III. 1999. Use of risk assessment panels during revision of the Tongass Land and Resource Management Plan. General Technical Report PNW-GTR-460. U.S.D.A. Forest Service, Pacific Northwest Research Station, Portland, OR. 43pp.

Smith, F.D.M., R.M. May, R. Pellew, T.H. Johnson and K.R. Walter. 1993. How much do we know about the current extinction rate? Trends Ecol. Evol. 8:375-378.

Stiassny, M.L.J. 2002. Conservation of freshwater fish biodiversity: the knowledge impediment. Verh. Ges. Ichthyol. 3: 7-18.

Stewart, T.J. 1992. A critical survey on the status of multiple criteria decision making theory and practice. OMEGA Int. J. Manage. Sci. 20(5/6):569-586.

Taylor, B.L. 1995. The reliability of using population viability analysis for classification of species. Cons. Biol. 9:551-558.

Taylor, B.L., P.R. Wade, U. Ramakrishnan, M. Gilpin, and H. R.Akçakaya. 2002. Incorporating uncertainty in PVAs for the purpose of classifying species by risk, p. 239-252. *In* S.R. Beissinger and D.R. McCollough, (eds.), Population Viability Analysis.

Taylor, B. L., P.R. Wade, R.A. Stehn, and J.F. Cochrane. 1996. A Bayesian approach to classification criteria for spectacled eiders. Ecol. Appl. 6:1077-1089.

Tobin, R.J. 1990. The expendable future: U.S. politics and the protection of biological diversity. Duke University Press, Durham, NC.

U.S.D.A. Forest Service. 2001. Appendix R – Assessment of species vulnerability and prioritization. Final Environmental Impact Statement, Sierra Nevada Forest Plan Amendment. U.S.D.A. Forest Service Region 5, Vallejo, CA.

U.S.D.A. Forest Service. 2002. Draft environmental impact statement for the draft land management plan for Midewin National Tallgrass Prairie. U.S.D.A. Forest Service Region 9, Milwaukee, WI.

U.S. Fish and Wildlife Service. 1983. Endangered and Threatened Species Listing and Recovery Priority Guidelines. Federal Register 48(184):43098-43105.

U.S. Fish and Wildlife Service. 1994. Endangered species listing handbook: procedural guidance for the preparation and processing of rules and notices pursuant to the Endangered Species Act. 4th Edition. U.S. Fish and Wildlife Service, Endangered Species Division, Washington, DC. March 1994.

U.S. Fish and Wildlife Service. 1997a. Petition to list the Queen Charlotte Goshawk under provisions of the Endangered Species Act, 12-month finding. U.S. Fish and Wildlife Service, Washington, DC. August 28, 1997.

U.S. Fish and Wildlife Service. 1997b. Decision process for reviewing the status of wolves and Queen Charlotte Goshawks in southeast Alaska. Unpublished administrative record for the 12-month petition finding, U.S. Fish and Wildlife Service Regional Office, Anchorage, AK.

U.S. Fish and Wildlife Service. 1999. Eastern Massasauga elevation recommendation. Report for the administrative record, May 17, 1999. U.S. Fish and Wildlife Service, Endangered Species Program, Region 3, Ft. Snelling, MN.

Wade, P.R. 2002. Bayesian population viability analysis, p. 213-238. *In* S. R. Beissinger and D. R. McCollough, (eds.), Population Viability Analysis. University of Chicago Press

Whitten, A.J., K.D. Bishop, S.V. Nash and L. Clayton. 1987. One or more extinctions from Sulawesi, Indonesia? Cons. Biol. 1:239-246.

Wilson, E.O. 1987. The little things that run the world: The importance and conservation of invertebrates. Cons. Biol. 1:344-346.

Wilcove, D. S., M. McMillan, and K. C. Winston. 1993. What exactly is an endangered species? An analysis of the Endangered Species List, 1985–1991. Cons. Biol. 7:87-93.

Yaffee, S.L. 1982. Prohibitive Policy: Implementing the Federal Endangered Species Act. MIT Press, Cambridge, MA. 223pp.

APPENDICES

APPENDIX 1: LESSONS LEARNED FROM THE EXPERIENCE OF THE U.S. FISH AND WILDLIFE SERVICE

The U.S. Fish and Wildlife Service (USFWS) has considerably more experience with ESA listing determinations than NOAA Fisheries. Therefore, it is appropriate to summarize the approach to listing currently used by the USFWS. This summary was provided by one of the QWG members (JC).

The USFWS's approach to listing relies on case-by-case professional judgment. It is the opinion of the USFWS that Congressional drafters of the Endangered Species Act of 1973 (ESA) left the law intentionally vague, noting that the difference between threatened and endangered species was a "matter of degree" and they believed, impossible to determine *a priori* given the overwhelming diversity of species and their needs. Thus, Congress intended to "rely on the [agency] professionals" to make the distinctions.

> Many witnesses testifying at the hearings expressed concern that the legislation does not contain meaningful, objective standards to guide the Secretary when making a determination whether certain species are threatened with worldwide extinction. They feared...arbitrary and frivolous action. Your committee does not believe specific standards can be written into the legislation without harming the effect of the legislation. Existing species are so varied that a standard to fit all appears incredibly complex and cumbersome... (H.R. 382, 92st Cong., 1st sess., 1969 cited in Easter-Pilcher 1996).

The USFWS's listing process consists of an in-depth status review resulting in a narrative status report and listing recommendation, proceeding to rule-making or public notice documents. The Endangered Species Listing Handbook (USFWS 1994) provides lengthy guidance on the administrative and rule-making steps for listing such as how to prepare Federal Register notices, but does not establish policy upon which the USFWS bases listing decisions. In compliance with Section 4(h) of the ESA, the USFWS has published guidelines for identifying species for priority review (USFWS 1983). While technically these guidelines only address prioritization among candidate species, they make explicit that the magnitude and immediacy of threats (causes of population decline) are the key considerations in USFWS listing decisions.

The core parts of all status reviews and listing documents are a thorough review of the species' taxonomy, life history, habitat and ecological relationships, and population status followed by a "five-factor analysis" of threats. The threats analysis is organized around the bulleted "factors" A through E in Section 4(a) of the ESA, commonly referred to as the "five listing factors". These factors represent principle, external mechanisms that can cause populations to decline to extinction including habitat and range destruction, modification or curtailment, predation and disease, and overutilization. The inadequacy of existing regulatory mechanisms to prevent overutilization or other threats is the fourth "listing factor." The fifth "factor" or bullet in the printed law, is a catch-all: "other natural or manmade factors affecting [the species'] continued existence."

Together, these factors or "threats" serve as a checklist to be used in evaluating species status, but also an indication that at-risk species should be protected regardless of the source for endangerment—any cause for decline including "natural" sources can be an "other" factor. At the same time, this section of the law indicates that decisions about the overall status of a species stem from a potentially complex set of influences, all of which must be weighed together to determine the species' most likely fate. In other words, when considering the "weight of the evidence" (the presumed legal basis for ESA decision-making under uncertainty), *all extinction risk factors* for which information is available must be included in the evidence. Thus, what is traditionally called "five-factor analysis" is better described as "full-factor analysis," although endangerment may stem from only one or a few factors.

Advantages of the Case-by-Case Approach

The case-by-case approach allows for infinite accommodation to individual situations. In-depth, case-by-case evaluation does not over-simplify analysis to fit into predetermined rules or protocols that may rely on only a limited range of information. The strength of the USFWS status review process is using information from a wide range of sources, documented in a lengthy written record, thus supporting both the 'full factor' analysis and 'weight of the evidence' decision making standards.

Disadvantages of the Case-by-Case Approach

Although the case-by-case approach provides great flexibility, the lack of explicit and transparent standards can turn flexibility into arbitrary and capricious decision-making. Consistent treatment across taxa relies on the presumption that the biologist or decision-maker is adjusting an overarching listing standard (implicitly known) to the particular situation of each species. Numerous reviews have revealed inconsistencies across past listing decisions resulting from subjective assessments in face of high scientific uncertainties (Yaffee 1982, General Accounting Office 1989, Tobin 1990, Rohlf 1991, Wilcove et al. 1993, National Research Council 1995, Easter-Pilcher 1996). Inconsistent results do not necessarily mean that species are being listed inappropriately (Carroll et al. 1996, Eisner et al. 1995), but they do reduce the efficiency and effectiveness of the listing program (Wilcove et al. 1993).

A key weakness of the USFWS' current listing process is its failure to discriminate between risk analysis, which estimates and describes the individual species' projected status, and risk management, the standards for how much and how soon species must be at risk of extinction—and how certain we want to be of those estimates—for the species to qualify for ESA protection. The case-by-case process of the USFWS blends the risk analysis and protection preferences indistinguishably. When listing standards are implicit, it's not possible to identify the likelihood of extinction, time to potential extinction, and the scope and reliability of data needed to classify species—and these considerations are probably quite blurred in the decision-maker's mind.

In sum, the decision standards are opaque, subjective, and not consistently repeatable; for example, if different staff or offices complete status reviews using the same information, the resulting classification recommendations will vary, particularly for less-obviously or less-imminently endangered species. Although by law only scientific and commercial (e.g., overutilization) information may be used in listing decisions, the lack of transparency in the case-by-case approach also allows for other considerations—even unintended or unrecognized biases—to infiltrate decisions.

APPENDIX 2: REVIEW OF EXISTING APPROACHES TO DEFINING "ENDANGERED"

Population Viability Analysis

Population viability analysis (PVA) was originally conceived as a method to incorporate the range of threats facing a species to estimate the minimum population size for a species considered in a reserve design context. However, it quickly became apparent that PVAs could be used for many purposes, including estimating the risk of extinction in a given period of time, which can be used in the listing process. The IUCN criteria include a category that would be based on PVAs, but this criterion is seldom used. A review of the current usage of PVA can be found in Beissinger and McCullough (2002). Here, we restrict our discussion of PVA to its use for listing purposes and to PVA models that go beyond extrapolation of recent trends to include population processes.

Population viability analysis is the only method that attempts to integrate all risks into a single quantity. This has strong benefits when the desired output is an absolute measure of risk of extinction because it is known that many risk factors interact. For example, a population decline caused by an extreme environmental event could result in reduced genetic variability, which could in turn compromise the population's ability to respond to disease. There is no question that small populations face a host of challenges that together make them much more vulnerable to extinction than large populations. There is also no question that simple proxies (like population size or rate of decline) do not capture these biological complexities. So why are PVAs seldom used in practice for listing purposes?

The primary reason is that PVAs require enormous amounts of data. Capturing biological complexity means building models with many parameters. Even when we have data for some parameters, we often lack data in the very areas where we have the greatest concern; that is, when abundance is very low. Ralls *et al*. (2002) suggest guidelines for using PVAs in endangered species management. However, most biologists feel very uncomfortable building a model that relies on parameters for which they have no direct data for the species in question. They feel much more comfortable using simple proxies, like a single estimate of abundance, even though that proxy ignores the biological complexity that we know to be true.

Population viability analyses are difficult to use in the listing process because the available models were created primarily to estimate relative rather than absolute extinction risks. Relative risk is sufficient when considering alternative management actions where all would be affected similarly by some unknown factor, such as reduced fecundity through inbreeding depression. In such an application, inclusion of genetic effects may be irrelevant to choosing a management option if all options would suffer the same, but unknown, negative genetic effects. Estimation of actual extinction risks, however, should attempt not only to include all risk factors, but also to include the uncertainty in such factors. There are several PVA programs available either commercially or as free software. These packages incorporate two types of uncertainty called "environmental stochasticity" and "demographic stochasticity". The latter assumes that birth and death rates vary more as abundance decreases simply because chance events cannot replicate probabilities at small abundance. For example, if adult survival rate averaged 0.97 each year, but only 10 adults remained, then the realized survival rate could only be 0.8, 0.9 or 1.0 and could not be 0.97. Environmental stochasticity incorporates the effects of the environment that are felt by the entire population. This allows the mean birth and death rates for the population to change according to the quality of the environment with respect to the species in a given year.

While allowing this type of uncertainty to be captured, most available PVA models do not incorporate uncertainty in parameter estimates (Taylor 1995, Ralls and Taylor 1996). However, several case-specific models have been developed to do so (Taylor et al. 1996, Taylor et al. 2002, Wade 2002). The reason why it is essential to capture this type of uncertainty for listing purposes is that it allows all species to be ranked on an equal basis in theory, regardless of the magnitude of our ignorance, and it allows PVAs to be conducted regardless of the amount of data available (Goodman 2002). Three important tasks are accomplished by direct incorporation of uncertainty: 1) denying the option of inaction in the face of uncertainty, 2) encouraging precaution by making the probability of listing higher for species for which there is greater uncertainty, and 3) rewarding gathering knowledge by reducing risk estimates as the magnitude of uncertainty/ignorance decreases.

Another benefit of conducting a formal risk analysis (PVA) is that the same model can be used not only in the listing decision, but later to evaluate and prioritize conservation actions. Thus, although there may be considerable investment in creating more appropriate and flexible PVA models to estimate absolute extinction risks (including all the uncertainties), these models can be used repeatedly in the species' conservation. For example, the PVA can be used to prioritize which actions result in the greatest decrease to extinction risk, how long it may take to recover the species under various options, and even to evaluate how to define critical habitat. None of these tasks can be accomplished using proxy listing criteria.

The final reason PVAs are also seldom used to list species is that because available models seldom address estimating absolute extinction risk, new models need to be created. There is a real shortage of scientists with the skill needed to create these models and there has been little institutional support to train such modelers.

IUCN, CITES and AFS Rule-Based Approaches

IUCN approach

The International Union for the Conservation of Nature has used a quantitative rule-based approach for listing species on its Red Lists since 1994. Recently, this approach was revised and expanded and has now been adopted by IUCN for current and future listings (IUCN 2001). The revised approach is extremely detailed and comprehensive, containing both quantitative elements (Appendix Table A2.1) and qualitative elements (e.g., in addition to quantitatively defining a restricted extent of occurrence, at least two of the following three factors must pertain: severe fragmentation, continuing decline in population size or area of distribution, or extreme fluctuations in population size or area of distribution).

In some respects, the current IUCN rule-based approach is too complex, in that it defines too many boundaries and somewhat arbitrary cut-off points (Appendix Table A2.1). In other respects, it is too simple, in that it uses a single set of numeric guidelines that cannot possibly apply across all taxonomic groupings and, in fact, are probably relevant for only a small proportion of extant taxa. For example, the population sizes signifying "critically endangered", "endangered", and "vulnerable" are, respectively, 250, 2,500, and 10,000 mature individuals (Appendix Table A2.1), yet for some taxa (e.g., many commercially exploited marine fish) the numbers triggering concern would be much higher than this, while for others (e.g., some of the large whales), numbers in the range 2,500 - 10,000 are likely to be considered as representative of a "healthy" population. Thus, the Quantitative Working Group (QWG) agreed that such an approach would not ensure that different taxa would receive equivalent levels of protection. Nevertheless, the *concept* of specifying extinction risk criteria in terms of categories like A-E (but using taxon-specific quantities) is worthy of further consideration.

CITES approach

Because IUCN acts as an advisory body to CITES, it is not surprising that the approach adopted by CITES in 1994 (CITES 1994) was based on the IUCN approach at that time (Appendix Table A2.1). There are, however, three important differences: (i) CITES Parties decided not to "hard-wire" numeric thresholds into the criteria; therefore, the descriptive criteria and numeric guidelines are contained in separate annexes (Annexes 1 and 5 of CITES 1994, respectively), (ii) the numeric cutoffs are suggested simply as guidelines, not thresholds, and (iii) the CITES descriptive criteria and numeric guidelines (Appendix Table A2.2) are much simpler and much more flexible than the IUCN approach (Appendix Table A2.2). One reason for these differences is that the direct consequences of listing species on CITES Appendices (restrictions or outright bans on international trade) are much greater than those resulting from an IUCN Red List listing.

AFS approach

The IUCN system was also considered when the American Fisheries Society (AFS) developed criteria to be used to define the risk of extinction in marine fishes (Musick 1999). The American Fisheries Society concluded that decline criteria were the most appropriate for marine species of fish, but that the numeric rates used by IUCN (Categories A in Appendix Table A2.1) would "grossly overestimate the extinction risk for many if not most marine fish species". They therefore proposed much greater decline thresholds (i.e., thresholds that would trigger concern about extinction risk at much lower relative population sizes). Perhaps more importantly, they introduced the idea of incorporating population resilience as a factor to consider when developing criteria for extinction risk. Assuming population productivity is a reasonable surrogate for resilience (based, in part, on the empirical record, which shows that high productivity species have experienced far lower extinction rates than low productivity species), AFS defined four categories of productivity based on values of the intrinsic rate of natural increase, growth rates, fecundity, age at maturity and maximum age. They then suggested different decline thresholds for each of the categories, with the relative threshold population decline being inversely related to productivity, as follows:

Productivity	Decline (over the longer of 10 years or 3 generations)
High	0.99
Medium	0.95
Low	0.85
Very Low	0.70

Musick et al. (2000) subsequently used this system, along with consideration of other "risk factors" to classify populations within 82 non-salmonid species or subspecies as vulnerable to extirpation. Most were nearshore reef-fishes affected by habitat degradation, or species with low to very low productivity (e.g., some sharks, sturgeons, Pacific coast rockfish and tropical groupers).

Recent developments

When the current CITES criteria and guidelines were adopted in 1994, it was also decided that they would be considered for possible revision in time for the 12th CITES Conference of the Parties (12th COP, held in November 2002). Therefore, considerable effort has been put into examining these criteria during the last 2-3 years. For example, CITES formed a Criteria Working Group with regional representation of member Parties that met several times and produced a set of recommendations for revision; NMFS set up and chaired a U.S. interagency working group that also critically evaluated the criteria and recommended several substantive changes (Mace et al. 2002); and FAO contracted experts to develop relevant background papers and held two major Technical Consultations focused specifically on criteria appropriate for commercially exploited species in marine and large freshwater bodies (FAO 2001a, b). However, the CITES criteria and guidelines were not revised at the 12th COP.

The recommendations that have emerged from reviews of the CITES criteria conducted to date are that Annex 1 of CITES (1994), which consists of qualitative descriptions of listing criteria to be used for an Appendix I listing (Appendix Table A2.2), requires little if any revision; Annex 2, which consists of qualitative descriptions of listing criteria to be used for an Appendix II listing (Appendix Table A2.2), needs to be more objective and specific; and Annex 5, which contains definitions, notes, and numeric guidelines, requires substantial elaboration in order to make it more useful to Parties developing proposals for listing or downlisting. In particular, the numeric guidelines (Appendix Table A2.2) are not very useful at present because they specify single absolute numbers that are only applicable for a small number of taxa.

Thus, CITES is still considering a rule-based approach, focused on one or more of three fundamental indicators of extinction risk: (i) small population size, (ii) restricted area of distribution, and (iii) a marked extent or rate of decline. There are currently no formal recommendations to change the numeric guidelines for (i) or (ii), although future proposals may recommend that guidelines on taxa-specific ranges of numbers be developed (e.g., a range of about 1,000-5,000 may constitute a small population for large whales, while a range of about 10,000-100,000 may constitute a small population for many commercially exploited marine species). Much of the current debate is centered on characterizing the decline criteria, and the utility of generation times to assess the extent or rate of decline. Current decline criteria (Appendix Table A2.1) only consider the rate of decline and use generation times as the period of assessment. Recommended changes include addition of, and emphasis on, the historical extent of decline (in population numbers, population biomass, area inhabited, percent coverage, or other relevant variables) with the period of assessment going back to a reasonable or potential baseline determined on a case-by-case basis. It has also been recommended that the period of assessment should not be based on generation time because, although it makes sense to look far back (or forward) in time for long-lived species, it also makes sense to look far back (or forward) for short-lived species. (In other words, it doesn't make sense to restrict the period of assessment to, say, 15 years for a species with a generation time of 5 years when the evidence suggests that there has been a long-term declining trend over the past few decades.) See Section 3 and Appendices IV and V of Mace et al. (2002) for further elaboration and examples supporting this argument.

It was further suggested by Musick (1999) and Mace et al. (2002) that the threshold extent of decline should be a function of the productivity of the species (where productivity is assumed to be a surrogate for population resilience. For a very high productivity species (e.g., one with high fecundity and a rapid turnover of generations), consideration for listing in CITES Appendices might not be triggered until the species has declined to relatively low levels; for example, about 5% of the baseline. For a very low productivity species (e.g., one with low fecundity and a long period between generations), consideration for listing in CITES Appendices might be triggered at much higher levels of relative population size; for example, a decline to about 30% of the baseline level. Support from the peer-reviewed fisheries scientific literature for the applicability of the 5-30% range and the role of productivity is summarized and discussed in Appendix I of Mace et al. (2002).

In subsequent evaluations, the Food and Agricultural Organization (FAO; 2001a, b) concluded that for the majority of commercially exploited fish and invertebrates in marine and large freshwater bodies, a narrower range of 5-20% would be more appropriate, with a range of 5-10% being used for relatively high-productivity species, 10-15% for species with medium productivity, and 15-20% for species with low productivity (where these three categories of productivity levels were characterized based on ranges of life history parameters).

Some CITES Parties have suggested that for species with lower productivity than commercially exploited marine species, it may be more appropriate to choose from the upper end of the 5-30% range. However, the 5-30% range seems a little high when considering an ESA listing. Even though both CITES and ESA are concerned with risk of extinction, CITES simply bans or regulates international trade and has no authority over domestic consumption. It seems reasonable that international trade in a given species should cease before severe restrictions are needed for domestic or personal use (e.g., international trade should cease at a higher population size than that triggering severe restrictions on domestic consumption).

At present, there is little impetus within CITES to adopt a PVA approach, or even to include it as one of the fundamental indicators of extinction risk as IUCN does (category E in

Appendix Table A2.1). The main reasons are likely the lack of data to formulate comprehensive and realistic PVAs in most cases, the lack of experience observing and experimenting with populations at low levels, and a worldwide paucity of modeling expertise. There is no doubt that model projections of future population size provide a better basis for decision-making than rule-based approaches, but only when the model incorporates population processes, and not necessarily when the models are simply mechanical extrapolations of recent trends, or when they are used to project far beyond the foreseeable future.

Point-scoring, Hybrid Approaches, and the Heritage Ranking Process

Description of point-scoring and hybrid methods

The previous section described the well-known species classification approaches from IUCN and CITES, as well as the more recent AFS approach, all of which are "rule-based" protocols. Another approach to classifying species may use the same or similar criteria, but assigns points to different answers rather than using the answers to sort the species into discrete categories as in rule-based methods. The points are then summed across all criteria, giving rank scores for each species. The summary scores may be simple addition of the individual criteria scores or in more complex systems may be adjusted by weights given to different criteria (i.e., the scores from some criteria may be increased by some value so they count for more in the final sum). Scoring can be used to sort species ordinally (by relative score) such as for priority setting, or to divide them into categories by assigning point ranges to each category. Hybrid systems combine some of rule, point, and even subjective ranking approaches.

Examples of point-scoring methods

Examples of point-scoring systems that address species extinction risks, as well as other conservation issues, include Millsap et al.'s (1990) protocol for setting conservation priorities for vertebrate species in Florida and the Partners in Flight protocol for ranking North American land birds by conservation priority (Carter et al. 2000, Beissinger et al. 2000). In Millsap et al.'s (1990) system, scores are assigned based on regional status for: population trend, range size, distribution trend, population concentration, reproductive potential and ecological specialization. Criterion levels are defined by either numerical thresholds or subjective criteria. Scores are summed to give a 'biological score' for the species. Separate criteria rank the current state of knowledge of distribution, population trend, limiting factors and current conservation efforts, to produce a summary (additive) 'action score.' Recent and potential threats are not explicitly considered. Lunney et al. (1996) adapted the Millsap method for use in a larger geographical area, New South Wales, Australia.

The Partners in Flight protocol (Carter et al. 2000) uses seven parameters to rank bird status within geographic regions: breeding distribution, non-breeding distribution, relative abundance, threats to breeding, threats to non-breeding, population trend, and area importance. Scores are based on objective, quantitative criteria where possible; however, subjective assessments from expert opinion are permissible. Recent and predicted threats are considered.

Scores may be additive or the parameters may be weighted to reflect their relative importance to users or regions.

Strengths and weaknesses of point-scoring methods

Point-scoring methods have the benefits of being highly explicit and allowing detailed weighting of different factors. This approach is also intuitively appealing; many decision-making groups will automatically develop point-scoring schemes to resolve difficult, multiple criteria problems (Stewart 1992). As long as the scoring systems are not too complex (e.g., if simple additive scoring is used), they are easily understood.

When scoring systems become complex through weightings or more complex mathematics, the process can become obscure and may even become unreliable in not reflecting actual priorities or preferences (Stewart 1992). Scoring systems that seem appealing when developed often become cumbersome to implement and may be rejected in favor of simple rule systems—for example, Millsap et al.'s (1990) protocol has been replaced by an IUCN-like rule system for Florida species ranking. To be effective, scoring schemes need to be transparent and scores or criterion weights must be as objective as possible.

Hybrid rule-based and point-scoring: the Heritage ranking method

Background -- A principal example of a hybrid rule-based and point-scoring process is the Heritage ranking method. The Heritage ranking system was designed originally to help set conservation priorities among "element occurrences", which are recorded locations of species and ecological communities (Master 1991). Subsequently the ranking system has evolved and is now more explicitly linked to rating species by their extinction risk (Master et al. 2000, 2002). Initially the Heritage system was a rule-based mixture of qualitative and quantitative criteria, with overall ranks estimated subjectively by experienced staff. To allow sufficient flexibility across a spectrum of species, Heritage proponents initially had no desire for 'rigid' numerical cutoffs on the factors or combinations that 'must be satisfied' (Master 1991). With repeated revisions, however, the system is evolving to become much more quantitative, prescriptive, and detailed, and to converge on the criteria and thresholds used by IUCN, particularly the major overhaul in 2002 (Master et al. 2002). Very recently, the subjective process for overall species rankings has been converted into a draft point-scoring system that overlays mostly quantitative rule-based criteria, in an attempt to make the final Heritage rankings fully 'repeatable, transparent, and explicit' (Regan et al. in press).

Description of the Heritage method -- Appendix Table A2.3 lists the Heritage rank factors and brief definitions. The thresholds for levels within the criteria are "somewhat arbitrary," but the 2002 version resembles IUCN criteria and levels (Master et al. 2002). Until Regan et al.'s (in press) point-scoring protocol, these 12 factors were combined into a summary species (or community) ranking based on the "overall fact pattern", using the "adjudicated expert judgment" of experienced, national staff. The factor thresholds and point allocation system from Regan et al.'s draft hybrid point- and rule-based system. Uncertainty in rankings is treated through rank modifiers or range ranks (ratings that cover a spread of ranks) with supporting documentation in the Heritage database (Master et al. 2000).

The Heritage system considers all key factors that may constitute deterministic or stochastic causes for extinction (Master et al. 2000), yet compared with other systems has evolved a stronger focus on element rarity, perhaps reflecting the system's origins in reserve site selection and related conservation priority setting. Rarity is seen as a "key predictor of extinction potential" (Master et al. 2000) because even without known threats very rare elements can be threatened rapidly (Master 1991). Rarity is reflected in four of the ranking factors: number of populations, population size, range breadth, and amount of occupied habitat, and the number of extant and viable populations heavily influences the rankings (Regan et al. in review). In present form, the Heritage ranking contains a substantial subjective element in the determination of individual population (element occurrence) viability. These "EO ranks" are imbedded in the 'number of occurrences with good viability' ranking factor. NatureServe (2002) provides more than 200 pages of guidance and training on ranking the current status of (not future threats to) occurrences.

Reagan et al.'s (in press) proposed hybrid system attempts to make explicit the expert judgment previously used in overall Heritage rankings. The system considers combinations and dependencies among the ranking factors, including factor weighting. Not all factors are used in any rating, with preferred factors used if relevant data are available and alternatives or surrogates used when the primary factors are not known. The preferred factors are number of occurrences, their viability (EO ranks), population size, short-term trends, and threats. Long-term trends, geographic distribution, environmental specificity, intrinsic vulnerability, and site protection and management are alternative factors.

Strengths of the new hybrid Heritage method -- The strengths of the proposed new hybrid Heritage approach are that it is highly explicit, transparent, and mostly repeatable (see preliminary tests in Regan et al. in press), with an exception that the imbedded EO ranks are still subjective. Generally the Heritage protocol is derived from more than two decades of experience with applications in state and provincial programs across North America. The protocol is also readily available for testing and modification with NatureServe support. NatureServe intends for the Heritage approach to be consistent with IUCN rankings as it is further refined over the next few years. The complex, hybrid ranking approach could be suitable for taxa-specific applications depending on further development and analysis. The ranking factors encompass all possible information sources of relevance to ESA species classification (i.e., the five ESA ranking factors, including "other").

Weaknesses of the new hybrid Heritage method -- As with all the other rule- and point-based systems, Heritage rankings may not reflect actual correlations with extinction risk (Regan et al. in review). In particular, it is not clear that the higher weighting given to number of occurrences (e.g., rarity) is appropriate for the shorter time frame extinction risks relevant to ESA species classification. The new hybrid approach needs more development and will continue to require technical support and training for application because it is a fairly complex ranking protocol. Portions of the protocol (e.g., EO ranks) remain subjective and somewhat obscure.

Research and development needs for the Heritage method -- The hybrid protocol needs to be tested against models and diverse taxa examples (Regan et al. in review, National Research Council 1995). Testing should consider the minimum set of required information for reliable classification under explicit extinction risk levels. NatureServe intends to continue working toward coherence with the IUCN system, especially in risk category levels and ranking factor cutoff levels, but both of these systems need reliability testing. If the hybrid approach is pursued farther, the next major step should be developing explicit, objective criteria for EO ranking or deleting this subjectivity from the protocol.

Structured Expert Opinion

Introduction to use of expert opinion

Expert opinion is sometimes the only source of information for assessing the viability of species and, even when objective data are available, it will often be an important input in various species classification methods (Maguire and Cochrane 2001). Ideally, "expert-opinion elicitation should not be used in lieu of rigorous reliability and risk analytical methods, but should be used to supplement them and to prepare for them" (Ayyub 2001:234). Endangered Species Act time frames for decision-making (e.g., petition findings) and the number of species potentially warranting ESA consideration may induce a greater reliance on expert opinion than would be optimal in a data- and time-rich environment. The key to defensible use of expert opinion is appropriate structuring, whether in combination with or in lieu of objective analysis.

A structured approach for eliciting expert opinion and incorporating it into species risk assessment is essential to gain the most benefit from existing knowledge, to minimize errors and biases in subjective estimates, and to make the process as transparent, repeatable, and hence, defensible as possible. The discipline of decision analysis provides tools and approaches for using expert opinion. Maguire and Cochrane (2001) summarize the steps in eliciting expert opinion and the proper use of experts for species viability analysis, as Cleaves (1994) does more generally for natural resource planning. More in-depth general references on decision analysis include Clemen and Reilly (2001) and Goodwin and Wright (1999), while the Ayyub (2001) and Cooke (1991) texts give excellent guidance on eliciting expert scientific opinions involving uncertainty. The following section briefly describes what it means to structure expert opinion.

Description of structured expert opinion

Structuring expert input encompasses many steps such as identification of appropriate experts, decomposing the decision problem into specific questions that can be addressed by relevant expertise, controlling for bias through design of questions, interactions with experts, and response analysis, quantifying responses or otherwise making them transparent, training in elicitation and response, and other measures. Depending on the complexity of the listing decision, expert knowledge can be elicited informally from individuals up to highly organized panels or integration with quantitative modeling. For ESA application, the ultimate purpose in estimating risk is to determine whether species fall above or below classification thresholds, including incorporation of uncertainty. This task is less difficult than estimating absolute extinction risk (McCarthy et al. 2003). Depending on the circumstances and level of detail or rigor desired, it can take from a day to many weeks to complete a risk analysis using structured expert opinion.

Since human judgment is easily influenced by how questions are posed, what external information is available at the time, and by interactions between people—even among experts—expert opinion will be most reliable when its elicitation is carefully facilitated. The examples described below illustrate some valuable tools for controlling bias and increasing rigor in expert opinions, such as modified Delphi approaches for working with groups and how to elicit quantitative responses. Steps preceding the actual elicitation can be critical, including selecting the experts and preparing them in to address probabilities and uncertainties. The references cited in the Introduction to Expert Opinion give additional resources for facilitating expert opinion processes.

One tenet of decision analysis is that complex problems can be analyzed more accurately if they are "decomposed" into component questions before considering the final, overarching issue. The benefits of decomposition include improving the performance of contributing experts both by tapping into the particulars of their expertise and by helping them to "think through" a large, messy problem that taxes even the most experienced of human brains. Another advantage of problem decomposition is improving decision-makers' and ultimately the public's understanding of the issues involved. Dissecting the causes of extinction risk, for example, is more informative than simply estimating risk without distinguishing minor from major factors, or revealing the areas of greatest uncertainty or disagreement and their influence on the conclusions. Decomposed and structured analysis illuminates the experts' chain of logic behind risk estimates, which can provide a more defensible basis for agency decision-making

Since species risk analysis is a complex problem, it is typically best not to rush experts to produce a single estimate of a species' comprehensive extinction risk without preparation, nor to simply ask how a species should be classified under the ESA.[1] Intuitive, summary judgments of a species' overall risk level may be biased either from lack of knowledge, limited analysis, or personal motivation (anticipating the listing result). Expertise is derived from experience. Because extinction is a very rare event, scientists cannot develop true expertise in extinction likelihood prediction.

Extinction risk problems can be decomposed by developing an 'influence diagram' or mental model of how environmental and intrinsic forces affect populations (Maguire and Cochrane 2001). By revealing and then discussing these details, scientists can "think through" their analysis more carefully, while providing documentation for the decision-making record. Experts estimate the likelihood of probabilistic events as well as uncertainties in their own projections based on their "degree of belief" in different outcomes. Scientific and personal

[1] Under the recommendations in this report, establishment of classification criteria for the levels of risk associated with ESA threatened and endangered categories is separated from the scientific task of projecting the species' future status to be compared with the classification standards.

uncertainties can be elicited and documented in various ways, such as asking experts to allocate "likelihood points" among a range of possible outcomes (e.g., time to expected extinction or different future scenarios). The final, comprehensive extinction risk estimate may result from mathematical computations in a model such as an "event tree" (Goodwin and Wright 1999) or Bayesian belief network (e.g., Lee and Rieman 1997), or the experts may be asked to provide the overall risk estimates directly for use in a classification decision framework (i.e., a rule system). Thus, structured expert opinion can stand alone when decisions are to be made by "best professional judgment," or can be complementary to more quantitative methods.

In summary, structured expert opinion is a useful approach for species risk analysis under one or more of these conditions:

- time is insufficient to complete more objective, quantitative techniques (e.g., petitions),
- data are limited and simple classification guidelines are not available or appropriate,
- data are inadequate for quantitative assessment yet detailed species-specific analysis is desired (for example, projecting future threats),
- data are substantial, but still inadequate to capture all relevant factors or their interactions for a risk analysis,
- quantitative analyses have been conducted, but need to be integrated into a structured decision-making process.

Examples of the use of structured expert opinion

U.S. Forest Service examples -- The U.S. Forest Service pioneered the use of structured expert opinion for species viability analysis. Expert panels were first convened to consider the impacts of 10 Northwest Forest Plan alternatives on habitat for 1,120 species (FEMAT 1993). The panel process was used again to evaluate 11 Tongass National Forest Plan revision alternatives for 16 species (Shaw 1999), 8 alternatives each on three national forests in Minnesota and Wisconsin (176 species; Mighton et al. 2000, Schenck et al. 2002), and also for the Midewin National Tallgrass Prairie management plan (USDA Forest Service 2002) and the Interior Columbia Basin broad-scale assessment process (Lehmkuhl et al. 1997). Experts on the Forest Service panels individually assigned 100 "likelihood" points among alternative potential outcomes, such as "suitable ecological conditions are broadly distributed and of high abundance across the historical range of the species within the planning area. …[These] conditions provide opportunity for continuous or nearly continuous intraspecific interactions for the species" over a specified time frame (Schenck et al. 2002). Some evaluations focused only on habitat, while others were decomposed into environmental and population questions; and either one or two planning horizons were evaluated (e.g., 10 and 100 years).

The results represent the expert's degree of belief in the alternative future outcomes (rather than statistical probabilities), with their uncertainty expressed by how much they spread the points among the choices. Each expert completed the exercise independently followed by either open discussion or a modified Delphi process (Cleaves 1994) and an opportunity to individually revise point assignments, but without any attempt at consensus. The results were later averaged across the experts, although all individual ratings were included in the record to emphasize the uncertainties in the belief statements. Generally both the process and results were peer reviewed.

In other planning efforts, the Forest Service used expert opinion without convening panels. For the Sierra Nevada Forest Plan covering 11 national forests in California, experts provided outcome likelihood ratings similar to the panel process but individual experts provided input for single species (USDA Forest Service 2001). In the Interior Columbia Basin broad-scale assessment, individual expert opinion was combined with empirical data to analyze species viability with Bayesian belief network models (Raphael et al. 2001; Marcot et al. 2001).

NOAA Fisheries salmonid example -- During the 1990s, NOAA Fisheries conducted a series of reviews of the status of west coast populations of Pacific salmon and steelhead (*Oncorhynchus* spp.) that relied substantially on expert opinion. Initially, reviews were completed in response to petitions, but in 1994 the agency began a series of proactive, comprehensive ESA status reviews of all populations (evolutionarily significant units or ESUs) of anadromous Pacific salmonids from Washington, Idaho, Oregon, and California. Status reviews are presently being updated as part of recovery planning. In its biological status reviews, the Biological Review Team (BRT) draws scientific conclusions about the current risk of extinction faced by ESUs, under the assumption that present conditions will continue into the future (recognizing that natural demographic and environmental variability are inherent features of "present conditions"). The team does not evaluate possible future effects of protective efforts, except to the extent the effects are already reflected in metrics of population or ESU viability, because those efforts are taken into account in a separate process by the NOAA Fisheries regional offices prior to making listing determinations. Therefore, the BRT does not make recommendations as to whether identified ESUs should be listed as threatened or endangered species.

Salmon face a bewildering array of potential threats throughout every stage of their complex life cycle and it is difficult to evaluate the relative importance of a wide range of interacting factors. The BRT does not attempt a rigorous analysis of each factor that has contributed to historical salmon population declines. Instead, the status reviews question whether an ESU is presently at risk, regardless of the reasons for its initial decline. They consider contemporary factors for decline and the extent to which these have been alleviated by existing protective efforts.

Salmon risk assessment is addressed at two levels—population and overall ESU—since salmonid ESUs are typically metapopulations. Individual populations are assessed according to the four viable salmonid populations (VSP) criteria: abundance, growth rate/productivity, spatial structure, and genetic diversity (McElhany et al. 2000). Larger-scale issues are considered in evaluating the status of the ESU as a whole, such as total number, geographic distribution, and connectivity of viable populations. The results are compiled in a risk matrix (table) that allows comparison within and across ESUs.

After reviewing all relevant biological information for a particular ESU, each BRT member assigns a risk score on a scale of 1 (very low risk) to 5 (very high risk) to each of the four VSP criteria (risk levels are defined qualitatively). Scores are also provided for "recent events" such as floods that will have predictable consequences for ESU status in the future but have occurred too recently to be reflected in the population data or VSP criteria. Recent events are scored on a subjective, five-level scale from "expect strong improvement" (++) to "expect strong decline" (--) of the ESU.

The BRT analysis of overall risk to the ESU uses categories that correspond to the narrative endangered and threatened species definitions in the ESA, reflecting professional judgments by each BRT member. This assessment is guided by the results of the risk matrix analysis as well as expectations about likely interactions among factors. Although the VSP and recent event scores helps to integrate and quantify a large amount of diverse information, there is no simple way to translate the risk matrix scores directly into an assessment of overall risk. For example, simply averaging the values of the various risk factors would not be appropriate; an ESU at high risk for low abundance would be at high risk even if there were no other risk factors. Each BRT member distributes 10 likelihood points (see FEMAT, above) among the three ESA risk categories (endangered, threatened, not threatened), reflecting their opinion of how likely that category correctly reflects the true ESU status. This method has been used in all status review updates for anadromous Pacific salmonids since 1999 and in slightly different form in the status reviews during 1991-1999.

U.S. Fish and Wildlife Service examples -- The USFWS has limited experience with convening expert panels to assist with listing decisions. In 1997, Region 7 convened panels for risk analysis for two species petitioned for listing in Southeast Alaska, the Alexander Archipelago wolf (*Canis lupus ligoni*) and Queen Charlotte goshawk (*Accipiter gentiles laingi*) (U.S. Fish and Wildlife Service 1997a,b).

USFWS employed structured decision analysis techniques in a panel process to
(1) Elicit objective and informed, expert opinions on biological status and threats in the absence of rigorous, quantitative analysis and supporting data;
(2) Express clearly the full range of relevant biological information and qualitative assessments, especially addressing and documenting uncertainties;
(3) Involve the principle decision-maker throughout the deliberations to ensure full understanding of expert opinions and their uncertainties;
(4) Guide the group through serial evaluations and create a concise written record that documented adherence with ESA stipulations; and
(5) Produce a decision where human judgments were explicit and directly related to quantified decision standards—hence, a rational and defensible decision.

Panels of three to five biologists with knowledge of the species or habitat completed the analysis for each species in one day of structured exercises.[2] While using the likelihood points concept from FEMAT, the USFWS risk analysis panels were both more detailed in their analysis than the Forest Service panels and customized to ESA questions. Panelists were not asked for summary judgments on the species' extinction risk or recommended listing status. Instead, the exercises were developed to lead the panel through a series of questions about the availability of

[2] Because of FACA concerns, all panelists were Federal employees. Similarly, only Federal and State employees worked on the Tongass National Forest panels.

information and the species' status tied to the legal analysis steps in a petition finding. By considering a single species rather than the dozens or hundreds covered by the Forest Service panels, the analysis could be decomposed to reduce motivational bias ("jumping to conclusions") and reveal the underlying issues and uncertainties to the decision-maker (regional director) who observed most of the panel process. Appendix Table A2.4 outlines the series of exercises completed by the Alaskan panels.

In the USFWS exercises, experts assigned 100 likelihood points to whether specific potential threats (previously discussed and defined carefully) would occur at all, would cause population declines below a defined endangered level of extinction risk, and would cause declines below the endangered level within a defined foreseeable future which was based on timber harvest and forest regeneration cycles. The threat ratings were then compiled into event trees showing all possible combinations of threats (Goodwin and Wright 1999), producing summary figures for the likelihood that species would decline beyond key thresholds due to any/all of the threats (the likelihood of threats in combination is derived by multiplication).[3] In addition to distributing points between "yes" and "no" answers, in each exercise the experts could express their uncertainty around those estimates in various forms (numerical point ranges or drawing visual distributions). Where appropriate, the exercises were repeated for separate portions of a species range as well as across the entire range.

All exercises were conducted with a modified Delphi process where initial ratings were independent and the results presented anonymously for group discussion, with the opportunity to independently revise the ratings after the discussion. Because of the small panel size and single species analysis, it was possible to retain all experts' ratings separately in the summary charts and reports; no attempt was made to average or merge the answers. The panel results were thus quite variable and not intended to prescribe a decision, yet the results combined with the in-depth discussions stemming from the structured exercises were highly influential in the petition-finding decisions.

In 1998, USFWS Region 3 completed a similar panel process to evaluate whether the eastern Massasauga rattlesnake (*Sistrurus catenatus catenatus*) should be added to the candidate species list (U.S. Fish and Wildlife Service 1999). Because this action would be discretionary and was not subject to a petition, the panel process was less formal than the Alaskan panels. A dozen panelists from various organizations and agencies participated, bringing expertise from different parts of the species' range in the Midwestern United States and Ontario, Canada. While the exercises were similar to the sequence in Appendix Table A2.4, they were abbreviated to rating likelihood of extinction in a 10-generation time frame due to four threats (because of the larger group size and lesser advance preparation and motivation). The experts were uncomfortable providing ratings outside their geographical region of personal familiarity (in part because of professional "turf" issues), so the results were collected primarily by separate "evaluation units," leaving the USFWS to assess range-wide status.

In a candidate species status review for the Dakota skipper (*Hesperia dacotae*) (a prairie butterfly), expert opinion was gathered individually by telephone without convening a panel, but

[3] For these species, the experts determined that the potential threats were independent events so contingencies or order of occurrence did not need to be addressed in the event trees or likelihood ratings.

following a structured series of questions (Cochrane 2002). The ranking process used FEMAT-style likelihood point ratings, but varied from the previously described analyses by having the choice of "outcomes" be likelihoods of a specific decline or extirpation. For each of 8 distinct threats at specified clusters of occupied sites (local metapopulations), experts assigned 100 points among five levels of likelihood that the cluster would decline at all over 20 years.[4] For each cluster they assigned 100 points among five levels of likelihood that the cluster would be extirpated within 20 years due to all threats cumulatively.[5] Experts also peer-reviewed ratings of population status at all known occupancy sites using a quantitatively defined scale (Appendix Table A2.5).

Strengths and weaknesses of structured expert opinion

By using expert opinion we can take full advantage of existing knowledge. Structuring the elicitation is important to maximize the amount of useful information attained while fully exposing uncertainties and minimizing errors and biases inherent to subjective assessments. Structuring can help separate specific aspects of risk analysis from risk management. When parameters and their uncertainty can't be directly measured, experts can provide quantitative estimates that improve decision-making. Indeed, "quantitative analysis may not be any more accurate or more useful than direct human judgment. The appearance of precision given by numerical outputs of [quantitative] analyses can mask the fact that the analyses and their interpretations reflect human judgment." (Cleaves 1994:1). Depending on the number of experts consulted and the structuring (e.g., panels or not), the approach can be time- and cost-effective compared with relying strictly on data collection and quantitative analysis.

By definition expert opinion is subjective so results usually will not be precisely repeatable. Expert opinion approaches are limited by lack of knowledge and susceptible to personal and institutional biases. In application, expert opinion approaches often involve challenges of working with strong personalities and facing non-biological concerns that technically should not be part of species risk analysis. Selection of appropriate experts is critical to the development of appropriate listing determinations. To aid decision makers, participants in such processes must exhibit expert professional judgment in addition to expert knowledge.[6]

Research and development needs for structured expert opinion

As with most other available methods for species viability assessment, results from structured expert opinion have not been tested for reliability (e.g., against models or other independent risk predictors). The testing should focus on classification accuracy and consistency rather than absolute extinction risk prediction (compare with recent analysis of how well PVA performs in classification despite substantial uncertainty in extinction risk estimates; McCarthy et al. 2003). We need to better understand how to decompose viability analysis to take full advantage of expert knowledge while controlling variability and biases as much as possible. A

---

[4] The 'outcome' likelihoods were >0-25%, >25-50%, >50-75%, >75-95%, and >95%.
[5] In this exercise, the outcome likelihoods were 0%, >0-5%, >5-20%, >20-50%, and >50%.
[6] Following Cleaves (1994), rational judgment thoroughly uses available information with an awareness of implications, is consistent with similar judgments, agrees with general laws of probability, and is understood by others.

key component of improving performance will be training and repeated experience linked to feedback on performance—for both experts and facilitators.

Appendix Table A2.1. Summary of quantitative features of the IUCN rule-based approach. See IUCN (2001) for a much more complete description of the criteria. Bolding indicates the differences between the classifications of "critically endangered", "endangered", and "vulnerable". (wl) = "whichever is longer, up to a maximum of 100 years".

Critically endangered

A. Reduction in population size
- ≥ **90%** decline in past 10 years or 3 generations (wl), if understood and reversible and stopped
- ≥ **80%** decline in past 10 years or 3 generations (wl), if not understood or reversible or stopped
- ≥ **80%** decline projected for next 10 years or 3 generations (wl)
- ≥ **80%** decline including past and future 10 years or 3 generations (wl), if not understood or reversible or stopped

B. Geographic range
- extent of occurrence < **100** km^2
- area of occupancy < **10** km^2

C. Population size < **250** mature individuals and:
- continuing decline ≥ **25%** in future **3** years or **1** generation (wl)
- no subpopulation with > **50** mature individuals, or ≥ **90%** mature individuals in one subpopulation

D. Population size < **50** mature individuals

E. Quantitative analysis showing Pr(extinction) ≥ **50%** within **10** years or **3** generations (wl)

Endangered

A. Reduction in population size
- ≥ **70%** decline in past 10 years or 3 generations (wl), if understood and reversible and stopped
- ≥ **50%** decline in past 10 years or 3 generations (wl), if not understood or reversible or stopped
- ≥ **50%** decline projected for next 10 years or 3 generations (wl)

- ≥ **50%** decline including past and future 10 years or 3 generations (wl), if not understood or reversible or stopped

B. Geographic range
- extent of occurrence < **5000** km^2
- area of occupancy < **500** km^2

C. Population size < **2500** mature individuals and:
- continuing decline ≥ **20%** in future **5** years or **2** generations (wl)
- no subpopulation with > **250** mature individuals, or ≥ **95%** mature individuals in one subpopulation

D. Population size < **250** mature individuals

E. Quantitative analysis showing Pr(extinction) ≥ **20%** within **20** years or **5** generations (wl)

Vulnerable

A. Reduction in population size
- ≥ **50%** decline in past 10 years or 3 generations (wl), if understood and reversible and stopped
- ≥ **30%** decline in past 10 years or 3 generations (wl), if not understood or reversible or stopped
- ≥ **30%** decline projected for next 10 years or 3 generations (wl)
- ≥ **30%** decline including past and future 10 years or 3 generations (wl), if not understood or reversible or stopped

B. Geographic range
- extent of occurrence < **20,000** km^2
- area of occupancy < **2000** km^2

C. Population size < **10,000** mature individuals and:
- continuing decline ≥ **25%** in future **10** years or **3** generations (wl)
- no subpopulation with > **1000** mature individuals, or **100%** mature individuals in one subpopulation

D. Population size < **1000** mature individuals

E. Quantitative analysis showing Pr(extinction) ≥ **10%** within **100** years (wl)

Appendix Table A2.2. Summary of quantitative features of the CITES rule-based approach. (CITES 1994)

Biological Criteria for Appendix I (international trade banned)

A. The wild population is small (numeric guidelines: < 5,000 for a population; < 500 for subpopulations)
B. The wild population has a restricted area of distribution (numeric guidelines: < 10,000 km^2 for a population; < 500 km^2 for subpopulations)
C. The number of individuals in the wild has declined (numeric guidelines: > 50% in 5 years or 2 generations, whichever is longer; or, for a small wild population, > 20% in 10 years or 3 generations, whichever is longer)
D. The status of the species is such that it is likely to satisfy one or more of the above criteria within a period of 5 years

Criteria for Appendix II (international trade regulated)

A. Unless trade is strictly regulated, it is likely that at least one of the above criteria will be met in the near future
B. Harvesting for international trade is likely to have a detrimental impact because the harvest rate is not sustainable, or because the population has been reduced to a size at which its survival could be threatened by other influences

Appendix Table A2.3. Definitions of the factors used in assessing Heritage conservation status (from Reagan et al. in press and Master et al. 2002). Further details are in Master et al. (2002) and NatureServe (2002).

Factor	Definition
Number of occurrences	Number of distinct populations or subpopulations.
Viability of occurrences or ecological integrity of communities	Relative viability or likelihood of persistence, based on their size, condition, and landscape context.
Population size	Number of mature individuals (species only).
Area of occupancy	Total area of occupied habitat across the range.
Range extent	Extent of overall geographic range.
Trends	Short and long term increase or decrease in population size, area of occupancy, or condition of occurrences.
Threats	Known or suspected current threats, or likely future threats.
Protected occurrences	Number of adequately protected and managed populations.
Intrinsic vulnerability	Inherent susceptibility to threats due to intrinsic biological factors.
Environmental specificity	Vulnerability or resilience of the element due to habitat preferences or restrictions or other environmental specificity.

Appendix Table A2.4. Structured exercises from the southeast Alaska wolf and goshawk panels (abbreviated from U.S. Fish and Wildlife Service 1997b).

Sequential Steps and Exercises	**Description**
Natural history profile	Use conservation biology principles to assess species' inherent vulnerability and resiliency to decline (e.g., ecological specialization, lifespan). **Exercise**: likelihood ratings on how important each defined factor is to the species (outcomes: low, medium, high).
Information review—listing factors	Review and reaffirm status report information, particularly extrinsic factors like habitat trends.
Define foreseeable future	Establish quantitative definition (years) based on species life history and relevant land use or management planning time frames.
List factors that may be threats	Enumerate all natural and anthropogenic factors that *could* cause or contribute to a long-term population decline (excluding highly conjectural or unlikely events); demonstrate for the record that the risk analysis has been comprehensive.
Evaluate whether factors are threats (will cause declines beyond natural range)	Reduce list to key factors (build influence diagram). The following two exercises indicate whether information is sufficient to determine that a decline could occur, thus affirming the 90-day finding that the petition "may be warranted." **Exercise**: likelihood ratings for whether specific factors would cause a long-term decline in the population *if they occurred* (outcomes: yes, no).
Estimate likelihood threats will occur	**Exercise**: likelihood rating for whether the threats will occur in specified time frame (outcomes: yes, no). (First confirm assumption of independent likelihoods for all threats or perform ratings on specified combinations). Threats that "survived" to this point (i.e., were judged to be both possible and to cause a decline by at least one expert) were transferred to scoring sheets for subsequent exercises on ranking magnitudes of threats in relation to listing criteria.
Define 'endangered'	Establish working, quantitative definition in terms of extinction likelihood and time (IUCN definition used for lack of alternative).
Estimate likelihood species is endangered	**Exercise**: likelihood ratings for whether specific factors would cause a decline to or exceeding the endangered threshold (outcomes: yes, no).

Estimate likelihood species will become endangered in foreseeable future	**Exercise**: likelihood ratings for whether specific factors would cause a decline to or exceeding the endangered threshold within the foreseeable future (e.g., would have a 20% probability of extinction in 20 years, at any time within 100 years) (outcomes: yes, no).
Cast hypotheses to test in decision monitoring	Discussion about the risks of making a "wrong" decision and what research questions or hypotheses could be tested to increase certainty about population decline or status. A wrong decision from a strictly biological perspective—the only perspective explicitly allowed in the ESA—would happen when a species is truly declining toward extinction, but the agency failed to propose listing. Also, how long would it take to gather sufficient information to resolve the species' true biological status?
Projected hindsight (potential consequences of decision error)	Discuss the likelihood that irreversible harm or long-term threats would occur during the time needed to improve certainty about the species' status (see hypotheses, above), if the decision was not to classify the species under the ESA.

Appendix Table A2.5. Status level definitions from the Dakota skipper status review (from Cochrane 2002).

Level	Definition
Secure	Inherently viable by size or other characteristics; no active threats (< 5% probability extinction within 50 years).
Vulnerable	Possibly not viable due to isolation or other factors; threats may affect (not secure, but < 20% prob. extinction within 20 years).
Threatened	Active threats and/or high inherent vulnerability (≥ 20% prob. extinction within 20 years).
Extirpated	Converted habitat or degraded habitat and no recent observations despite active searching.
Unknown	Site information lacking.

APPENDIX 3: ALTERNATIVE DEFINITIONS OF "EXTINCTION RISK"

Three alternative approaches to defining extinction risk will be considered here. All three approaches involve computing the area under a curve. In each approach, the curve consists of the product of a probability function and a weighting function. The approaches are described in detail below. Before presenting these descriptions, however, it will prove helpful to develop several concepts.

Two hypothetical species, designated "A" and "B", will be used to illustrate the concepts needed to describe the three alternative approaches.

Species Dynamics

Other than the case where a catastrophic event leads to an immediate extinction event, a species' abundance must decline over time prior to an extinction event. The rate of change in abundance is therefore a fundamental concept in species dynamics and conservation biology. In the simplest models, the rate of change is proportional to species abundance. In more complicated models, the relationship between rate of change and species abundance is nonlinear. In particular, it is possible for the expected rate of change to become negative if species abundance falls below a *depensatory threshold*. When abundance falls below the depensatory threshold, the species' dynamics undergo a qualitative change and the species is drawn (though perhaps not inexorably) toward extinction.

Appendix Figure A3.1 shows how the expected rate of change varies with abundance in the cases of Species A and Species B. For both of these hypothetical species, the expected rate of change is negative for all abundances less than 1,000 individuals and positive for all abundances greater than 1,000 individuals. An abundance of precisely 1,000 individuals thus represents the depensatory threshold for both of these examples. On either side of the depensatory threshold, the rate of change varies more appreciably in the case of Species A than in the case of Species B.

Probability Functions

Among the things that complicate attempts to model the process of extinction is random natural variability. While it might be possible to compute the most likely time for extinction to occur, this estimate will almost always be highly uncertain. In other words, extinction could occur at any time, but some extinction times are more probable than others. This idea is conveniently expressed by a *probability function*. A probability function shows the relative likelihood of alternative states or outcomes. Appendix Figure A3.2 shows how the probability of extinction varies with time for Species A and B. The probability functions in Figure A3.2 are based on the dynamics illustrated in Appendix Figure A3.1 for the special case in which the initial abundance of Species A is 1,100 individuals and the initial abundance of Species B is 10,000 individuals. Appendix Figure A3.2 shows that an extinction time of about 25 years is more likely than any other value for both Species A and Species B.

Another factor that complicates attempts to model the process of extinction is statistical imprecision. For example, although a census, survey, or statistical model may provide an *estimate* of a species' abundance, the species' *true* abundance will typically be uncertain. Appendix Figures A3.3a and 3b illustrate this type of uncertainty for Species A and B (note that the scales on the axes differ between these two figures). The curves in these figures relate each potential abundance to a probability. In Appendix Figure A3.3a, the range of likely abundances is fairly narrow, whereas the range of likely abundances in Appendix Figure A3.3b is quite large.

It should be noted that the probability functions in Appendix Figures A3.3a and 3b are completely independent of the probability functions in Appendix Figure A3.2. However, the fact that abundance is uncertain means that extinction is somewhat more difficult to predict than indicated by the probability functions shown in Appendix Figure A3.2. Once the probability functions in Appendix Figure A3.2 have been adjusted so as to reflect the uncertainty in abundance described by Appendix Figures A3.3a and 3b, the probability functions shown in Appendix Figure A3.4 result. For Species A (solid curve), the probability function in Appendix Figure A3.4 is barely distinguishable from the one in Appendix Figure A3.2 because the probability function in Appendix Figure A3.3a is so narrow. For Species B (dotted curve), on the other hand, the probability function in Appendix Figure A3.4 is noticeably broader than the one in Appendix Figure A3.2.

The area under any given portion of a probability function represents the probability associated with the corresponding interval on the horizontal axis. In the special case where an interval starts at the lower end of the horizontal axis, the probability associated with the interval is called the *cumulative probability*. Appendix Figure A3.5 shows the cumulative probability that extinction will occur before any given point along the time axis. For example, there is a 5% chance that Species A will go extinct within 20 years and a 5% chance that Species B will go extinct within 20 years. Similarly, Appendix Figures A3.6a and 6b show the area under the curves in Figures A3.3a and 3b, respectively. For a given point on the horizontal axis, the height of the curve in Appendix Figure A3.6a or 6b represents the cumulative probability that the true value falls between zero and the given point. For example, Appendix Figure A3.6a shows there is a 50:50 chance that the abundance is between zero and 1,100 individuals, and Figure A3.6b shows there is a 50:50 chance that the abundance is between zero and 10,000 individuals.

Weighting Functions

A *weighting function* describes the relative importance of alternative states or outcomes, independent of the likelihood of those states or outcomes. A weighting function is thus very different from a probability function. A probability function answers the question, "How *likely* is it that outcome X will occur?" whereas a weighting function answers the question, "How *important* would it be if outcome X were to occur?"

The weighting function labeled WF1 in Appendix Figure A3.7a is a simple example. WF1 assigns a weight of 1 to all values below a specified time horizon and a weight of 0 to all values above the time horizon. For illustrative purposes, the time horizon for WF1 has been set equal to 20 years. In practice, of course, the location of the time horizon would be a policy decision. The important feature of WF1 is that it is purely dichotomous: all extinction times are assigned a weight of either 1.0 or 0.0 depending on the time horizon.

The weighting function labeled WF2 in Appendix Figure A3.7a varies much more smoothly than does WF1. Rather than treating all extinction times lower than the time horizon as equally important and all extinction times above the time horizon as completely unimportant, WF2 views the range of extinction times as a continuum. The example shown in Appendix Figure A3.7a is governed by a single parameter (though more complicated continuous functions could also be used). As with the location of the time horizon in Appendix Figure A3.7a, the value of this parameter would be a policy decision.

The weighting function labeled WF3 in Appendix Figure A3.7b is similar to WF1 in AppendixFigure A3.7a in terms of overall shape, in that all values below some point are assigned a weight of 1 and all higher values are assigned a weight of 0. The two main differences are as follow: First, the variable on the horizontal axis in Appendix Figure A3.7a is time, whereas the variable on the horizontal axis in Appendix Figure A3.7b is abundance. Second, the value of the parameter in WF1 (the time horizon) is a policy decision, whereas the parameter in WF3 is the depensatory threshold, whose value is determined by the biology of the species.

Description of Alternative Approaches

The following table shows how probability functions associated with Species A and B can be combined with weighting functions WF1, WF2, and WF3 to yield the curves shown in Appendix Figures A3.8a and 8b. These curves, in turn, distinguish the three alternative approaches described in the next subsections.

Weighting function	Probability function: extinction time		Probability function: abundance	
	Species A	Species B	Species A	Species B
WF1 (traditional)	Appendix Figure A3.8a (black solid)	Figure A3.8a (black dashed)	n/a	n/a
WF2 (comprehensive)	Figure A3.8a (gray solid)	Figure A3.8a (gray dashed)	n/a	n/a
WF3 (threshold)	n/a	n/a	Figure A3.8b (solid)	Figure A3.8b (dashed)

Traditional Approach

The traditional approach defines extinction risk as the area under the product of WF1

(Appendix Figure A3.7a) and the extinction time probability function (Appendix Figure A3.4). In Appendix Figure A3.8a, this corresponds to the area under the red curve for Species A and the area under the blue curve for Species B. Assuming a time horizon of 20 years, the extinction risk under the traditional approach is 5% for both Species A and Species B. However, if the time horizon were 40 years, Species A would have a much higher extinction risk than Species B (38% vs. 14%), and if the time horizon were 15 years, Species A would have a much lower extinction risk than Species B (0.4% vs. 2.4%).

Threshold Approach

The threshold approach defines extinction risk as the area under the product of WF3 (Appendix Figure A3.7b) and the abundance probability function (Appendix Figure A3.3). In Appendix Figure A3.8b, this corresponds to the area under the red curve for Species A and the area under the blue curve for Species B. The threshold approach is distinguished from the traditional approach by use of a different weighting function and a different probability function. The extinction risk under the threshold approach is 5% for both Species A and Species B.

Comprehensive Threshold Approach

The comprehensive threshold approach defines extinction risk as the area under the product of WF2 (Appendix Figure A3.7a) and the extinction time probability function (Appendix Figure A3.4). In Appendix Figure A3.8a, this corresponds to the area under the pink curve for Species A and the area under the aqua curve for Species B. The only difference between the traditional and comprehensive threshold approaches is that the traditional approach uses a dichotomous weighting function whereas the comprehensive threshold approach uses a smooth weighting function. Assuming that WF2 has the shape shown in Figure A3.7a, the extinction risk under the comprehensive approach is 31% for Species A and 13% for Species B.

Evaluation of Alternative Approaches

Some characteristics of the three alternative approaches ("Trad." = traditional, "Thresh." = threshold, "Comp." = comprehensive threshold) are summarized in the table below and discussed in the paragraphs which follow.

Characteristic	Trad.	Thresh.	Comp.
Well established in the conservation biology literature	Yes	Yes	No
Directly related to extinction	Yes	No	Yes
Capable of giving non-zero weight to all possible outcomes	No	No	Yes
Free from assumptions regarding depensatory dynamics	No	Yes	No
Free from explicit policy decisions regarding an "endangered"	No	No	No
Free from explicit policy decisions regarding other parameter	No	Yes	No

Well established in the conservation biology literature

The traditional approach and the threshold approach are familiar concepts in the conservation biology literature, whereas the comprehensive threshold approach is not. Similarly, the traditional approach and the threshold approach are reminiscent of hypothesis testing, and so will be easily communicated to practitioners of that particular methodology. On the other hand, the comprehensive threshold approach is firmly rooted in decision theory, and so will be easily communicated to practitioners of that particular methodology.

Directly related to extinction

The traditional approach and the comprehensive threshold approach are both directly related to extinction because both are based on the extinction time probability function. The threshold approach, on the other hand, is not directly related to extinction because it is based on the abundance probability function rather than the extinction time probability function. The threshold approach is, however, *in*directly related to extinction because the abundance probability is evaluated relative to the depensatory threshold, being the abundance level below which the species is drawn toward extinction. However, it should be noted that the presence of a tendency toward extinction does not necessarily mean that extinction is imminent or even inevitable.

Capable of giving some weight to all possible outcomes

The comprehensive threshold approach is the only approach capable of giving non-zero weight to all possible outcomes. In principle, the weighting function used by the comprehensive threshold approach can take whatever shape is needed to reflect society's attitudes toward extinction. The traditional approach and the threshold approach, in contrast, both rely on weighting functions that regard some range of possible outcomes as equally important and the remaining possible outcomes as completely unimportant.

Free from assumptions regarding depensatory dynamics

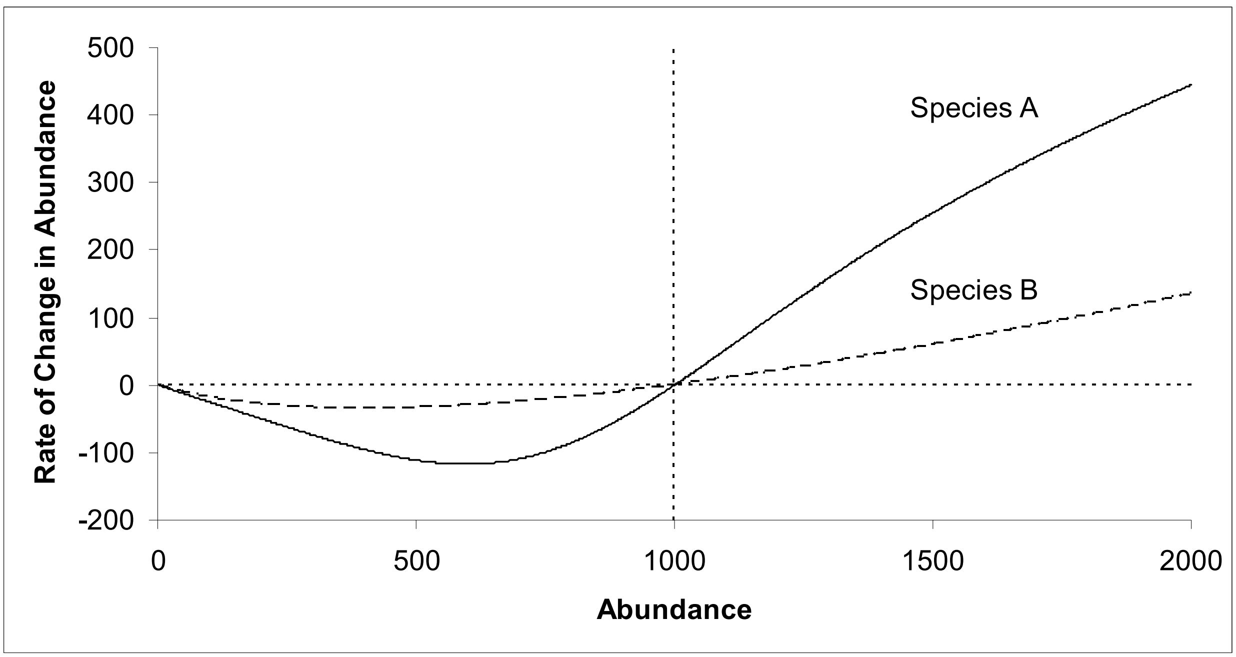
The threshold approach is the only approach that does not require modeling of species dynamics at abundances below the depensatory threshold, although it does require estimating the location of the depensatory threshold. The traditional approach and the comprehensive threshold approach likewise require estimating the location of the depensatory threshold, at least implicitly, but they also require estimating how the species will respond in the event that it falls below the depensatory threshold.

Free from explicit policy decisions regarding an "endangered" risk level

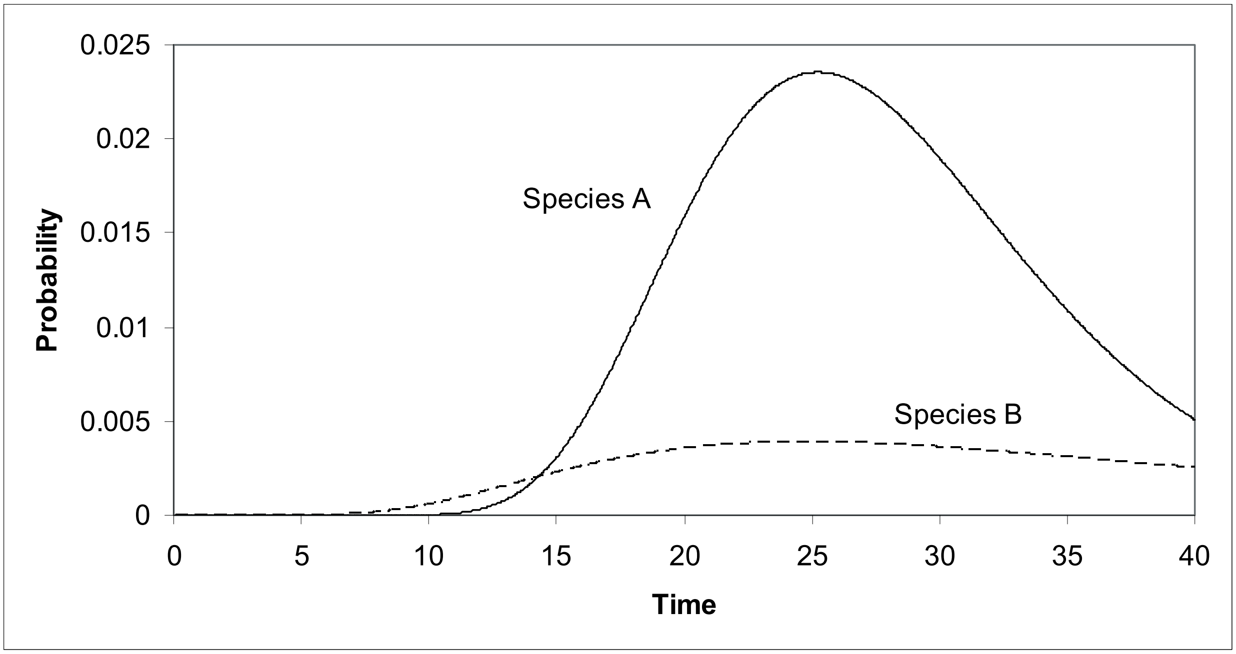
None of the approaches is free from the need to make a policy decision regarding the amount of risk that corresponds to "endangered" in the sense of the ESA. All three approaches result in a measure of extinction risk that ranges between 0 and 1, but the meaning of this measure differs between the approaches.

Free from explicit policy decisions regarding other parameter values

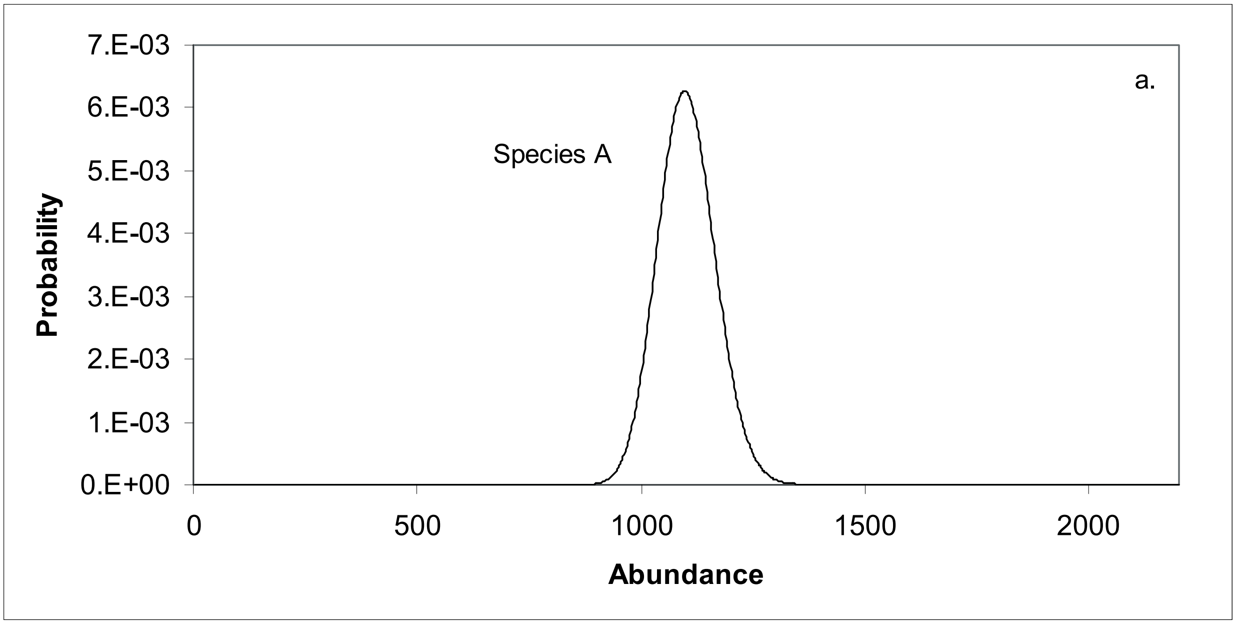
The threshold approach is the only approach that does not require explicit policy decisions regarding the value of one or more other parameters (i.e., in addition to the parameter defining the “endangered” risk level). The traditional approach requires a policy decision regarding the location of the time horizon and the comprehensive threshold approach requires a policy decision regarding the parameters determining the weighting function (note: the number of parameters need not be large; only one parameter is required for WF2, for example).

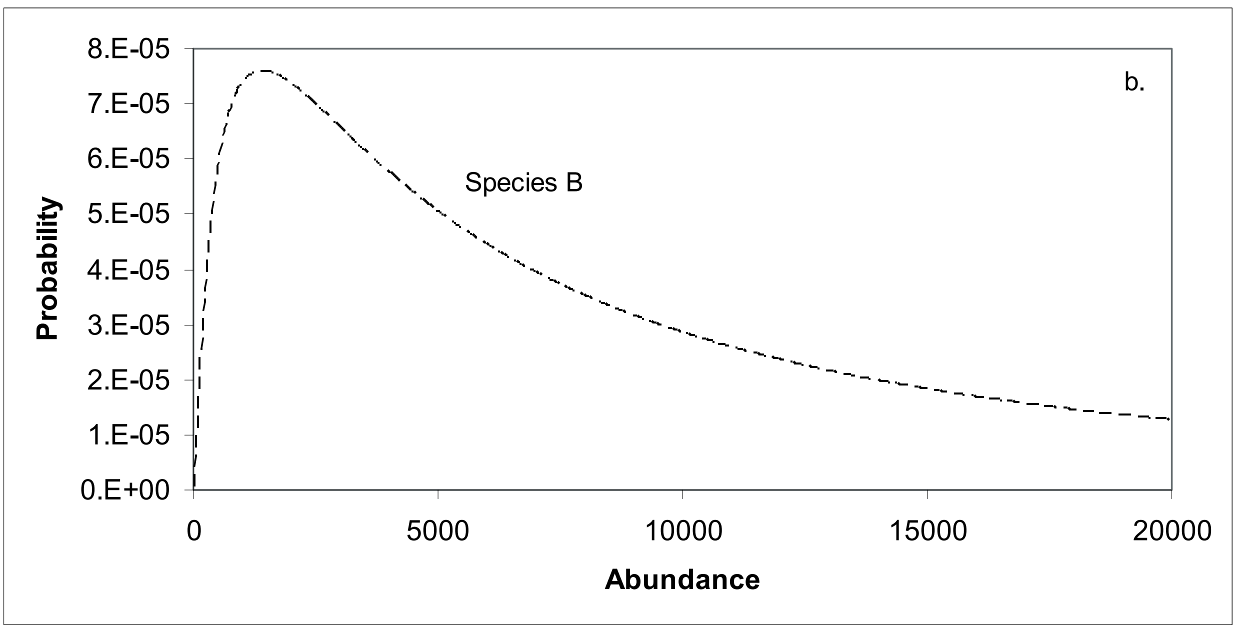


Appendix Figure A3.1. Rates of change in abundance for Species A (solid) and Species B (dashed). A depensatory threshold at 1000 individuals exists for both Species. Below the depensatory threshold, the expected rate of change is negative.

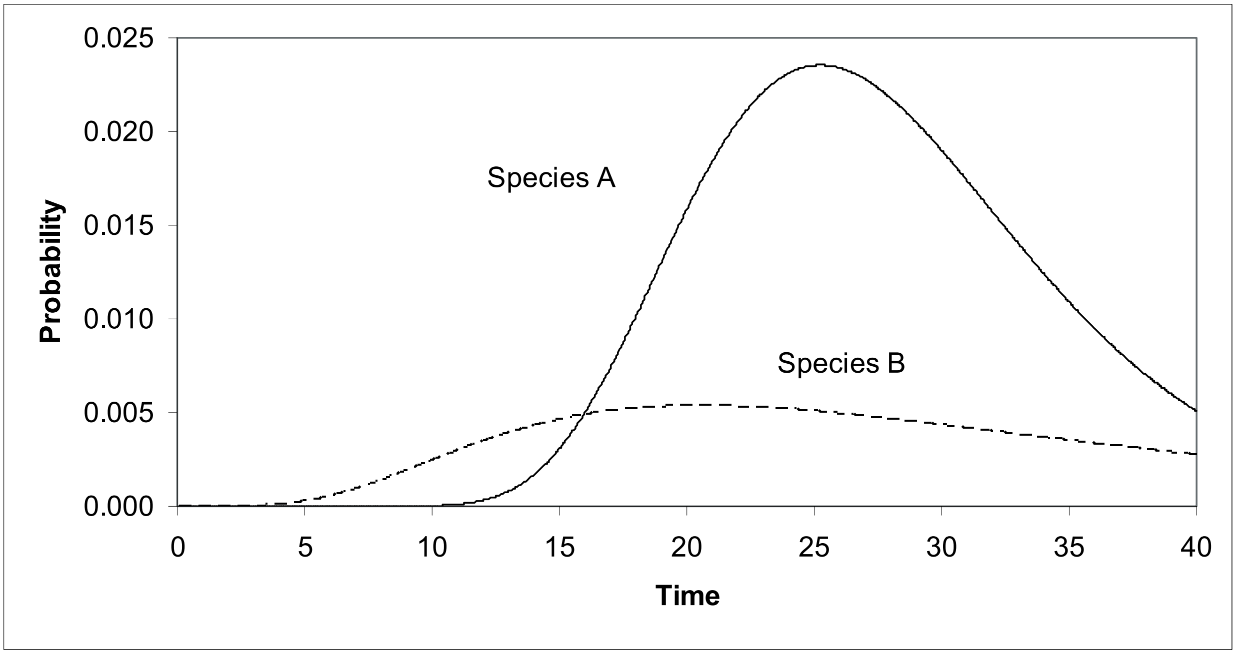


Appendix Figure A3.2. Extinction time probability functions for Species A (solid) and Species B (solid). Each curve shows the relative likelihood that extinction will occur at any given time.

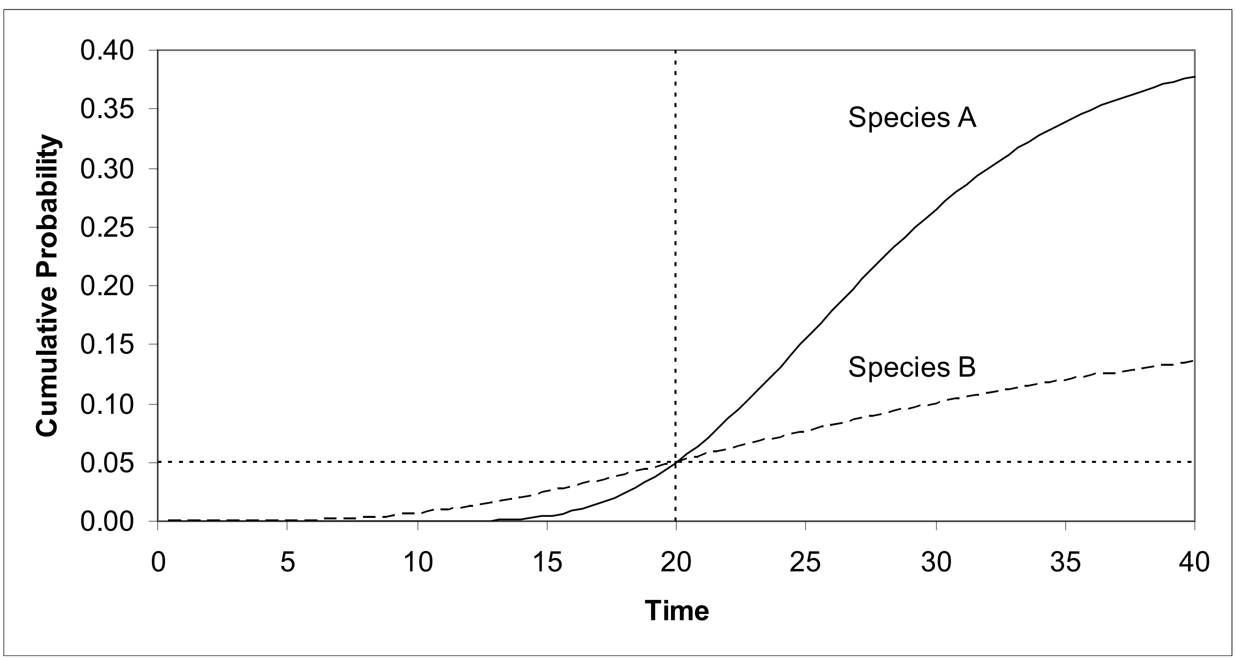




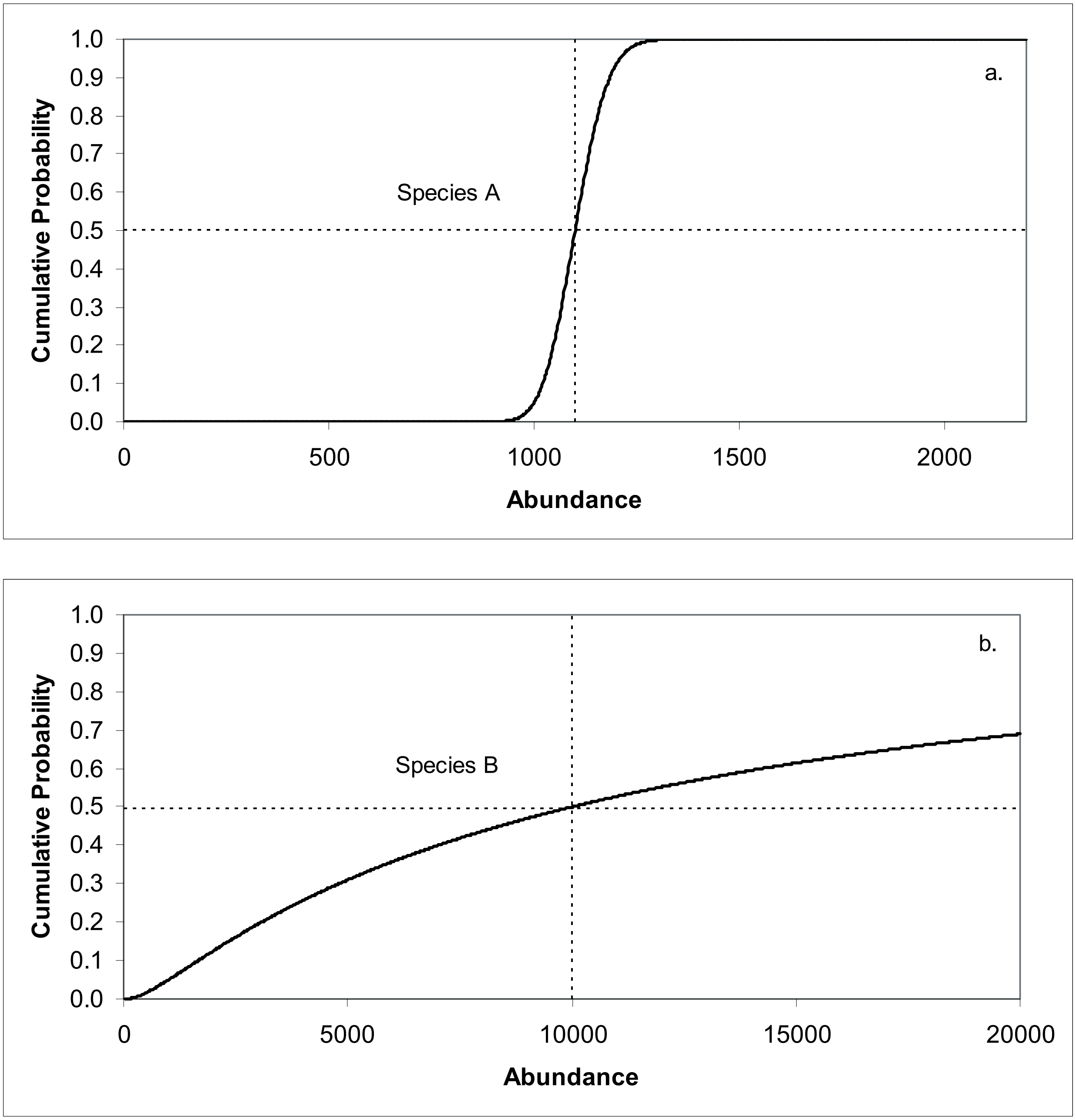
Appendix Figure A3.3. Abundance probability functions. Appendix Figure A3.3a (top): Abundance probability function for Species A. Figure A3.3b (bottom): Abundance probability function for Species B. Note that the scales on the axes differ between the two panels.



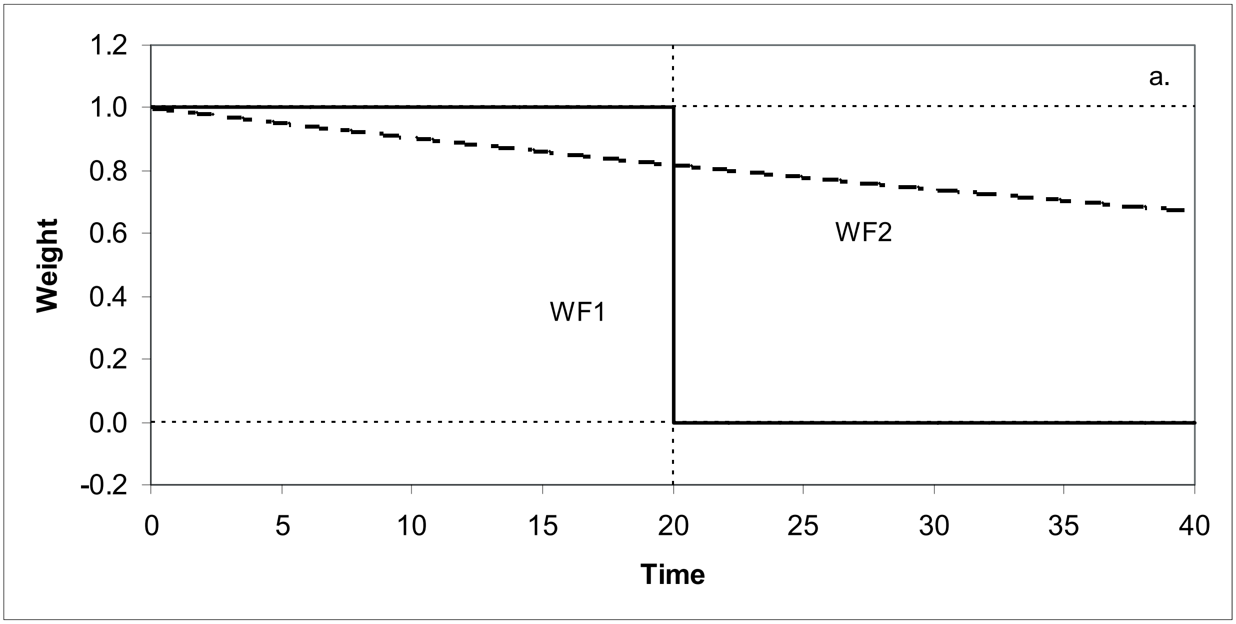
Appendix Figure A3.4. Extinction time probability functions for Species A (solid) and Species B (dashed), adjusted to reflect the uncertainty in abundance shown in Appendix Figures A3.3a and 3b.

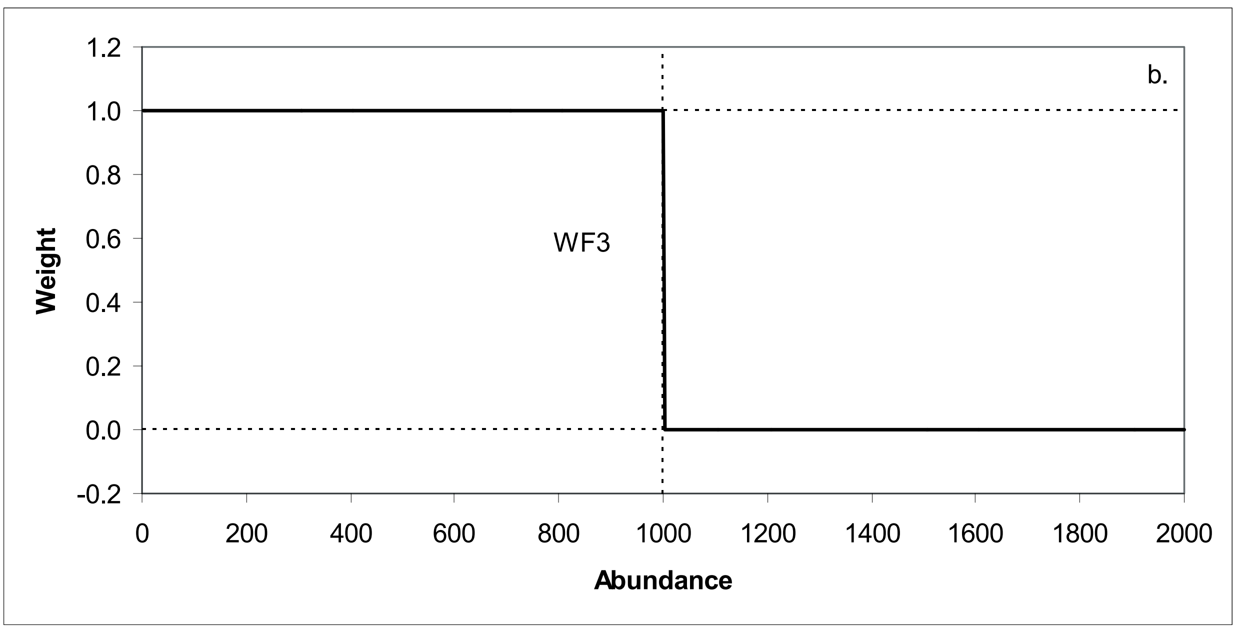


Appendix Figure A3.5. Cumulative probability of extinction as a function of time for Species A (solid) and Species B (dashed).

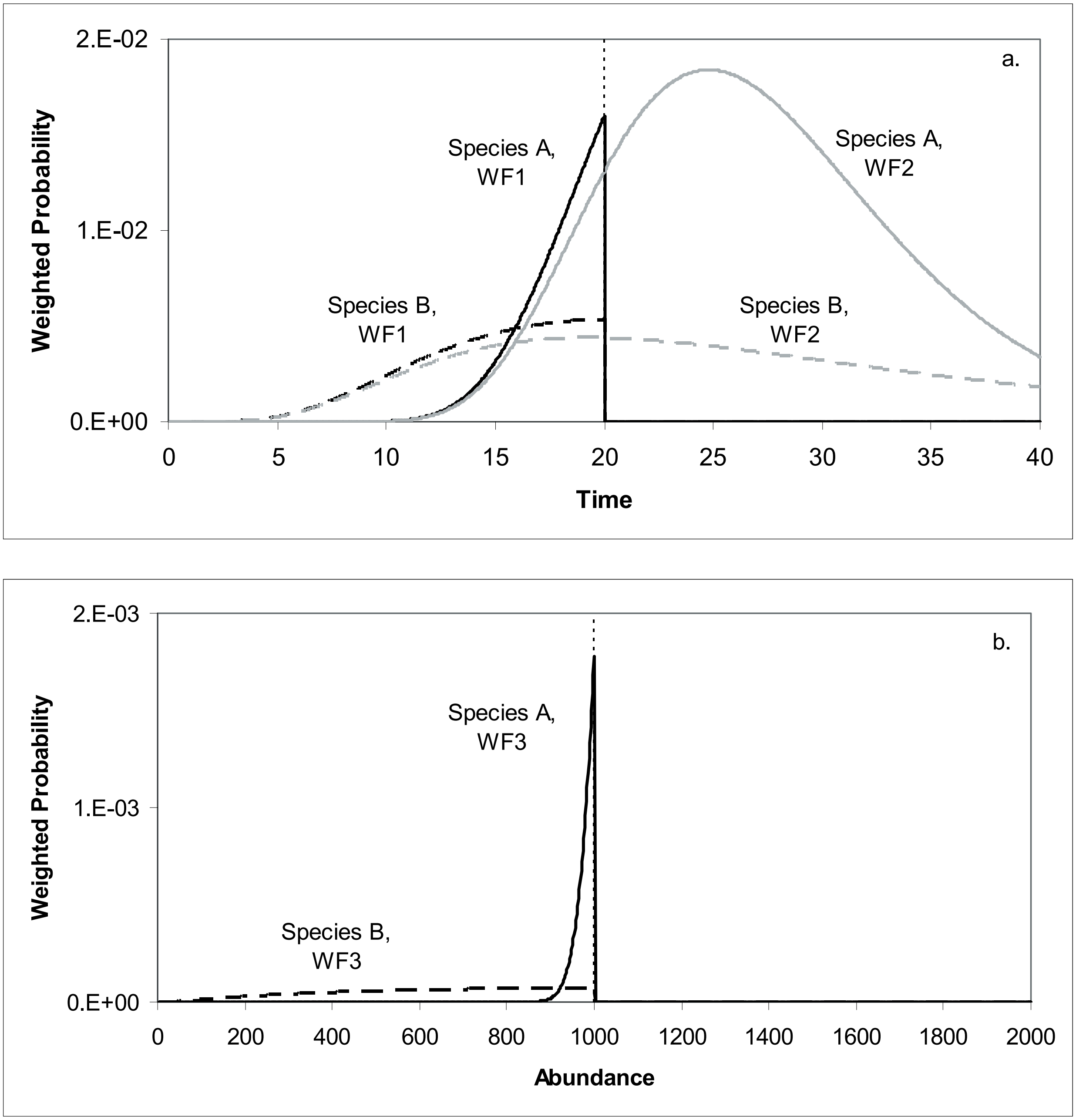


Appendix Figure A3.6. Cumulative probability of abundance. Appendix Figure 6a (top): Cumulative abundance probability for Species A. Figure A3.6b (bottom): Cumulative abundance probability for Species B. Note that the scales on the axes differ between the two panels.





Appendix Figure A3.7. Weighting functions. Appendix Figure A3.7a (top): Two weighting functions with extinction time as the variable. Figure A3.7b (bottom): A weighting function with abundance as the variable.



Appendix Figure A3.8. Definitions of extinction risk for three approaches. For each approach and species, the area under the curve is the extinction risk. Appendix Figure A3.8a (top): Traditional (black solid; black dashed) and comprehensive (grey solid, gray dashed) approaches. Appendix Figure A3.8b (bottom): Threshold approach.

APPENDIX 4: PROBABILITY OF EXTINCTION THRESHOLD EXAMPLE – MODEL DETAILS

Example proxy population viability analyses (PVA) are developed for two pinniped cases (equal adult sex ratio and skewed adult sex ratio) and two cetacean cases (early age at first reproduction (age 4) and late age at first reproduction (age 10)). For this exercise, no research has been done on what values are appropriate for specific PVAs. For example, one important parameter in PVAs is how much birth and death rates vary through time (also called environmental stochasticity). When these proxies are really developed by Life History Groups, one of their tasks will be to review the literature to arrive at appropriate proxy values. Here, we simply chose values to illustrate the process of developing proxies that correspond to over-arching risk levels. The base case standard deviation was ½ the maximum growth rate (R_{max}). This nicely correlates variability in growth rate to growth rate, which is an expected evolutionary correlation and has been observed in marine mammals. For example, sea lions have high growth rates (for marine mammals) but are strongly affected by environmental events such as El Niño. Animals with a slow growth rates, like bottlenosed dolphins, do not have such volatile growth rates as poor years often are reflected in lowered birth rates, which has a smaller effect on growth rates than actual loss of juveniles and adults.

This example concentrates on two proxy criteria likely to be sufficient for all marine mammals: 1) abundance and trends, and 2) abundance alone. In the case of abundance and trends, much of the data most crucial to a PVA are already being used. We pursued two different strategies with respect to time scale: 1) a population reduction of at least x% within 100 years, and 2) a population reduction of at least y% within 15 years. Both involve the same simulation. The 15 year time horizon allows for a broad range of possible rates of decline. This time scale also has the appeal of appearing more at a management time scale of decades rather than a century. Further, it is likely that data will be available for most or all of such a period, which lessens the need to extrapolate. The time period of 15 years was chosen as a time when a moderately high rate of decline (5%/year) would be detectable 90% of the time (assuming $\alpha = 0.10$) with a common level of precision for marine mammals (CV = 0.25).

In the case of abundance only, a different approach was considered where the question is whether the population was "stable" at that abundance ($r = 0$), would it go below the extinction threshold with a >1% chance within 100 years. Or put another way, is it likely that a population that is now at a small size will by chance events alone drift down to dangerous levels? Thus, the threat is the property of being a small population that experiences normal fluctuations in abundance because of environmental stochasticity.

Pinnipeds

The proxy PVAs use a very simple model for an "average" pinniped species with a maximum growth rate of 12%/year. Two major social structures are separated because these reproductive strategies are likely to significantly affect risk: equal adult sex ratios and unequal (female biased) adult sex ratios. The unequal sex ratio is likely to be at higher risk for the same number of adults because the effective number of adults will be much lower from a genetics perspective. Thus, decompensatory responses (through lack of suitable mates, inbreeding depression, etc.) are likely to begin at higher abundances for this mating strategy. For this exercise the social structure differences are represented through different extinction thresholds: 50 adults for equal sex ratios and 75 adults for unequal sex ratios.

Cetaceans

The proxy cetacean PVAs use two different growth rates to depict different life history strategies with maximum growth rates of 2%/year and 8%/year. Several of the large baleen whales have demonstrated growth rates near 8%/year with inter-birth-intervals around 2 years. Several of the social odontocetes, however, have inter-birth-intervals of over 4 years with substantially more care and investment in offspring. We use the same extinction thresholds for equal adult sex ratios as for pinnipeds (50 adults).

Model form: For simplicity, the population dynamics model uses only abundance (N), growth rate (λ or ln(r), and variance in growth rate (s^2). Although the details may be useful in discussing plusses and minuses of the effort to create proxies, a more realistic model would no doubt be much different.

Results

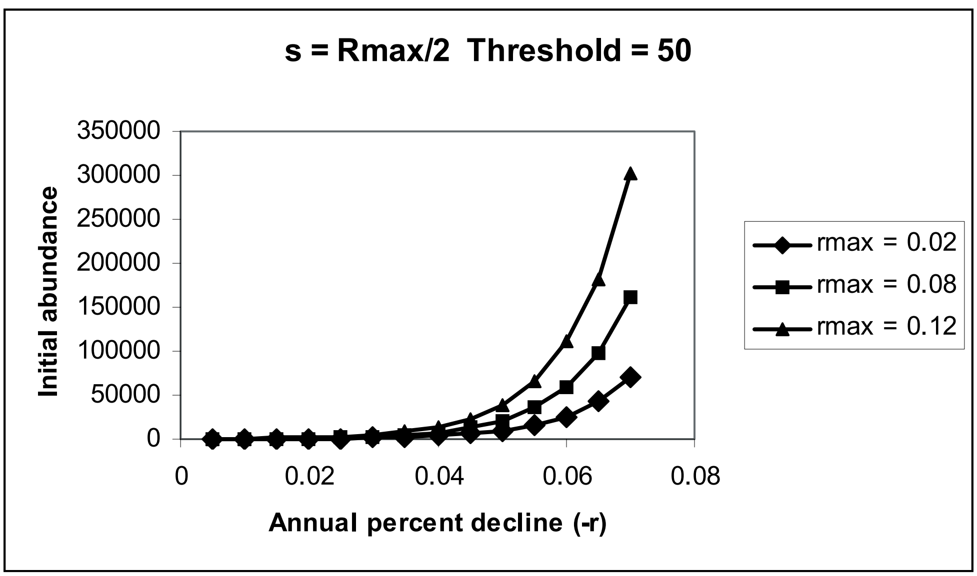
Criterion A—abundance and trend

The IUCN criteria use single numerical values that are supposed to apply across all taxa to categorize species. For example, the criterion using abundance and trend data (IUCN 2001: Criterion C) uses only one number for abundance (2,500) and one for percent decline in 5 years (20%). However, when one considers a 1% chance of reaching an extinction threshold of 50 in 100 years it is clear that there are many ways to arrive at this state of risk. For example, a small population of 100 could decline at a very slow rate or a large population of 100,000 could decline at a very fast rate. Both are actually at the same risk of extinction and should receive the same treatment. The IUCN criteria have sacrificed the principle of equal treatment for equal risk for rules-of-thumb that may only be applicable to a relatively small number of taxa.

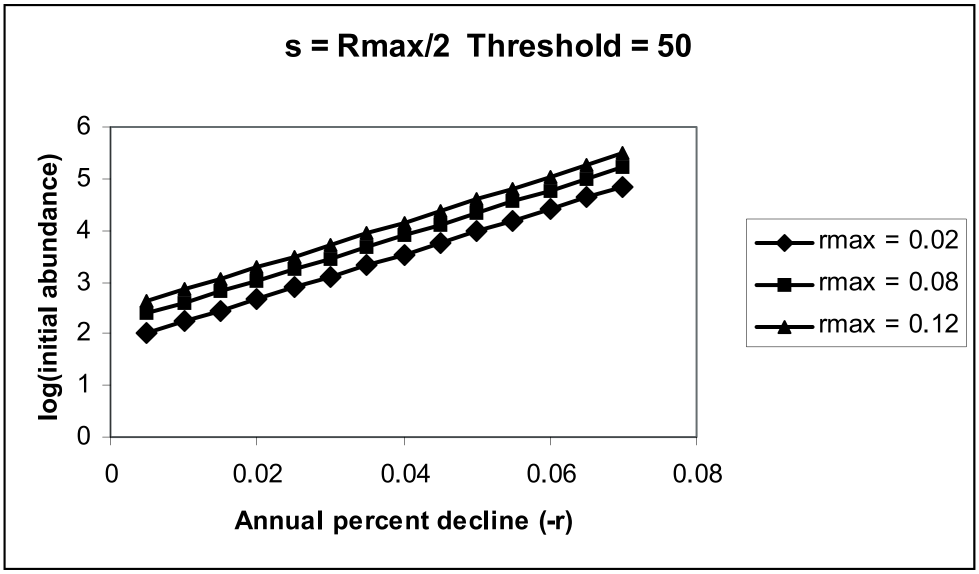
The most direct output of the abundance and trends model shows trend against initial abundance that results in a 1% chance of reaching the extinction threshold in 100 years (Appendix Figures A4.1a and b). Using this figure we can re-examine IUCN Criterion C. A 20% reduction in 5 years equates to an annual rate of decline of 4.5%. This would result in listing the two cetacean examples (R_{max} = 0.02, initial abundance (N_0) = 5,077; R_{max} = 0.08, N_0 = 13,265) as endangered but not listing the pinniped examples (R_{max} = 0.12, N_0 = 23,855). The reason that the pinnipeds require a higher abundance for the same average rate of decline is because the variance in their growth rate is much higher and the observed growth rate of a population is actually the geometric mean growth rate not the arithmetic mean growth rate. The IUCN criterion currently in use is therefore inconsistent in the way species with high variance in growth rate are managed relative to species with low variance in growth rates.

The large influence of the magnitude in variability in growth rates is shown in Appendix Figures A4.2a and b. For these simple simulations, the only way that R_{max} is employed is as a direct correlate of s (which is usually set at $R_{max}/2$). This simple relationship results in a perfect relationship between initial abundance and s (Appendix Figure A4.2b). As indicated earlier, these nice relationships will not be maintained once more realistic PVAs are created. Nevertheless, the basic form of these relationships will be maintained; that is, for a given average r the variance in r will have a large impact on extinction probabilities. Thus, if the principle of equal treatment for equal risk is upheld, then criteria must take variance in r into account.

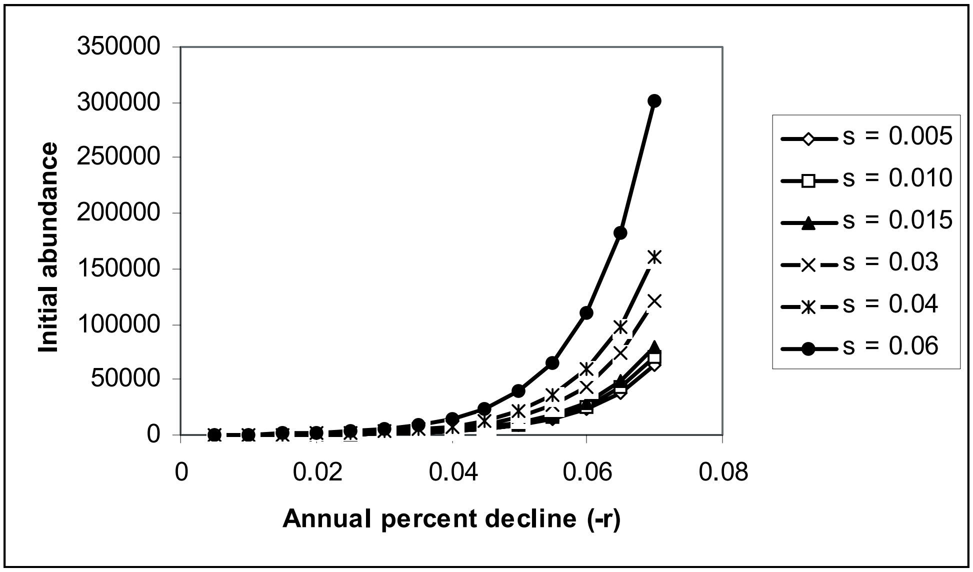
The abundance/reduction criteria contours for a 15 year period (Appendix Figures A4.3a and b) show the same properties as the annual rate of decline figures; i.e., different initial abundances coupled with different rates of decline lead to the same overall risk in 100 years. An alternative criterion is the percent reduction in 100 years (Figure A4.4). Unfortunately, most of the simulated populations in this analysis have declined by >99% in 100 years, so there is little resolution at this time scale for rates of decline greater than 2% per year. If a simple proxy was used, for example something like >90% in 100 years, then the result would be a large over-protection error.



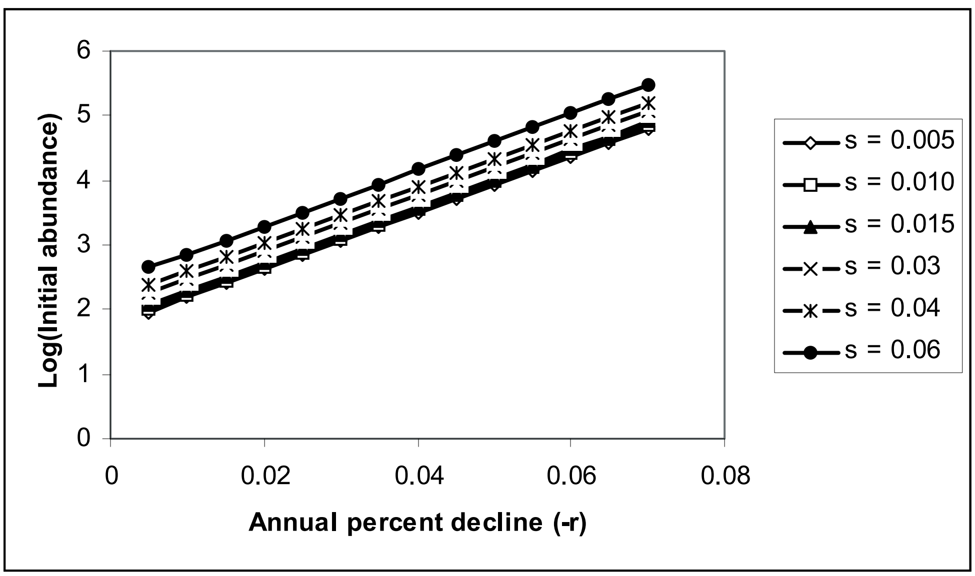
Appendix Figure A4.1a. Initial abundances together with annual percent decline that all result in a 1% chance of reaching an extinction threshold of 50 adults in 100 years assuming that the standard deviation (s) is $R_{max}/2$. Any combination of Initial abundance and annual percent decline falling below the appropriate R_{max} line would qualify for an Endangered listing.



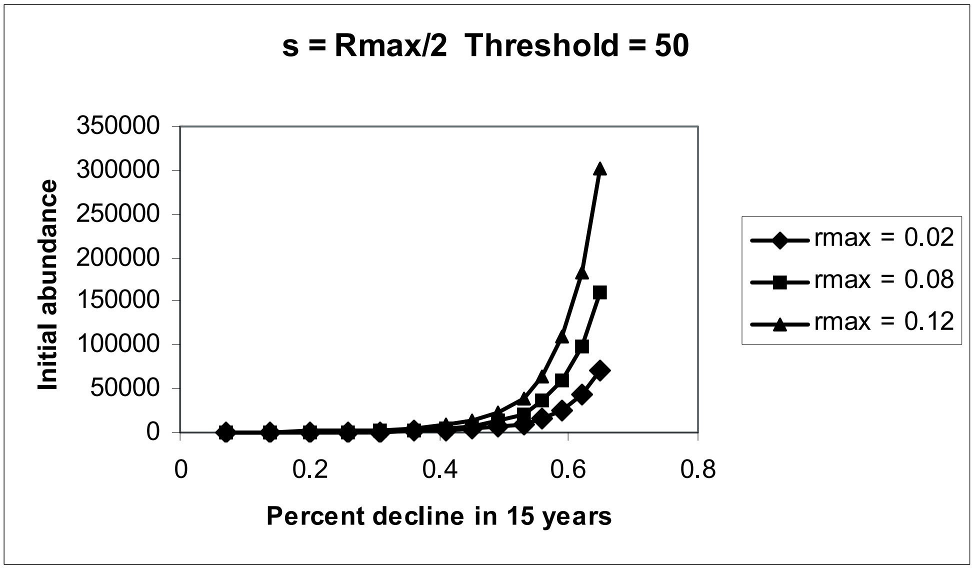
Appendix Figure A4.1b. The same data presented in Figure A4.1a with abundance on a log scale.



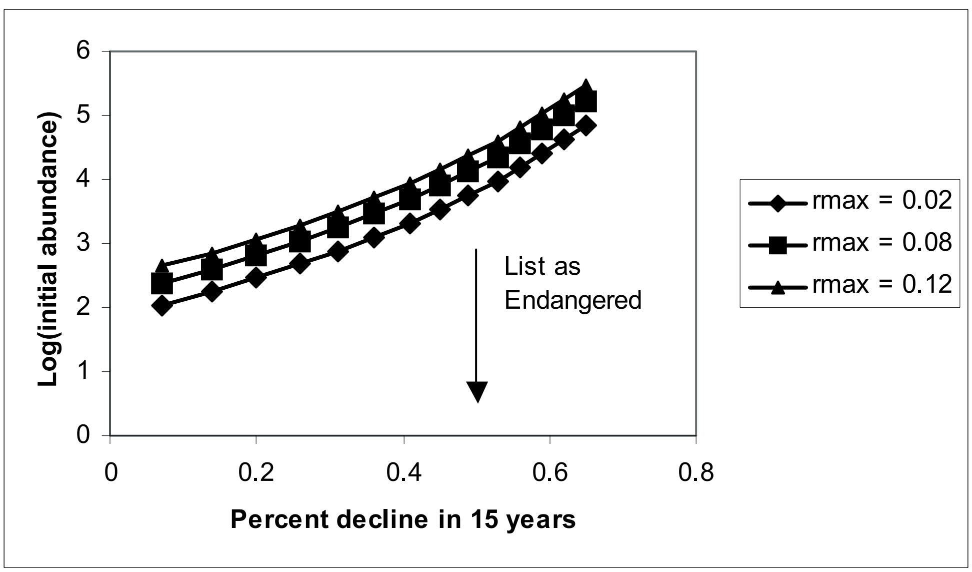
Appendix Figure A4.2a. Initial abundance as a function of annual percent decline for different values of the standard deviation in r (s).



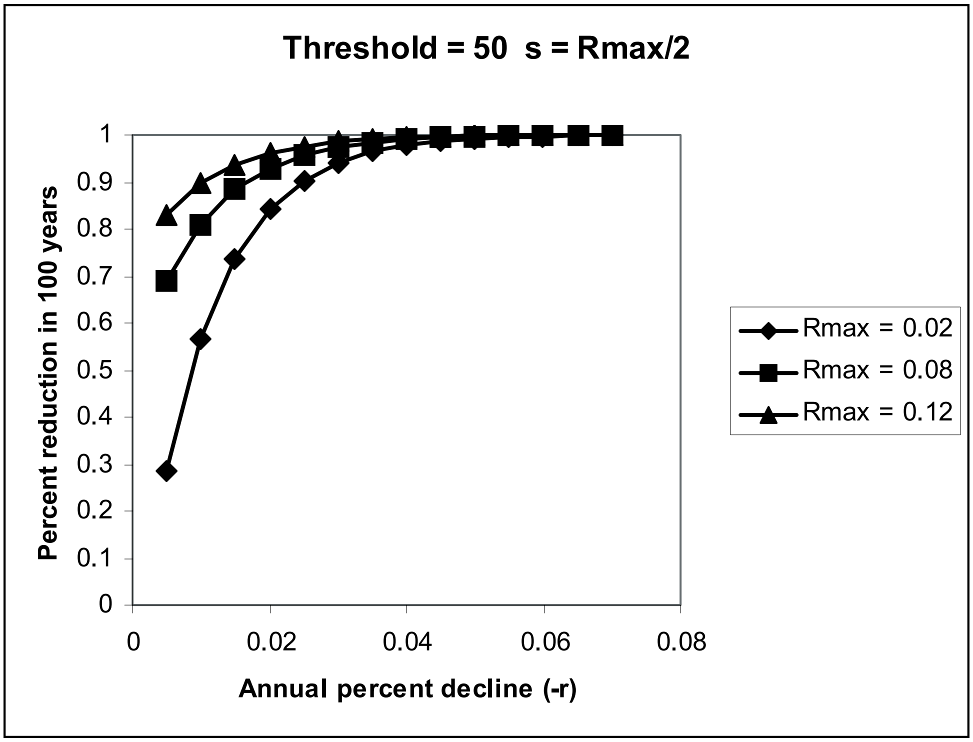
Appendix Figure A4.2b. As in Appendix Figure A4.2a with the log of initial abundance.



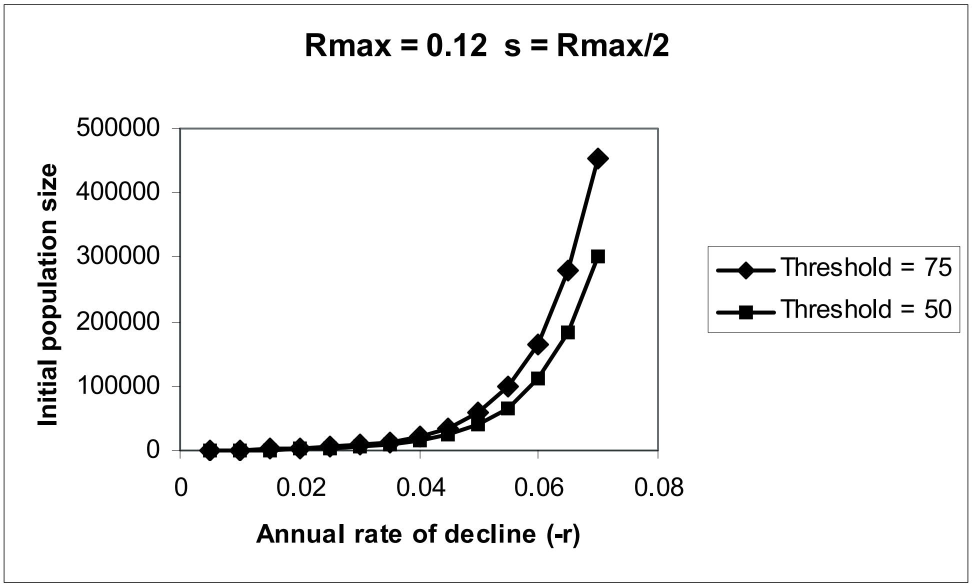
Appendix Figure A4.3a. Given the same threshold for extinction (50) and standard deviation in growth rate ($s = R_{max}/2$), different EN criteria contours are shown for different R_{max} values. An EN listing would be warranted for cases that fell below the appropriate contour.



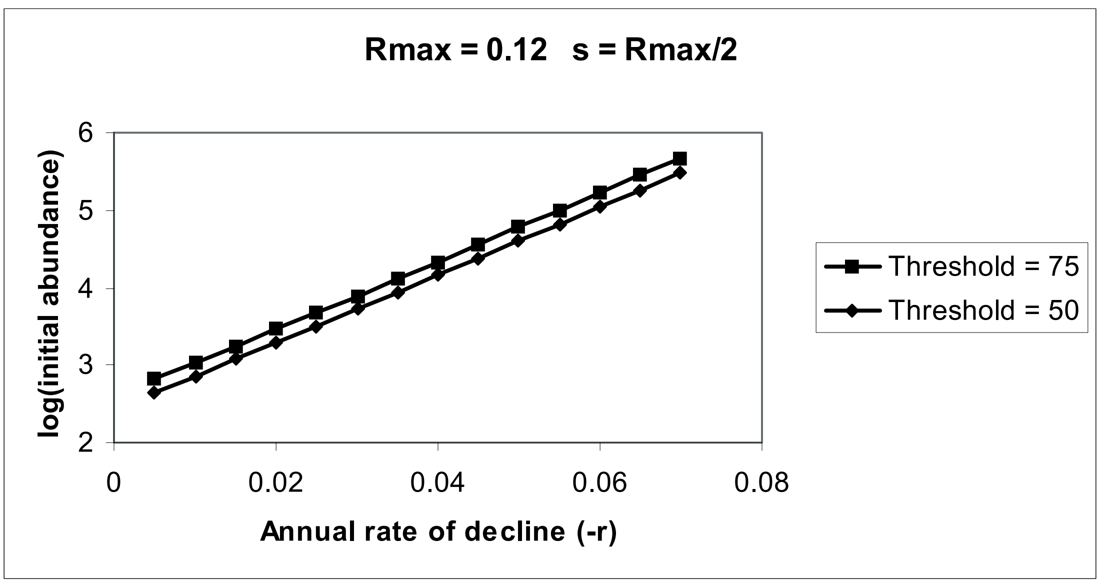
Appendix Figure A4.3b. As in Appendix Figure A4.2a with the log of initial abundance.



Appendix Figure A4.4. The percent reduction in 100 years for different annual rates of decline and for different levels of variability in growth rate.



Appendix Figure A4.5a. Comparison of different extinction thresholds for an even adult sex ratio mating system (threshold = 50) and a skewed adult sex ration (threshold 70).



Appendix Figure A4.5b. As in Appendix Figure A4.3a with the log of initial abundance.

APPENDIX 5: BASELINE EXTINCTION RATES

The fossil record indicates that throughout the history of life on earth, species experienced periods of relative stability punctuated by episodes of mass extinction (Wilson 1987, Raup 1992, cited in Primack 1998). Current extinction rates due to anthropogenic factors are 100 to 1,000 times faster than the background extinction rates predicted by the fossil record (Primack 1998).

The extinction rates of birds and mammals are better understood than those of any other taxa. However, even these are uncertain for a number of reasons (Diamond 1988, Whitten et al. 1987, cited in Primack 1998). Since 1600, 2.1% of mammals, 1.3% of birds, and 0.1% of fishes are known to have become extinct (Reid and Miller 1989, Smith et al*.* 1993, Heywood 1995, cited in Primack 1998). (The percentage reported for fishes is representative of freshwater and marine fishes in North America and Hawaii). The majority of extinctions for all three classes occurred in the last 150 years (Smith et al. 1993).

For birds and mammals, Primack (1998) reports the following rates of extinction for the periods indicated: 1 species per decade (1600-1700); 1 species per year (1850-1950); 4 species per year (1986-1990). The current observed rate of extinction for birds and mammals translates to 1% of all known species per century according to Primack (1998), who also notes that at a natural rate, only 1 species per century would become extinct.

Extinction rates of fishes are difficult to determine for many reasons including insufficient research and ambiguous taxonomy (Harrison and Stiassny 1999; Abell 2002, cited in Stiassny 2002). Harrison and Stiassny (1999) categorized extinctions of freshwater fishes based upon how much evidence of extinction was available. Only 3 species could be termed "unequivocally extinct"; however, as many as 245 freshwater fish species may be extinct (Harrison and Stiassny 1999). Of the 245 presumed extinctions, 54% occurred in Lake Victoria. Nelson (1994) estimated that 245 species equates to 2% of the world's freshwater fishes, including anadromous species. Bruton (1995) suggested that 245 may be a gross underestimate of the number of extinct freshwater fishes. According to Harrison and Stiassny (1999), excluding the extreme case of Lake Victoria, 93% of freshwater fish extinctions occurred in the past 50 years, with approximately equal numbers of events in tropical and temperate areas. They quantified the proportions of these extinctions for different global regions as follows: 5% (Africa and Madagascar); 30% (Asia); 18% (Central America); 17% (North America); 16% (Europe); 10% (South America).

Extinction rates for marine fishes do not seem to be well-understood or quantified. In terms of natural extinction, it is possible that marine fish species are highly resilient and that they endure for prolonged periods until a major tectonic upheaval or climate change wipes out large numbers of species at once (L. Kaufman, pers. comm. Boston University, Boston, MA, 1994). Roberts and Hawkins (1999) suggest that human impact on marine fishes, however, has been extensive and that the risk of extinction is now increasing more quickly than many suspect. Hutchings (2001) indicates that the previously accepted premise that marine fishes are significantly less susceptible to extinction than most other taxa lacks evidence. However, the empirical record actually provides such evidence.

While the Quantitative Working Group was not able to develop and evaluate the merits of specific ways to apply information on baseline extinction rates in the Endangered Species Act (ESA) listing process, one possible approach was identified. Here, a threshold for an ESA listing of endangered would be established where the estimated rate of extinction for a given taxa was greater than some value that was acceptable to the public. For example, if 90% of U.S. citizens polled believed that a value of 100 times the baseline rate of extinction for that taxa was reasonable, then such an approach could be implemented using 100 as the threshold value. However, this approach would require reliable estimates of baseline extinction rate for a variety of taxa, as well as information on current rates of extinction for a specified period of time for the same taxa. How to incorporate uncertainty in such an analysis would be challenging. Further, for those taxa where information on baseline extinction rates do not exist, alternate approaches have to developed. If the Steering Group is interested in further developing this concept, a separate working group should be established.

APPENDIX 6: USE OF PERFORMANCE TESTING TO ADOPT A SET OF DECISION METRICS

One simple approach to adopting a set of decision metrics begins by noting that the outcomes of the simulations follow a 3-by-3 factorial design defined by three possible *true* listing decisions (endangered, threatened, and no listing) and three possible *estimated* listing decisions (endangered, threatened, and no listing). Of course, the ideal result would be for the estimated listing decision to equal the true listing decision in all cases. In practice, however, this will not be possible. What is needed, then, is a means by which the cost of an erroneous listing can be measured. Presumably, NOAA Fisheries leadership would view a one-category-removed error (e.g., determining that a truly endangered species is only threatened) as less costly than a two-categories-removed error (e.g., determining that a truly endangered species does not merit listing at all). Also, it is possible that NOAA Fisheries leadership might value errors in the positive direction (e.g., determining that a truly endangered species is only threatened) differently from errors in the negative direction (e.g., determining that a truly threatened species is actually endangered). Let the matrix of possible costs be designated *L*. An example *L* matrix is shown below. In this example, a correct listing decision (unformatted text) is assigned a cost of zero. Each error in the negative direction (*italics*) is assigned a cost equal to 75% of the corresponding error in the positive direction (**bold**), and a two-categories-removed error within a color group is assigned a cost twice that of a one-category removed error within the same color group.

True decision	Estimated decision Endangered	Threatened	No listing
Endangered	0	**100**	**200**
Threatened	*75*	0	**100**
No listing	*150*	*75*	0

To use this approach, of course, NOAA Fisheries leadership would have to replace the hypothetical amounts in the off-diagonal elements of the above table with amounts that adequately characterize the agency's values. Once this table has been finalized, however, no further management inputs are required (i.e., all of the remaining inputs to the procedure are supplied by the design and results of the simulations).

The frequency of true listing decisions can be expressed as a vector of proportions summing to unity. Let this vector be designated *P*. An example *P* vector is shown below:

True decision	Proportion
Endangered	40%
Threatened	20%
No listing	40%

The frequency of estimated listing decisions can be expressed as a matrix, each row of which represents the distribution of outcomes for a given category of true listing decisions. Let this matrix be designated *Q*. An example *Q* matrix is shown below:

True decision	Estimated decision Endangered	Threatened	No listing
Endangered	50%	**40%**	**10%**
Threatened	*40%*	50%	**10%**
No listing	*10%*	*40%*	50%

The element-by-element product of the *P* vector and the *Q* matrix gives another matrix showing how the estimated listing decisions are distributed across all possible outcomes. Let this matrix be designated *R*. The *R* matrix corresponding to the example *P* vector and *Q* matrix is shown below:

True decision	Estimated decision Endangered	Threatened	No listing
Endangered	20%	**16%**	**4%**
Threatened	*8%*	10%	**2%**
No listing	*4%*	*16%*	20%

The element-by-element product of the *R* and *L* matrices gives another matrix showing the expected loss associated with each possible outcome. The resulting matrix corresponding to the example *R* and *L* matrices is shown below:

True decision	Estimated decision Endangered	Threatened	No listing
Endangered	0	**16**	**8**
Threatened	*6*	0	**2**
No listing	*6*	*12*	0

The sum of the elements in the above matrix gives the expected loss associated with the proposed set of decision metrics. The expected loss in this example is 50. As noted above, this value only has meaning when compared to the expected loss value for other decision metrics.

To avoid undue attention to the absolute value of the expected loss, it would be convenient to normalize this quantity. A convenient normalization begins by forming a vector from the maximum values in the rows of the *L* matrix. Let this vector be designated *Lmax*. The *Lmax* vector corresponding to the example *L* matrix is shown below:

True decision	Maximum loss
Endangered	200
Threatened	100
No listing	150

The element-by-element product of the *P* and *Lmax* vectors is shown below:

True decision	$P \times Lmax$
Endangered	80
Threatened	20
No listing	60

The elements in the above vector sum to 160, representing the maximum possible expected loss for any set of decision metrics in this example. The expected loss (50) divided by the maximum possible expected loss (160) gives the relative expected loss, which in this example has a value of 0.3125. The set of decision metrics that yields the lowest relative expected loss is the set that should be adopted.