The IUCN Criteria Review:
Report of the Range Areas
and Uncertainty Workshop

Report of a workshop held at the Quarantine Station Conference Centre, Manly, Sydney, Australia on 3-5 May 1999, part of the review of the IUCN Criteria for listing threatened species.

Edited by: Georgina M. Mace
Institute of Zoology
Zoological Society of London
Regent's Park
London NW1 4RY, UK
Tel: +44 171 449 6692
Fax: +44 171 483 2237
Email: Georgina.Mace@ioz.ac.uk

Additional copies of this report may be obtained by contacting the SSC Red List Officer. Contact Craig Hilton-Taylor <craigh@wcmc.org.uk>. The SSC has a website with information about all activities see http://iucn.org/themes/ssc.

TABLE OF CONTENTS

1.

Introduction

C. Hilton-Taylor opened the meeting by welcoming all the participants and thanking David Keith and Berin Mackenzie for making the arrangements for the workshop. The meeting was sponsored by the New South Wales Parks and Wildlife Service and the New South Wales Scientific Committee. C. Hilton-Taylor also thanked these bodies for their support.

The participants were invited to introduce themselves and outline their interests and connections to the subject under discussion.

G. Mace then outlined the background to this workshop in the context of IUCN's Red List Programme. She explained that the Red List Programme was evolving from a somewhat subjective and *ad hoc* system that had been in place in the 1980s to a more systematic and transparent programme planned for 2000 to 2003. This evolution was important for the new uses that were being required from IUCN's threatened species lists for overall conservation planning and management. An important component of the new programme was the introduction in 1994 of new categories and criteria for threatened species listing. This new system was applied to around 20,000 animal species in preparation for the *1996 IUCN Red List of Threatened Animals*, which included over 5000 threatened species. Subsequently, in 1998, the same system was applied in a wide-ranging analysis of trees leading to the listing of 10,000 threatened species of trees in the *World List of Threatened Trees*. Meanwhile a number of other plant groups were evaluating species using the new system.

In 1996 the Species Survival Commission of the IUCN had committed to reviewing the new system of categories and criteria, and appointed a Criteria Review Working Group as part of the Red List Programme Committee. The application of the new system to a wide variety of species led to a number of queries and concerns being raised, and the review process was planned to take account of all these as well as seeking input from the membership of IUCN and SSC.

The review process was summarised in a separate document (The Criteria Review Process). After seeking written input from the membership of IUCN and SSC a scoping workshop was held in 1998 to discuss all the topics raised and to then focus the review. This led to the recommendation that there were four major areas that needed further investigation, which should each be the focus of separate workshops involving experts on these topics. The first of these, a workshop on marine issues, was held in January 1999. The present meeting, concerning range area and uncertainty issues, would be followed by a workshop relating to criterion A in July 1999. A final review meeting would be held to make final decisions on the recommendations from all three of these meetings. The purpose of the current workshop therefore was to review the issues that had been raised relating to range area measures and the handling of uncertainty in the criteria, and make recommendations to the review group.

Finally, G. Mace reminded the group that it was important to discuss these topics and the Red List categories and criteria in the right context. The Red List is a global list of threatened species and cannot be expected to give the high precision classifications that might be expected within taxonomic groups or within regions. The aim of the Red List is to provide overall information on the status of the world's species and to guide actions made at national and local levels. Most of the world's species are very poorly known, and for such a system to be effective it needs to be able to deliver classifications from limited information. A simple, inclusive but effective system is needed, rather than a detailed, specific and predictive one. However, it was of real concern to IUCN if the current criteria were often inaccurate, or were biased towards the inclusion or exclusion of certain kinds of species. The main issues concerning the discussion of range areas at this meeting were comments on possible biases. Secondly, because the system is broad and often based on data of poor quality, there was a general concern that some measure of uncertainty should accompany the listings. This would be the focus of the discussion on day 2.

G. Mace then reviewed the issues to be addressed under the range areas discussion.

2. Range area thresholds and problems of scale

2.1 Review of issues to be addressed

a) Issue 1: Regional applications

Although the IUCN categories and criteria were designed for global species assessments, many people are interested in applying the system at regional, national or local scales. This has led to a number of problems. IUCN is addressing this through a separate process, and a Regional Applications Working Group has been established under the chairmanship of Ulf Gardenfors. Nevertheless, since many issues are common to the theme of this workshop, it was decided to include a short presentation on this topic.

b) Issue 2: Size of area thresholds

It has been suggested that maintaining a single set of thresholds for area in criterion B is undesirable as the values are inappropriate for some taxa, and that variable thresholds, relevant to different taxa, should be adopted. There are two aspects to this problem. In the first place, all the criteria are not intended to be appropriate for all cases. The range area criterion is formulated to allow listing of species that are threatened as a result of the small areas that they can use, and not for species where population size or ongoing declines are the major issue. As such, criterion B should apply particularly to small-bodied species living at high or variable population densities. However, there have been suggestions that for flowering plants and invertebrates, the current threshold values are too high. In contrast, for marine species the threshold values have been reported to be too small. The group was asked to review whether there should be different thresholds for different taxa, and if not, whether there should be modifications to the current values for range area thresholds.

c) Issue 3: Issues of scale

Issues 1 and 2 are also related to problems of scale. The scale at which a range is measured affects the area estimated, since grids are scale dependent. This becomes a problem for species measured at fine scale, which will always have an extremely small area of occupancy. Some assessors are finding the thresholds too inclusive or not inclusive enough because interpreting AOO thresholds measured at a biologically appropriate scale (as indicated by the definition in IUCN (1994) page 12) means that the thresholds are inappropriate (for example, too large for the scale grid). Variable range thresholds are one solution to problems of scale. Alternatively, more guidance on the choice of scale of measurement should be provided. In any case this issue needs to be explained more clearly for assessors to understand the implications of alternative methods used to derive values for the assessment.

d) Issue 4: Highly skewed metapopulation structure

Criterion C excludes listing of species where the majority of the mature individuals are found in one population, but some members are found outside the main population. Related problems are that in criterion B, individuals could be unevenly distributed between locations, and that the differing status of population sub-units is not considered. This means that species close to the current thresholds do not qualify although their status could be as threatened as others which do qualify. Potential solutions to this might be as follows: (1) The term sub-population could be substituted by ‘location’ with greater flexibility applied to the latter term (i.e. occurrence of a few individuals does not qualify as ‘locations’). (2) Criterion C2b could be changed from “all individuals in one sub-population” to "more than 90% of the mature individuals in one sub-population". The group was asked to consider the importance of this problem and these, and other, solutions.

e) Issue 5: Parity between criteria B and C

There is currently a lack of parity between criteria B and C in a number of aspects. For example, B may be more exclusive than C, B2 is more flexible and environmental fluctuations are not included in C. The different number of sub-criteria in B and C is also confusing. There are no quantitative thresholds for fragment size in B1, whereas there are for C2. There is no equivalent of criterion C1 in B, and C1 is rather rarely used. The group was asked to consider whether these differences are necessary given the different biological phenomena being identified by criteria B and C? If not, could the criteria be simplified by introducing more common structures, e.g. eliminate C1, use B sub-criteria for C, add a sub-criterion C2c encompassing fluctuation, include B1 within the basal part of criterion B, etc.

f) Issue 6: Extremely restricted populations

There are a number of problems in the listing of extremely restricted populations, which relate to criterion D, and especially to criterion D2. (1) Under the current criteria there is no way of listing very small but stable populations, measured only by range, at any level above VU, without invoking declines from unpredictable threat processes. Whilst it would be possible to create a D2 equivalent for EN, and maybe even CR, the levels would need to be validated to ensure extinction risks are appropriate. Such a solution must exclude poorly known taxa or regions. This could alternatively be addressed by reworking the definition of B1+2c. (2) For a variety of small-bodied species, especially some plants and invertebrates, the current D2 is perceived to be too inclusive, since many species distributed and measured at small scales qualify for threatened status despite the fact that assessors perceive them to be much less threatened than related forms. The suggestion has been made that the area threshold for D2 should be reduced. (3) For oceanic species the current thresholds in D2 are thought to be too exclusive, and the suggestion has been made that the thresholds should be increased. (4) There is a more general view that criterion D leads to the listing of many species that are naturally rare and that this is therefore a very different emphasis from the other criteria that identify species that are thrreatened and declining. The group was asked to consider whether there was a case for a reformulation of criterion D, especially D2, or the extension of D2 into categories higher than Vulnerable?

g) Issue 7: Unusual ranges

All areas in the Red List criteria are in km^2. There have been a number of requests to account for the three-dimensional ranges of marine fishes and birds and for the linear (or even 3D) ranges of riverine species. The group was asked to consider whether these species were really distinct and, if so, what was the best way to deal with their ranges in the criteria.

h) Issue 8: Definition of location

The terms location and sub-population are used in various places throughout the criteria. A review is needed of where each (or both) is appropriately used. A clearer definition of the term location is also needed (IUCN 1994, p12, note 11). Some suggestions already made are to: (1) Replace 'event' in definition with 'extinction process' or equivalent, (2) Include text distinguishing sub-populations and locations. The group was asked to make suggestions for clarifying or generalising the term location, and distinguishing between the use of location and sub-population.

i) Issue 9: Linking range declines to population declines

In criteria A, B and C, there are a number of references to declines in area of occupancy, extent of occurrence and quality of habitat. In criterion A it is clear that population reductions hould be deduced from these declines, but no guidance is given on how to do this. Elsewhere the intention is to reflect declining habitat quality in the area measures, but more guidance on this is needed.

2.2 General discussion

It was agreed that issues 2 and 3 were closely related. U. Gardenfors pointed out that there was a contradiction between 'appropriate biological scale' (for measuring area of occupancy) and the implicit grid scales in the cut-off values, especially for Critically Endangered. D. Keith agreed and said that km^2 was meaningless for species with very small ranges and that apart from this problem with assessment there was a difficulty with measuring very small ranges. He suggested that there were three solutions to this problem: (1) to live with it, (2) to adopt different thresholds scaled somehow for different species, (3) to use a ratio between the scale at which an area is measured and the areas of occupancy. R. Akcakaya commented that there were various alternative ways to scale the threshold values; among vertebrates they may be related to body size, but perhaps not among plants or invertebrates. A meaningful measure might relate to dispersal distances or among animals, the day range length. D. Given said that there are different kinds of risk and that some should scale to the biology of the species, but some, especially the stochastic events would not. M. Burgman said that the criteria should be inclusive and that he was comfortable with the thresholds erring on being too inclusive. R. Akcakaya said that body size is known for almost all species being assessed and suggested that this could be used to scale the threshold values. Y. Sadovy pointed out that some large-bodied species depended upon very small areas for survival, and R. Akcakaya replied that the body size correlates would be precautionary for these. H. Marsh asked how the thresholds would be scaled with body size. N. Collar agreed that people would find this very difficult, and emphasised the importance of keeping the methods simple. H. Marsh said that there were conflicting needs here since the capacity of experts to use the criteria was important. If they are too complicated they get used inappropriately by people on both sides of a debate.

W. Ponder raised the issue of over-listing under criterion D2. He stated that this leads to many species of smaller-bodied organisms being over-listed. This was especially a problem when there was a general lack of data. R. Akcakaya suggested that criterion B should be used to help to solve the criterion D problem. K. Gaston observed that in fact we do not actually know about the relationship between area and risk of extinction, so precise solutions are not possible. D. Keith suggested that this also related to the use of the term 'location'. A. Burbidge stated that the adequacy of surveys was important here. U. Gardenfors suggested that one solution would be to allow assessors not to list a species even if it qualified. G. Mace said that this was not allowed under the current rules. M. Burgman said that a sensible system would allow assessors to choose to not list species. G. Mace replied that this would have many implications for the interpretation of lists and would also cause problems where people disagreed about a species' status. R. Akcakaya agreed that this was not appropriate in a system based on standard rules. G. Mace pointed out that the wording of the D2 criterion in the rules does allow assessors flexibility, in that it defines the circumstances under which species should be listed and presents the cut-offs more as guidelines that simple quantitative limits (by using the qualifier 'typically'). In practice assessors have chosen to apply the D2 criterion literally but the sense of the text is in fact more flexible.

U. Gardenfors asked whether there was any intention to alter criterion E. T. Yahara said that for listing of Japanese plants a combined A, C and E criterion had been developed which was useful and which would help to solve these taxon-specific problems.

There followed a discussion about the relationship between criteria B and C. D. Keith asked why more sub-criteria were required in B than C. K. Gaston explained that the range-based measures were always weaker indicators of extinction risk than the population size measures and that there was a lot of empirical evidence that fluctuations led to increased local extinction. G. Mace said that fluctuations are implicit in the C criterion because the measure is of the smallest number of mature individuals in a fluctuating population. T. Auld said that the differences in the structure of criteria B and C were difficult to understand and D. Keith asked why the number of locations changed with threat level in criterion B, but the sub-population numbers in B did not. It was agreed to form a working group to discuss the details.

2.3 Discussion of issues and conclusions from working groups

It was agreed that three working groups would consider issues 2 to 9. Issue 1 would be covered by a plenary presentation by U.Gardenfors.

The working groups were formed as follows:

Group A, considering issues 2 and 3.
A. Burbidge, M. Burgman, A. Butler, K. Gaston, D. Given, D. Keith, J. Paxton.

Group B, considering issues 4, 5, 7 and 8.
R. Akcakaya, T. Auld, N. Collar, S. Hopper, B. Mackenzie, Y. Sadovy, G. Mace.

Group C, considering issues 6 and 9.
P. de Lange, T. Gerrodette, P. de Lange, H. Marsh, W. Ponder, J. Wang, T. Yahara, C. Hilton-Taylor

a) Issue 1: Regional applications

U. Gardenfors presented the draft Guidelines for Regional Assessments using the IUCN categories and criteria. This document is obtainable from U. Gardenfors or C. Hilton-Taylor. The discussion centred around the incorporation of actual and potential immigrants in the listing of regional populations. R. Akcakaya commented that more consideration should be given to source populations and less to sinks. J. Wang agreed that immigration might be increased through invasion from a less-well adapted population. P. de Lange commented that in new Zealand they distinguish between real local populations, local populations that are peripheral and vagrants which have less conservation implications. D. Given suggested that a better way of dealing with the unknown status of species outside the region was needed. D. Hopper said that an understanding of the breeding system would help to determine whether invading populations were actually rescuing the species or not. U. Gardenfors agreed to look at this issue again.

b) Issues 2 and 3: Size of area thresholds and issues of scale.

The working group concluded that there is an implicit scale in the criteria that is not appropriate for all species, and that the present threshold values are not (and cannot be) universally applicable, and so will introduce bias. They considered that range areas are use din different contexts:

- To reflect the possibility of catastrophic loss. Here the bias is not so importartant.
- As a surrogate measure for population size. Here the bias is important.

The group had considered how to scale the range area measures for different species. The two most obvious ways were with body size, but here there are many exceptions, and with dispersal distance, but here again there are many exceptions and the data are inadequate. The group concluded there is no simple way to scale the area-based thresholds. They considered that the criteria may over-list sessile forms, but in general they consider they are precautionary and there is no obvious solution to the problem.

The group thought there could be a case for larger thresholds in marine habitats because threatening processes may operate on larger spatial scales there than in terrestrial habitats. This needed further investigation and there was no recommendation.

c) Issue 4: Highly skewed metapopulation structure

- The group considered that uneven distribution of individuals across the species' range cannot be incorporated when locations are used, because locations are used for species where population size is unknown.
- However, when population sizes are estimable, it was considered appropriate to make the sub-population criterion more flexible, and to modify criterion C2b to 'at least 95% of mature individuals in one population'

> **RECOMMENDATION 1**: Alter the condition in C2b, to be 'at least 95% of mature individuals in one population'

- The group also considered altering the population number thresholds in C2b with threat level. However, this was decided against because there is no simple relationship between population number and extinction risk.

d) Issue 5: Parity between criteria B and C

The group had spent some time discussing this issue and concluded that the confusing structure was probably more important than the parity between the two. The lack of parity results from the fact that criterion B is based on range area which is always a more indirect measure of extinction risk, whereas criterion C is based on mature individuals which already includes much intraspecific variation in the formulation of the definition. However, there were various topics to address:

- The group discussed criterion C1 and wondered if this was really useful or necessary. It seems that species listed by this criterion would commonly be identified by other criteria and its removal would simplify the structure of criterion C.

> **RECOMMENDATION 2**: Review the use of sub-criterion C1 and consider eliminating it if few (no) species uniquely qualify for listing by this means.

- The group also considered moving fluctuations into a sub-criterion of C. However, since criterion C currently uses a minimum population estimate and this would require changes to the definitions of mature individuals, it was agreed that it was best left as it stands.
- Consideration was also given to making B2 part of the main condition for criterion B so as to have one of two sub-criteria as in the present C2. However, this would eliminate the possibility of qualifying by criteria B1+3, which is currently used for a number of plant species. In general 2 of three criteria is more precautionary and the recommendation was made to leave this as it stands.

e) Issue 6: Extremely restricted populations

The group ha discussed criterion D2 at some length. They concluded that D2 captures species for listing that are very rare and always susceptible to extinction, even though the risk may be very small. Ideally, the spatial scale in D2 should scale with threats to the taxon, and notes are needed to guide assessors on this point. They had discussed a number of other ways to deal with D2 species:

- By bringing area-based threatening processes into criterion E, rather than D, where they can be dealt with more reliably and explicitly.
- By moving D2 listed species into LR (nt).
- By increasing the flexibility to assessors by clarifying the definition. Linking the two sentences in the definition of D2 would make this more explicit.
- By removing the quantitative guidelines in D2 altogether.

> **RECOMMENDATION 3**: Leave D2 in the criteria and do not extend it into Endangered and Critically Endangered. However, different wording is needed to improve the sensible application of D2. First, by altering the definition of D2 to make the link between the two halves of the definition and allowing greater flexibility. Second, some extra advice and examples should be provided to encourage assessors to include area-based threats under criterion E. Third, sampling intensity needs to be considered by assessors to avoid the false listing of under-researched species.

f) Issue 7: Linear and 3 dimensional ranges

The groups recommended that the precautionary approach is to simply calculate the two dimensional areas and use this in the assessment against the criteria. In the case of linear ranges this would be by multiplying the length of the habitat by the average or typical width.

Issue 8: Definition of location

The current use of locations in the criteria is focussed around threats. Guidelines and examples are needed to help assessors to use it correctly. The issue was raised of how to deal with the fact that species facing more than one threat can have different location counts. It was agreed that the recommendations should use the value for the most serious threat.

> **RECOMMENDATION 4**: Provide more guidance and examples in the definition of location. When there are several threats that are relevant, assessors should use the number of locations indicated by the most serious threat.

g) Issue 9: Extrapolating loss of habitat and loss of area

The group had discussed a number of issues in the relationship between area, habitat and population, and agreed that this was a complex issue. However, there was agreement that notes needed to be included to emphasise the non-equivalence of different measures.

> **RECOMMENDATION 5:** Notes and examples are needed to illustrate the fact that loss of habitat does not equate to loss of population, and assuming that the two are the same will sometimes be non-precautionary.

3. Dealing with Uncertainty

3.1 Introduction

Despite the fact that the notes accompanying the current criteria recognise the problem of data uncertainty, there is no clear guidance on how to deal with it in either the assessment of species or the interpretation of listings. This is an important problem that constrains the use and interpretation of the Red List criteria and categories, and leads to irresolvable debates over particular issues. New methods and approaches offer a way. Many other problems with the criteria are related to this issue, e.g. the use of Data Deficient, the lack of criteria for 'near threatened', and the assessment of species whose status is known only from one small part of its range. We need to review these new approaches and proposed methods to see where they are generally helpful, how we can usefully integrate them into the practical guidelines for red listing, how to provide a better understanding of uncertainty for users of the red list (decision makers), and where there are still unresolved problems.

3.2 Presentations

Three presentations were made. Mark Burgman outlined the nature of uncertainty, stressing the differences between errors in measurement, ambiguity in terminology and implict vagueness. Even under the best of circumstances some uncertainty remains and may actually be an important part of the system. Resit Akcakaya presented the RAMAS Red List software package, which deals explicitly with measurement errors and allows assessor subjectivity in the listing process through later adjustments to results from choices about dispute tolerance and risk tolerance. Both these presentations are summarised in the paper on 'Uncertainty' included as Appendix A.

Tim Gerodette then gave a short presentation on methods for testing the criteria based on procedures used by the International Whaling Commission to assess the performance of various whaling management plans. Such simulation testing could provide a means of evaluating whether the Red List system is really classifying species as it is intended. Although undertaking this would be a large and complex task, many participants agreed that it would be an interesting and useful thing to do.

A general discussion of the issues raised followed, and two working groups were formed. The first first outlined procedures for evaluating uncertainty in the listing process, and their conclusions are presented below. The second worked on details of the RAMAS program to make it fully compatible with IUCN's procedures. Since the second of these does not form part of the criteria review only the first is reported upon in this document.

3.3 Recommendations on uncertainty

RAMAS Red list proved to be a practical tool for integrating uncertainty into the process of listing and the presentation of categories and criteria.

However, it was agreed that apart from this software which would aid the handling and presentation of data, assessors needed to be made more aware of the nature of the data they were using and the consequences of too rigorous or too casual an approach to maintaining a responsible attitude. Accordingly, a system was outlined that should be used for training assessors and should be included alongside the rules and procedures for the categories. This system is intended to provide guidance towards more ssytematic handling of data and information in the listing process.

> **RECOMMENDATION 6**: More guidance on the handling of uncertainty is needed. The scheme outlined in Figure 1 ('How to deal with uncertainty'), and the text attached as Appendix A, should be included alongside the criteria, their rules and definitions, to encourage a more consistent and knowledgeable handling of uncertainty in the listing of species.

4. Concluding comments

C. Hilton-Taylor thanked all the participants for their contributions and David Keith and Berin Mackenzie for arranging and supporting the meeting.

5. Summary of Recommendations

RECOMMENDATION 1: Alter the condition in C2b, to be 'at least 95% of mature individuals in one population'

RECOMMENDATION 2: Review the use of sub-criterion C1 and consider eliminating it if few (no) species uniquely qualify for listing by this means.

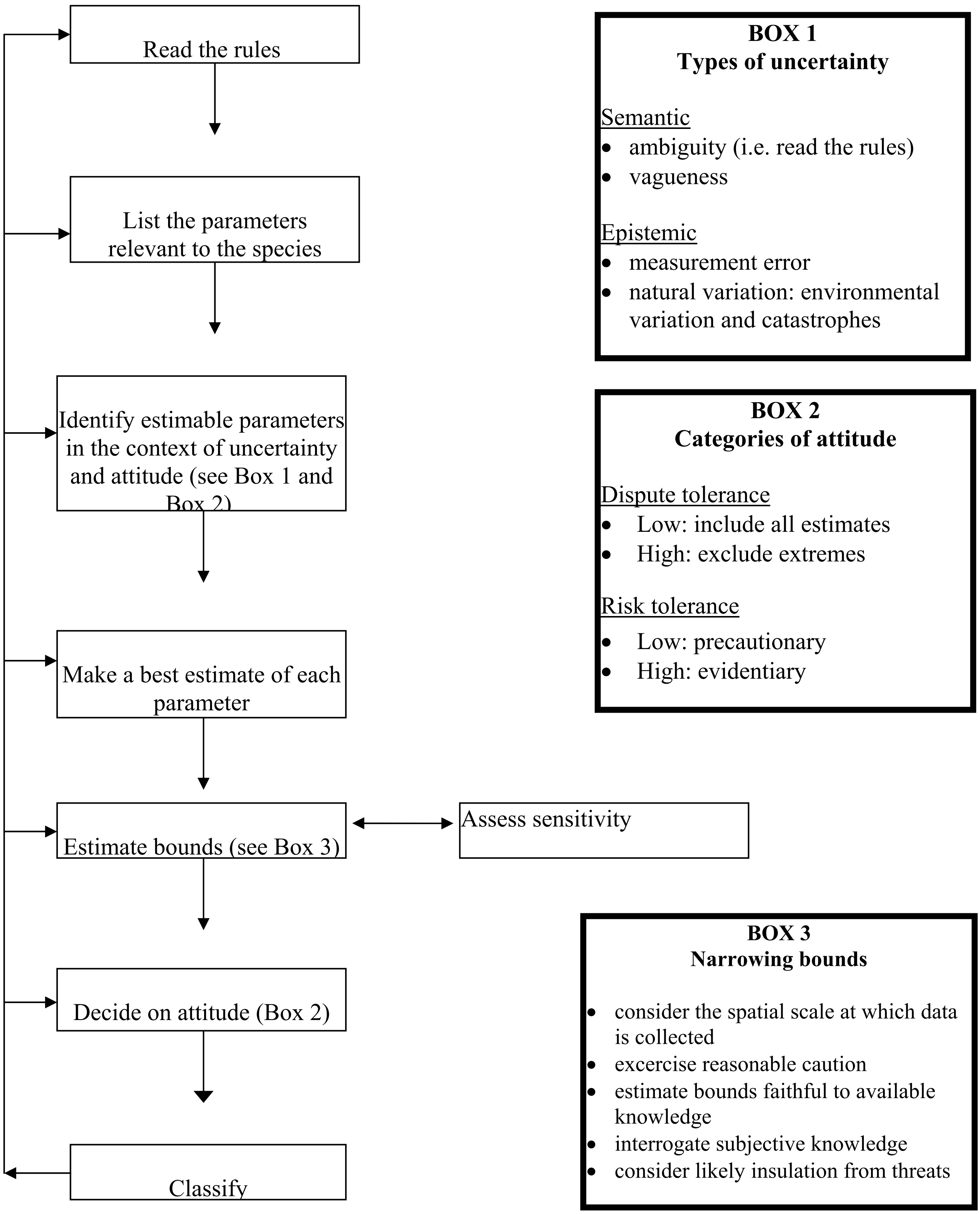
RECOMMENDATION 3: Leave D2 in the criteria and do not extend it into Endangered and Critically Endangered. However, different wording is needed to improve the sensible application of D2. First, by altering the definition of D2 to make the link between the two halves of the definition and allowing greater flexibility. Second, some extra advice and examples should be provided to encourage assessors to include area-based threats under criterion E. Third, sampling intensity needs to be considered by assessors to avoid the false listing of under-researched species.

RECOMMENDATION 4: Provide more guidance and examples in the definition of location. When there are several threats that are relevant, assessors should use the number of locations indicated by the most serious threat.

RECOMMENDATION 5: Notes and examples are needed to illustrate the fact that loss of habitat does not equate to loss of population, and assuming that the two are the same will sometimes be non-precautionary.

RECOMMENDATION 6: More guidance on the handling of uncertainty is needed. The scheme outlined in Figure 1 ('How to deal with uncertainty'), and the text attached as Appendix A, should be included alongside the criteria, their rules and definitions, to encourage a more consistent and knowledgeable handling of uncertainty in the listing of species.

FIGURE 1: A system for handling uncertainty



6. List of Participants

Akcakaya, Resit, Applied Biomathematics, 100 North Country Road, Setauket, NY 11733, USA, EMAIL: resit@ramas.com, TEL: +1 516 751-4350, FAX: +1 516 751-3435

Auld, Tony, Biodiversity Survey and Research Division, New South Wales National Parks and Wildlife Service, PO Box 1967, Hurstville, NSW 2220, Australia, EMAIL: tony.auld@npws.nsw.gov.au, TEL: +61 2 9585-?, FAX: +61 2 9585-66606

Burbidge, Andrew, Director, WA Threatened Species & Communities Unit, Dept Conservation & Land Management, PO Box 51, Wanneroo, WA 6065, Australia, EMAIL: andrewb@calm.wa.gov.au, TEL: +61 8 9405-5128, FAX +61 8 9306-1066

Burgman, Mark, School of Botany, University of Melbourne, Parkville, 3052, Australia, EMAIL: m.burgman@botany.unimelb.edu.au, TEL: ?, FAX: ?

Butler, Alan, CSIRO Marine Research, GPO Box 153, Hobart, 7001, Tasmania, Australia, 7001, EMAIL: alan.butler@marine.csiro.au, TEL: +61 3 6232-5491, FAX: +61 3 6232-5123

Collar, Nigel, Birdlife International, Wellbrook Court, Girton Road, Cambridge, CB3 0NA, United Kingdom, EMAIL: collar@birdlife.org.uk, TEL: +44 1223 277-318, FAX: +44 1223 277-200

De Lange, Peter, Science, Technology & Information Services, Dept of Conservation, Private Bag 68908, Newton, New Zealand, EMAIL: PDelange@doc.govt.nz, TEL: ?, FAX: ?

Gardenfors, Ulf, Threatened Species Unit , ArtDatabanken, Box 7007, Uppsala, S-750 07, Sweden, EMAIL: Ulf.Gardenfors@dha.slu.se, TEL: +46 18 67-2623, FAX: +46 18 67-3480

Gaston, Kevin, Dept. of Animal and Plant Sciences, The University of Sheffield, Sheffield, S10 2TN,United Kingdom, EMAIL: K.J.Gaston@Sheffield.ac.uk, TEL: +44 114 222-0030, FAX: +44 114 222-0002

Gerrodette, Tim, Southwest Fisheries Science Center, 8604 La Jolla Shores Drive, PO Box 271, La Jolla, CA 92038-0271, USA, EMAIL: timg@ucsd.edu, TEL: +1-858-546-7131, FAX: +1 858-546-7003

Given, David, 101 Jeffreys Road, Christchurch 5, New Zealand, EMAIL: givend@ibm.net , TEL/FAX: +64 3 351-6069

Hilton-Taylor, Craig, Red List Programme Officer, IUCN/SSC UK Office, 219c Huntingdon Road, Cambridge, CB3 0DL, United Kingdom, EMAIL: craigh@wcmc.org.uk, TEL: +44 1223 277-966, FAX: +44 1223 277-845

Hopper, Stephen, Chief Executive, Kings Park and Botanic Garden, West Perth, 6005, Australia, EMAIL: postmaster@kpbg.wa.gov.au, TEL: +61 8 9480-3600, FAX: +61 8 9322-5064

Keith, David, Biodiversity Survey and Research Division, New South Wales National Parks and Wildlife Service, PO Box 1967, Hurstville, NSW 2220, Australia, EMAIL: david.keith@npws.nsw.gov.au, FAX: +61 2 9585-66606

Mace, Georgina, Institute of Zoology, Zoological Society of London, Regent's Park, London, United Kingdom, EMAIL: Georgina.Mace@ioz.ac.uk, TEL: +44 171 449-6692, FAX: 44 171 483-2237

Mackenzie, Berin, Vegetation and Dynamics Unit, Biodiversity Survey and Research Division, New South Wales National Parks and Wildlife Service, PO Box 1967, Hurstville, NSW 2220, Australia, EMAIL: berin.mackenzie@npws.nsw.gov.au, TEL: +61 2 9585-6507, FAX: +61 2 9585-66606

Marsh, Helene, Environmental Science Postgraduate Studies, James Cook University, Townsville, Australia, EMAIL: helene.marsh@jcu.edu.au, TEL: +61 7 472-77502, FAX: +61 7 478-15581

Paxton, John, Fish Section, Australian Museum, 6 College Street, Sydney, NSW 2000, Australia, EMAIL: johnp@amsg.austmus.gov.au, TEL: +61 2 9320-6139, FAX: +61 2 9320-6059

Ponder, Winston, Australian Museum, 6 College Street, Sydney, NSW 2000, Australia. EMAIL: winstonp@amsg.austmus.gov.au, TEL: +61 2 9320-6120, FAX: +61 2 9360-4350

Sadovy, Yvonne, Dept. of Ecology & Biodiversity, University of Hong Kong, Pok Fu Lam Road, Hong Kong, China, EMAIL: yjsadovy@hkusua.hku.hk, TEL: ++852 2859-8977, FAX: ++852 2517-7997

Wang, John, Center for Marine Conservation, 1725 DeSales St., N.W., Washington, DC 20036, USA, EMAIL C/O: abrautigam@dccmc.org, TEL: ?, FAX: +1 202 872-0619

Yahara, Tetsukazu, Department of Biology, Kyushu University, Hakozaki 6-10-1, Fukuoka 812-81, Japan, EMAIL: tyahascb@mbox.nc.kyushu-u.ac.jp, TEL: ?, FAX: +81 92 642-2645

7. APPENDIX A

NOTES ON UNCERTAINTY

By R. Akcakaya, S. Ferson & M. Burgman

The criteria should be applied based on the available evidence on taxon numbers, trend and distribution, making due allowance for statistical and other uncertainties. These uncertainties arise from:

- semantic uncertainty
- measurement error
- natural variability.

Natural temporal and spatial variation in population size and distribution affects our ability to report on these parameters with certainty. Natural variation exists because populations vary in space and time in response to environmental variation and the vagaries of demographic processes. The effect of this variation on the application of the criteria is limited, because each parameter refers to a specific time or has a specific spatial scale. In addition, parameters related to fluctuations and fragmentation explicitly consider these types of variability. An important difference of natural variability from the other two types is that it cannot be reduced.

Semantic uncertainty arises from inexact definitions. Imprecise or unclear definitions should be improved to eliminate ambiguity as far as possible. However, even the best attempts to make the definitions exact will not succeed in eliminating semantic uncertainty altogether. In some cases this is simply not possible without the loss of generality. It would be impossible to create a sensible set of rules that would apply to all taxa, and that could be interpreted unambiguously. Consequently, some components of semantic uncertainty are irreducible, and are classified as 'vagueness'. However, that does not mean they should be ignored.

The second source of uncertainty derives from actual measures (factual or epistemic uncertainty). There are two distinct types in this category. **Measurement error** is often the largest source of uncertainty; it arises from the lack of precise information about the variables used in the rules. It is very unlikely to know, for instance, the exact numbers of mature individuals, or the exact rate of decline in the past 10 years. All numerical data are uncertain; qualitative information, such as whether there are extreme fluctuations, may also be uncertain. Measurement error is distinguished from **natural variability** because it may be reduced or eliminated, at least in theory, by acquiring additional data.

Representing uncertainty
One of the simplest ways to represent uncertainty when different kinds of uncertainty are confounded is to specify a best estimate and a range of plausible values. For example, we may represent an adult population size with a best estimate of 90 and bounds of 70 and 120. The ranges might be based on confidence intervals, the opinion of a single expert, or the consensus opinion of a group of experts. In cases when consensus is not possible, it may be necessary to pool different estimates from a number of experts or based on a number of assessment methods. In such cases the "best estimate" itself may be an interval (such as 85 to 95) instead of a point estimate. When data are very uncertain, the range for the best estimate may coincide with the range of plausible values. In this case, the resulting object can be represented as an interval.

Parameters for which there is no information can be represented by large intervals that include all possible values. For example, an unknown percent reduction should be represented as [0,100], and unknown number of mature individuals as [0, 8]. However, it is important to recognise that in many cases such an interval is likely to be overly uncertain. For example, even if no direct data are available for population reduction, it may be known by inference that the reduction was not 100%. If any information is available about the plausible range of a parameter, it should be used instead of "unknown" range.

Some of the data for the IUCN rules are required in the form of yes/no or true/false answers. These include, for example, whether there are extreme fluctuations, and whether all individuals are in a single population. It is not always possible to be certain about responses to these questions, for the same reasons that it is not possible to be certain about quantitative information. Measurement error may make the response to questions about the magnitude of fluctuations or the number of populations unreliable. Different people may hold different opinions. In addition, semantic uncertainty may make the uncertainty irreducible. When the information about these aspects of the status of a taxon are uncertain, the answer can be expressed as a number between 0 (False) and 1 (True), representing the truth value of the statement. In some cases, this number may represent the frequency with which the statement is true. More generally, it represents the degree of plausibility of the statement, or the reliability of the information about that aspect of the population. When assessments made by a number of experts are pooled, the result can be expressed as an interval. Parameters for which there is no information should be represented as [0,1].

Incorporating uncertain data
We suggest that the three rule sets (CR, EN, and VU) are applied to two sets of estimates: (1) a plausible range and (2) a best estimate, for each parameter. Best estimate itself might be a range, but in any case, the best estimate should be always nested in (included in) the plausible range. Both the plausible range and the best estimate should be based on the available empirical information or expert judgement and should be justified.
When the rule sets are applied to intervals and best estimates, the results of the rules sets (criteria) for CR, EN and VU are no longer simple boolean (True or False) results. They depend on the method of propagating uncertainties. One simple method is based on interval arithmetic, which results in 2 intervals of truth values (numbers between 0 and 1) for each rule set.

Reducing to a single category or range of categories
For practical reasons (e.g., reporting results for large numbers of taxa), it is necessary to reduce (condense) the uncertainty represented by these sets of truth values to a single category of threat or a range of categories. This step depends on preferences and attitudes towards risk and uncertainty. It is important the attitudes towards risk and uncertainty that led to a certain classification are explicitly stated.

Attitudes towards risk can be described in terms of ***risk tolerance*** (RT), ranging from low risk tolerance (risk-averse; precautionary), to high risk tolerance (risk-prone; evidentiary). Attitudes towards uncertainty can similarly be described in terms of ***dispute tolerance***. Dispute tolerance (DT) represents how the person interpreting the data feels about uncertainty. If the interpreter wants to be sure of encompassing all possibilities (thus avoiding disagreements and dispute), they would have a low DT. If they wish to be as precise as possible and rely only on people's best judgement, they would have a high DT. A high DT would mean disregarding the more extreme opinions (which might lead to dispute). If the tails of the distribution seem too extreme, but the best estimates are too small to bound uncertainty, the interpreter may choose an intermediate value of DT.

Based on DT, a single interval of truth values can be selected for each rule set ("splitting the difference" between upper and lower intervals for each rule set). In addition, it is necessary to specify a middle-point of this interval, in order to further condense the uncertainty. The location of this middle point within the interval determines another aspect of attitude towards risk, ***burden of proof***, which is closely related to risk tolerance.
Each of the three intervals (for CR, EN, and VU) can be compared to RT to decide, for each category, whether it will be in the final range of threat categories for the taxon.
Similarly, each middle point can be compared to RT to select a single category. The highest threat category for which the middle point is larger than RT is reported as the threat category for the taxon. However, the range of categories should also be reported in addition to this single category (see below).

Reporting results
Depending on uncertainties and attitudes, a taxon may be classified in a single category, and/or a range of categories. For example, a taxon may be classified as "EN (VU-CR)", which means that the most plausible category is Endangered, but because of the uncertainties, the range of plausible categories includes Vulnerable, Endangered, and Critically Endangered.

Where data are so uncertain that the range of categories is "(LR-CR)" and all three rule sets result in the interval [0,1], the category of "Data Deficient" may be assigned. However, it is important to recognise that this category indicates that data are inadequate to determine the degree of threat faced by a taxon, not necessarily that the taxon is poorly known. The category "Data Deficient" is not a threatened category, although it indicates a need to obtain more information on a taxon to determine the appropriate listing.

8. APPENDIX B

TESTING IUCN DECISION RULES WITH SIMULATION TRIALS FOR UNCERTAINTY

Tim Gerrodette and Barbara Taylor

Introduction

In a series of workshops, the IUCN is developing modifications to the decision rules (Red List Criteria) used to classify the threat status of species around the world. Many of the modifications under consideration involve additions, deletions or clarifications of the decision rule set. Such changes should allow the rules to be applied in a consistent and accurate manner. Other modifications address the issue that data quality may vary. The decision rules should not only take account of uncertainty, but deal with it in a precautionary way. Akcakaya et al., Keith et al. and Burgman et al. have proposed various changes to the decision rules that would deal with uncertainty in a better way than the current implementation does.

The following proposal draws on the experience of scientists associated with the International Whaling Commission attempting to come up with a set of rules under which whaling could resume on a sustainable and precautionary basis. The process was difficult and at times quite contentious, but the result was a management scheme that was tested very thoroughly, and that all sides agreed would be able to achieve its management objectives. Cooke (1994) gives a good summary of the process. A similar process could be applied to the IUCN Red List rules to test and refine methods of dealing with uncertainty.

Decision rules and performance goals

The IUCN decision rules (Red List Criteria) are applied to data to arrive at classification of risk. The process may be diagrammed as:

Data -----> Decision rules -----> Classification

The IUCN workshops are currently refining the decision rules part of this scheme: making them more inclusive, removing ambiguities and inconsistencies, etc. If this part of the effort were to be completely successful, the resulting "perfect" rules would, given good data, be applied consistently to all species and would yield accurate classification of risk.

However, even with "perfect" rules, we may not arrive at a correct classification of risk, because data will never be perfect. Any numerical estimate has some uncertainty associated with it. Therefore the decision rules must deal with uncertainty, but this means that the result of applying the decision rules, the risk classification, will also have uncertainty associated with it. Even with perfectly unambiguous and consistent decision rules, there is some chance that a wrong classification decision will be made. Under these circumstances, we want to come up with a set of decision rules whose behaviour under uncertainty is not only consistent but is understood in terms of the kinds of errors that the decision rules will make.

As noted in several of the background papers, there are some "arbritrary" choices that must be made in any system of rules. For example, Burgman et al. proposed using the lower limit of the 90% confidence interval instead of the point estimate for decisions. In the system of Akcakaya et al., there are choices about "dispute tolerance" and "risk tolerance" that may differ among people using the decision rule software. In fact, these choices represent important but implicit policy decisions about what kinds and rates of classification errors will be acceptable. It is important to make such policy decisions explicit.

The key idea behind the proposed simulation testing of the IUCN Red List Criteria is to embody the policy decisions in a separate set of performance goals, and then to test whether the proposed decision rules perform well relative to these goals. Thus, the management objectives are specified in a set of performance goals that the proposed decision rules must meet when they are applied to data. If the proposed decision rules do not perform well, they need to be modified until they do, or until it is concluded that they cannot meet the objectives and some other set of rules needs to be adopted. In this way, the parameters within a decision rule set can be tuned to meet the management objectives. Alternatively, several decision rule sets can be tested against each other to see which performs best.

The process could be diagrammed as

model ---> data ---> decision rules ---> classification ---> evaluation

The simulation model produces the data and can include any or all of the issues that are thought to cause difficulties – for example, estimates with high variance, missing data, biased data, unknown subpopulation structure, etc. Whether the classification of risk is correct or not is evaluated against the known parameters of the model. Because we are now viewing classification as a probabilistic process, we are concerned with the rates of misclassification, a surrogate for the probability of making a wrong decision. For example, consider a population whose true size is 2145 and whose true rate of decline is 22%/year. Under the current Red List rules, such a population should be classified as EN. However, natural variability combined with sampling errors may lead to the population being misclassified as VU, or perhaps CR. If we think of the classification process being repeated many times with variable data in a simulation, we can estimate the rates of misclassification.

The actual testing can be carried out as a "blind" experiment. One group runs the population model to produce the data, and another group receives the data and applies the proposed set of decision rules to them. The results of the classification exercise and the evaluation of the results against the performance goals are carried out in an open forum.

Examples

Burgman et al. proposed that the lower limit of the 90% confidence interval be used instead of the point estimate. Why 90% and not 80 or 95%? It is clear that using a lower number such as 80% would, other things being equal, lead to classification in lower categories of threat, and that we would run the risk of underprotection of species more often. On the other hand, choosing a high value such as 95% would lead to higher categories of threat, and we would run the risk of overprotection. The question is, how do we strike a balance? What, in fact, is an acceptable tradeoff between over- and underprotection errors? This is an example of a policy choice that must be made. The use of a 90% confidence interval already implies a policy choice of acceptable rates of mis-classification errors, but it is not clear what these are. The performance goals in the simulation trials force this choice to be made explicit and then to be tested.

Akcakaya et al. have proposed a well-developed system for dealing with uncertainty based on fuzzy logic that has many advantages over the present IUCN system. Among its advantages are that the choices about dispute tolerance and risk tolerance, which can be subjective choices and may differ among individuals or among management agencies, are explicitly stated. This raises a question: is a system that permits different classifications of species, given the same data, acceptable to IUCN? There are policy choices to be made, to be sure. What is proposed here, as a next step in the development process, is that the policy choices be debated and agreed on. Then the values of dispute and risk tolerance that are compatible with those policy choices can be incorporated into the implemention of the decision rules. In essence, such a system would be one way of making the precautionary approach, which at present we are exhorted to use but are not sure just how, an integral part of the IUCN classification system.

Robustness

Estimates may not only be imprecise (have high variance) - they may also be biased. Data may be missing. Ideally, we would like to have a set of decision rules that are robust to various kinds of errors. Robustness can be tested by simulating biased data, for example. In the case of the International Whaling Commission, this was an important part of the process. Competing decision rule sets were challenged with "bad" data, within what was thought to be a realistic range of "bad." The sensitivity of the decision rule sets to biased or missing data often varied in surprising ways.

Beyond classification to management performance

If the results of risk classifications result in management actions that have effects on populations, and which may therefore change risk classification in the future, it would be possible to model the management performance of a set of decision rules in a more dynamic sense. Such a system might be diagrammed as

```
model   --->  data --->  decision rules   --->  classification   --->  evaluation
    \                                                 /
     \                                               /
      \   <------   management actions   <------/
```

In this case, the performance goals could be, for example, recovery to a certain population size or delisting of the species within a specified period of time.

Summary

Decision rules such as the Red List Criteria contain implicit policy decisions about what rates of over- and underclassification are acceptable. It is important to know what these rates are. Testing the performance of the Red List Criteria can result in rules which are more likely to achieve the desired management objectives, and which can be robust to certain data deficiencies. In addition, the process requires a debate and agreement on explicit policy goals, which is highly beneficial in itself.

References

Akcakaya, HR, S Ferson, MA Burgman, DA Keith, GM Mace, and CR Todd. (in press) Making consistent IUCN classifications under uncertainty. *Conservation Biology.*

Burgman, MA, DA Keith, FJ Rohlf, and CR Todd. Probabilistic classification rules for setting conservation priorities. Working document at this meeting.

Cooke, J. 1994. The management of whaling. *Aquatic Mammals* 20: 129-135.

Keith, D, T Auld, M Ooi and B MacKenzie. Sensitivity analyses of decision rules in IUCN Red List criteria using Australian plants. Working document at this meeting.