

A framework for selecting a suite of indicators for fisheries management

Jake C. Rice and Marie-Joëlle Rochet

Rice, J. C., and Rochet, M-J. 2005. A framework for selecting a suite of indicators for fisheries management. — ICES Journal of Marine Science, 62: 516–527.

We develop a framework for the objective selection of a suite of indicators for use in fisheries management. The framework encompasses eight steps, and provides guidance on pitfalls to be avoided at each step. Step 1 identifies user groups and their needs, featuring the setting of operational objectives, and Step 2 identifies a corresponding list of candidate indicators. Step 3 assigns weights to nine screening criteria for the candidate indicators: concreteness, theoretical basis, public awareness, cost, measurement, historic data, sensitivity, responsiveness, and specificity. Step 4 scores the indicators against the criteria, and Step 5 summarizes the results. Steps 3–5 offer technical aspects on which guidance is provided, including scoring standards for criteria and a generalized method for applying the standards when scoring individual indicators. Multi-criterion summarization methods are recommended for most applications. Steps 6 and 7 are concerned with deciding how many indicators are needed, and making the final selection of complementary suites of indicators. Ordinarily, these steps are done interactively with the users of the indicators, thus providing guidance on process rather than technical approach. Step 8 is the clear presentation to all users of the information contained. The discussion also includes the special case in which indicators are used in formal decision rules.

Crown Copyright © 2005 Published by Elsevier Ltd on behalf of International Council for the Exploration of the Sea. All rights reserved.

Keywords: criteria, ecosystem, framework, guidelines, indicators, objectives, scoring.

Received 1 April 2004; accepted 27 October 2004.

J. C. Rice: Canadian Science Advisory Secretariat, Department of Fisheries and Oceans, 200 Kent Street, Stn 12S015, Ottawa, Ontario, Canada K1A 0E6. M-J. Rochet: Laboratoire de Mathématiques Appliquées à l'Exploitation des Ressources Halieutiques et Aquacoles, IFREMER, B.P. 21105, 44311 Nantes Cedex 03, France. Correspondence to J. C. Rice: tel: +1 613 9900288; fax: +1 613 9540807; e-mail: ricej@dfo-mpo.gc.ca.

Introduction

Many policy and management bodies with an interest in aquatic or marine systems have endorsed indicator-based approaches to management (OECD, 1998; FAO, 2002; World Bank, 2002). In all cases, the agencies note that ecosystems are so complex and unpredictable that suites of indicators are needed to give an adequate picture of their state. In fact, it is often noted that suites of indicators are needed for each of the major dimensions of sustainability: ecological, social, economic, and institutional (Charles, 2001; FAO, 2003). Indicators now have a prominent and legitimate role in monitoring, assessing, and understanding ecosystem status, impacts of human activities, and effectiveness of management measures in achieving objectives, and they have a growing role in rule-based decision-making. Given all these roles, the suites of indicators intended to fulfil them must be chosen wisely.

For evaluations of ecosystem effects of fishing, marine ecosystems have so many properties of concern and so few

proven general state measures that there is generally no shortage of proposals for indicators (e.g. CSAS, 2001; ICES, 2001; Link *et al.*, 2001). The task we undertake here is to outline the steps necessary to select wisely from the long lists of diverse, potential indicators.

Because each indicator implies monitoring, evaluation, and reporting costs, redundant indicators, at least, should be avoided. Both the capacity for meaningful dialogue and the processing ability of rule-based decision-making systems become saturated when overloaded with information from too many indicators (FAO, 2002, 2003). Most seriously, with even modest numbers of indicators, “current values” of different indicators are likely to support arguments for incompatible management actions. Therefore, indicators may simply become a new battleground for partisan arguments, with adversaries selecting the indicators whose values happen to support the decision they desire. For example, for single-species fisheries management, ICES advises largely within the comparatively simple context of annual estimates of spawning-stock biomass (SSB) and

fishing mortality (F). Fisheries commissions commonly have argued against implementing ICES advice to reduce exploitation when F is above its precautionary reference point, if SSB happens also to be larger than its reference point. They argue that, under those circumstances, the excessive F imposes no immediate conservation issue, and that they will have time to reduce F when SSB really requires such action (ICES, 2003).

Clearly, to be cost-effective and to provide clear management guidance, suites of indicators should be kept as small as possible while still fulfilling the needs of all users. The challenge is to identify the suite that best meets the needs in each particular application. Marine ecosystems differ in availability of historical data, monitoring capacity, prosecution of fisheries, other human uses, and governance system, as well as in their ecological properties. All these factors may affect the utility of a specific indicator (Belfiore, 2003; Olsen, 2003), making it obvious that no single suite of indicators is universally the best.

The framework presented is designed to be a guide for practice, and therefore consists of a series of steps and specific tasks to be performed at each step. In practice, governance processes often ensure that indicators are selected in dynamic, interactive exercises; rigid, stepwise algorithms are unlikely to be followed. Hence, the framework has to be flexible in its application. Whatever process is followed, however, the issues described in each step *must* be addressed to select the final suite, and for some of these steps, the order matters (e.g. criteria must be weighted before indicators are scored).

Step 1 – determine user needs

To determine the needs of the users involved in management or governance, it is, of course, necessary to identify who they are. Needs of both managers and stakeholders will be affected by the types of decisions required and the objectives being pursued. Both legislated and cultural governance considerations influence which aspects of the fishery (catch or effort, quota, gear, spatial, or seasonal restrictions) are amenable to regulation, and this may influence the practicality of different indicators (FAO, 2002, 2003).

Whether the indicators are intended to merely inform discussion or to support decision-making directly, the management objectives need to be clearly specified. Some jurisdictions are attempting to do this explicitly (Bergen Declaration, 2002; EC, 2003), in which case the operational objectives can be taken directly from the policy documents. Often, however, objectives either do not exist, or are so general and vague that they provide little guidance for selecting appropriate indicators. In those cases, management bodies must formulate operational objectives first. It is efficient to involve those participating in the selection process in the process of formulating operational objectives as well, to ensure that the final suite of indicators matches

the concerns behind the objectives, even when their wording reflects compromises among differing points of view.

At this initial stage, the major threats to achieving the objectives should be identified – the pressures in a driver-pressure-state-impact-response framework (e.g. OECD, 1993; Bowen and Riley, 2003). When fishing is placed in an integrated management framework with other human activities (Belfiore, 2003; FAO, 2003), it is even more important to specify the major avenues by which each of the activities may threaten achievement of objectives, because the indicators must inform managers about the effects of multiple uses. Indicators of effects of fishing need to be robust to other anthropogenic effects (or the latter need to be understood well), if they are to provide a sound basis for managing fisheries. Information on threats will be important when evaluating the sensitivity, specificity, and responsiveness of candidate indicators.

Universally applicable algorithms for identifying participants, objectives, and threats do not exist. However, general approaches for identifying stakeholders and developing consensus-based objectives provide useful guidance (Smith *et al.*, 1999; Walker *et al.*, 2002; FAO, 2003).

Step 2 – develop a list of candidate indicators

The next key consideration is that candidate indicators truly measure ecosystem status relative to the objectives. Knowledge of the ecosystem, characteristics of the fisheries, and societal values must all be considered. Where clear, system-specific operational management objectives have been set, this step can be as straightforward as listing reasonable ways to measure the property reflected in each of these objectives. Even in this simple case, substantial technical knowledge may be required to develop an initial, comprehensive list of candidate indicators. When objectives are only defined conceptually or developed without adequate technical expertise and full stakeholder representation, the process requires care and patience, because the whole array of potential ecosystem effects of fishing must be considered. Examples include status of non-target species (Bellail *et al.*, 2003), size structure of the fish community (Jennings *et al.*, 1999; Bellail *et al.*, 2003), the central node of a wasp-waisted foodweb (Rice, 1995; Cury *et al.*, 2003), and habitat features (NRC, 1994). Likewise, candidate fishery indicators such as target species, gears, spatial and temporal distribution, quantities and types of discards, and even levels of participation, should be considered (Garcia, 1996; Garcia and Staples, 2000). Candidate social and economic indicators may be no more straightforward to list comprehensively, particularly when policy goals in those respects are not well articulated, or in conflict (Bowen and Riley, 2003; Rice, 2003). Where other uses may affect opportunities available to fisheries, or

expected yields, the list might have to include indicators of the status of the activities (Gottret and White, 2001; Belfiore, 2003; Talaue-McManus *et al.*, 2003). Social scientists, economists, and community leaders may be required to contribute, if the inventory is to be sufficiently complete to allow the selection of a final suite that provides a basis for informed discussion and management support.

Step 3 – determine screening criteria

Published lists of criteria on which indicators should be evaluated (e.g. UNCSD, 2001; ICES, 2002; EEA, 2003) are generally similar. Table 1 lists a selection of nine that cover the concepts behind those proposed by all expert groups, although some agencies may list subsidiary considerations as full criteria, reflecting their particular priorities.

All nine criteria should always be considered, but they are not equally important in every case. Moreover, even in individual applications, different participants in the governance process are likely to value the importance of criteria differently. However, to keep the screening process objective, the relative importance of the nine criteria should be established *before* the screening is done.

Although complex weighting and scoring algorithms have been developed for specific situations (MSC, 2004), weighting the criteria on a refined scale would usually give a false sense of precision to an exercise generally lacking a quantitative basis. Moreover, the final steps in the framework are anyway sufficiently consultative to diffuse any great precision of inputs early in the process. Results of some comparative experiments (Rochet and Rice, 2005) suggest that ranking criteria according to three classifications (high \approx essential; moderate \approx useful; minor \approx inconsequential)

Table 1. Relative importance (Minor, Moderate, High) that three different user groups are expected to attach to the nine criteria used in screening candidate indicators (numbers in parenthesis are tentative rankings within each group, although these will deviate on a case-by-case basis; there is no basis for ranking criteria of Minor importance).

Criterion	Technical experts and advisors	Decision-makers and managers	General audience
Concreteness	<i>Minor</i>	<i>Moderate/High</i> : Decisions would be easy to explain to the public, and to relate to other management activities (5/6)	<i>High</i> : Low score means that it would be difficult to relate personal experience to indicator (2)
Theoretical basis	<i>High</i> : Inconsistency with established theory means low confidence, but a solid empirical basis (3/4)	<i>Minor</i> : Management generally based on values and performance, not ecological theory	<i>Minor</i>
Public awareness	<i>Minor</i>	<i>Moderate</i> : Valuable for getting compliance with management plans (5/6)	<i>High</i> : If general knowledge is lacking, a major education programme would be required (1)
Cost	<i>Minor</i> : In general, not their concern	<i>High</i> : Governance systems are budget-conscious (3)	<i>Minor to High</i> : Value for money (4)
Measurement	<i>High</i> : Low or unknown accuracy and precision; often sufficient grounds for rejection (1/2)	<i>Minor</i> : As long as technical advisors and public have confidence	<i>Minor</i> : Unless sampling design is not considered representative of personal experience (scientific survey debate) (5)
Historical data	<i>High</i> : For estimating reference points, and to have confidence in interpretation (2/3)	<i>Minor</i> : As long as technical advisors and public have confidence. May become <i>Moderate</i> to <i>High</i> when management has to function without technical support	<i>Minor to High</i> : Depends on how much context is needed to interpret changes in value
Sensitivity	<i>High</i> : Poor sensitivity may be sufficient reason for rejection (1)	<i>Moderate</i> : To interpret biological and economic importance of changes in value (4)	<i>Moderate</i> : To attach meaning to changes in value (3)
Responsiveness	<i>Moderate</i> (5/6)	<i>High</i> : For those wanting feedback on effectiveness of management plans (1)	<i>Minor</i>
Specificity	<i>Moderate</i> : To disentangle fishing effects from other impacts (5/6)	<i>High</i> : For those wanting to take proper actions to remedy problems in fishery (or other managed activities) (2)	<i>Minor to Moderate</i> : To understand how fishery relates to the “big picture”

generally should suffice. Sorting and ranking should be done interactively and systematically with the client groups involved.

We make a distinction between three major user groups that may be expected to attach differential importance to the nine criteria (Table 1). Technical experts and science advisors would use indicators to measure progress towards achievement of explicit objectives, often supported by the use of reference points (OECD, 1998; FAO, 2002, 2003; ICES, 2002). The criteria of major and moderate importance to this group would presumably be measurement, historical data, theoretical basis, sensitivity, and responsiveness/specificity. When indicators are used in formal decision rules, science advisors are likely to reject those performing poorly on any of these criteria.

Decision-makers and managers use indicators to support decision rules or, less formally, to guide management actions in addressing discrepancies of indicator status relative to an objective. If indicators are to be used in a structured decision-support context, their selection must be guided even more closely by suitable criteria. However, outside a decision-support context, application of the more stringent criteria might exclude those that are more cost-effective. Criteria valued in both rule-based and consultative decision-making include responsiveness, specificity, cost, concreteness/public awareness, and sensitivity. For rule-based decision-making, the indicators must also perform well on historical data, so that meaningful reference points and decision rules can be set.

When indicators are used to inform general audiences about ecosystem status or effects of management, they are mainly concerned with public awareness, concreteness, sensitivity, and cost, and sometimes measurement. However, a differentiation may be needed between situations where the role of the indicator is to inform an aware and engaged public, or to motivate an apathetic public. In the latter case, an explanation of the underlying theory in accessible language may become increasingly important, as will recent deviations from historical values. For specific users such as fishers, it is also important that personal experience can be linked to changes in indicator values.

Step 4 – score indicators against criteria

The scoring process has two components: the evaluation of the information content or quality of each indicator relative to each criterion, and the strength of the evidence by which information content or quality is judged. These “properties” will not necessarily co-vary, possibly resulting in different scores on different properties of a single criterion that subsequently will have to be reconciled.

With regard to scoring of the information itself, a quantitative evaluation may be made for a few properties of a few criteria only. For example, programme audits provide estimates of the cost of periodically obtaining indicator values. In general, however, attempts to provide

fully quantitative estimates of the value of all indicators on each criterion are likely to fail. Moreover, some criteria are multi-dimensional (e.g. bias, variance, accuracy, precision). Calibrating a criterion value for effects in different dimensions is almost certainly impractical, if not impossible. Hence, candidate indicators often have to be scored in the face of complex dimensionality of the criterion, and in the absence of sound quantitative measures of the properties of interest. Under such circumstances, detailed quantitative scores would give a misleading sense of discriminating power among indicators. In practice, an ordinal scoring on a scale of 3–5 ranks for each candidate indicator on each criterion would seem sufficient (e.g. low, fair, moderate, high). If a multi-dimensional criterion is of major importance, one practical option may be to retain the ordinal scores of the candidate indicators on the key dimensions of the criterion, and to deal with the added complexity in Step 5. The strength of the evidence supporting each evaluation of information quality is likewise rarely amenable to a fully quantitative scoring. Table 2 proposes a ranking of the inherent reliability of different information sources. Such rankings are straightforward to apply and generally adequate for the task. As long as the relative position of candidate indicators is carried forward with regard to their strength of evidence, subsequent steps can be performed with objectivity and rigour.

Existing experimental and analytical approaches permit direct testing of the effectiveness of indicators in supporting decision-making. Piet and Rice (2004) explored tools such as signal-detection theory (Helstrom, 1968) as a means of testing the performance of indicators of fishing effects based on responsiveness, sensitivity, and specificity. However, the interpretation of the performance error rates obtained required external information about costs that users would assign to different types of management errors (e.g. unnecessary TAC reductions vs. permitting overfishing). This suggests that even quantified error rates might not be comparable across criteria, and hence only useful for ranking within criteria (e.g. high “miss” rates for one indicator may arise from a lack of sensitivity; high “false alarm” rates for another from a lack of specificity; which type of error is more serious depends on many factors, including the uses and the objectives being supported by the indicators). Nonetheless, to ensure that sound indicators are chosen from among the candidates, retrospective analysis of their performance in supporting decision-making (ICES, 2003; Piet and Rice, 2004) should be fed directly into the evaluation process.

Step 5 – summarize scoring results

For the final evaluation, two matrices will be available: one with the weights assigned to the nine criteria for each user group, and one with scores of each candidate indicator on each criterion. Entries of the second matrix should contain

Table 2. For each of the screening criteria, the constituent considerations (sub-criteria) in conducting the scoring (H, high; F, fair; M, moderate; L, low) for an indicator (IND) and the methods by which the evaluation could be conducted. Stars on items labelled H and L indicate that, if the consideration (or method of evaluation) is relevant, scoring high there is of high importance, and scoring low is a nearly fatal flaw, respectively. Methods of evaluation are presented in decreasing order of confidence in the results (SI, conclusive published experimental research using Strong Inference; MP, Multiple independent Publications providing consistent findings; FS, Formal designed Surveys; MM, Multiple independent Models producing consistent results; IC, Interdisciplinary Consensus of weight of evidence; TJ, research Team professional Judgement).

Criteria and sub-criteria	Method of evaluation
Concreteness	
<ul style="list-style-type: none"> Concrete property of physical/biological world (H), or abstract concept (L)? 	FS; IC; TJ
<ul style="list-style-type: none"> Units measurable in the real world (H), or arbitrary scaling factor (L)? 	IC; TJ
<ul style="list-style-type: none"> Direct observations (H), or interpretation through model (L)? 	IC; TJ
Theoretical basis (number of competing theories to allow contrast is important)	
<ul style="list-style-type: none"> (i) Not contested among professionals (H); (ii) basis credible, but debated – can account for patterns in many data sets (H–F, depending on how other models fit the same data); (iii) credible, but competing theories have adherents and empirical support is mixed (M); (iv) adherents, but key components untested or not generally accepted (M–L) 	MP**, SI*, MM; IC; TJ
<ul style="list-style-type: none"> If IND derived from empirical observations: (i) concepts readily reconciled with established theory (H); (ii) concepts not inconsistent with, but not accounted for by, ecological theory (M); (iii) concepts difficult to reconcile with ecological theory (L) 	SI**, MP; MM; IC; TJ
<ul style="list-style-type: none"> Theory allows calculation of reference point associated with serious harm (M)* 	MP; MM; IC, TJ
Public awareness	
<ul style="list-style-type: none"> Is it a property with a high (H) or low (L) public awareness outside the use as an IND? 	FS*; IC; TJ
<ul style="list-style-type: none"> Does public understanding correspond well (H) or poorly (L) with technical meaning of IND? 	FS; IC; TJ
<ul style="list-style-type: none"> If awareness high, is public likely to demand action that is: (i) proportional to IND value as determined by experts (H); (ii) disproportionately severe (M); (iii) largely indifferent (L) 	FS; IC; TJ
<ul style="list-style-type: none"> Does the nature of what constitutes “serious harm” (used to define a reference point) depend on values that are widely shared (H) or vary widely across interest groups (L)? 	FS; IC; TJ
<ul style="list-style-type: none"> Internationally binding agreements, national or regional legislation require that a specific IND be reported at regular intervals (H), to agreements/legislation require environmental status reporting, but IND not specified (M) to no such requirements (L) 	IC; TJ (when IND not specified in legislation)
Cost	
<ul style="list-style-type: none"> Uses measurement tools that are widely available and inexpensive to use (H), to needs new, costly, dedicated, and complex instrumentation (L) 	IC; TJ
Measurement	
<ul style="list-style-type: none"> Can variance and bias of IND be estimated? Yes (H); No (L) 	MP; MM; IC; TJ
<ul style="list-style-type: none"> If variance can be estimated, is variance low (H) to high (L) 	MP; IC; TJ
<ul style="list-style-type: none"> If bias can be estimated, is bias low (H) to high (L)? 	MP; IC; TJ
<ul style="list-style-type: none"> If IND biased, is direction usually towards overestimating risk (H), or towards underestimating risk (L) 	MP; MM; IC; TJ
<ul style="list-style-type: none"> If both can be estimated, have variance and bias been consistent over time (H), or have they varied substantially (L) 	MP; MM; IC; TJ
<ul style="list-style-type: none"> Probability that IND value exceeds reference point can be estimated with accuracy and precision (H), to coarsely or not at all (L)** 	MM; IC; TJ (type of risk quantification is important)
<ul style="list-style-type: none"> IND measured using tools with known accuracy and precision (H), to unknown or poor/inconsistent (L) 	MP; MM; IC; TJ
<ul style="list-style-type: none"> Value obtained for indicator unaffected by sampling gear (H), to sampling methods can be calibrated (M), to calibration difficult or not done (L) 	SI; MP; IC; TJ
<ul style="list-style-type: none"> Seasonal variation unlikely or highly systematic (H) to irregular (L) 	SI; MP; MM; IC; TJ
<ul style="list-style-type: none"> Geographic variation irrelevant or stable and well quantified (H), through random (M) to systematic on scales inconsistent with feasible sampling (L)** 	SI; MP; IC; TJ
<ul style="list-style-type: none"> Taxonomic representivity: IND reflects status of all taxa sampled/modelled (High), through ecologically predictable subset of species (M), to only specific species with no identifiable pattern of representivity (L) 	SI; MP; IC; TJ

(continued)

Table 2 (continued)

Criteria and sub-criteria	Method of evaluation
Availability of historical data	
<ul style="list-style-type: none"> Necessary data are available for: periods of several decades (H) to only relatively recent period (M), to opportunistic or none available (L) 	MP; IC; TJ
<ul style="list-style-type: none"> Necessary data are: from the full area of interest (H), to restricted but consistent sampling sites (Moderate), to opportunistic and inconsistent sources, or none (L)** 	MP; IC; TJ
<ul style="list-style-type: none"> Necessary data have high contrast, including periods of harm and recovery (H), to high contrast but without known periods of harm and recovery (M), to uninformative about range of variation expected (Low) 	MP; IC; TJ
<ul style="list-style-type: none"> The quality of the data and archiving is known and good (H), to data scattered with reliability but not systematically certified, and archives not maintained (L) 	MP (e.g. environmental IND); IC; TJ
<ul style="list-style-type: none"> Data sets are freely available to research community (H), to private or commercial holdings (L) 	IC
Sensitivity (length of time-series used for testing important)	
<ul style="list-style-type: none"> IND responds to fishing in ways that are: (i) smooth, monotonic, and with high slope (H)**; (ii) smooth, monotonic, and with low slope (M); (iii) smooth, monotonic over a restricted range of effort characteristics (M–F); (iv) unreliable (M–F, depending on when it fails to inform about fishing effects); (v) insensitive or irregular. Magnitude of response does not depend on magnitude of signal in effort (L) 	SI; MP; MM; MC; TJ
Responsiveness (length of time-series used for testing important)	
<ul style="list-style-type: none"> IND changes within 1–3 years of implementation of measures (H), to IND only reflects system responses to management on decadal scales or longer (L) 	SI; MP; MM; MC; TJ
Specificity (contrast in data set used for testing important)	
<ul style="list-style-type: none"> Is impact of environmental forcing on IND known, and small (H) or strong (L)? 	SI; MP; MM; MC; TJ
<ul style="list-style-type: none"> If environmental forcing affects IND, effect systematic and known (H), to irregular or poorly understood (L)** 	SI; MP; MM; MC; TJ
<ul style="list-style-type: none"> Relative to other factors, IND: (i) known to be unresponsive (H); (ii) responds to specific factors in known ways (M); (iii) thought to be unresponsive (F); (iv) responds to many factors in only partly understood ways (L)** 	SI; MP; MM; MC; TJ

both the score for information quality and some designation of the weight of evidence for that score, and sometimes may have multiple pairs of these for different dimensions of a criterion. This step describes how these two matrices are converted into information that can be used in the final selection process.

Of course, it would be possible to compute a final score of each indicator as the sum of the matrix products of weights by scores. Although this procedure provides unique scores that would facilitate subsequent crossing off the lower ranks, there are several reasons for advising against such a simple procedure:

- (i) Weighted averages could give moderate scores to candidate indicators that were strong on some criteria, but fatally flawed on others, and it is by no means certain that a few attractive properties balance other severe shortcomings.
- (ii) The approach would tend to give similar scores to indicators with similar properties, fostering selection of redundant rather than complementary indicators.
- (iii) It would be tempting to make the scoring on individual criteria more finely differentiated, without

sufficient information to justify such scores and neglecting the fact that scores are comparable within criteria only.

- (iv) Information on the strength of evidence, which we stressed as an important part of the evaluation process, would be disregarded.

Graphic methods such as radar plots have been proposed for display of ecosystem status, using semi-quantitative information on a selection of ecosystem indicators (Collie *et al.*, 2001). Conceptually, displaying the status of a candidate indicator on multiple criteria would be similar: the relative performance on each criterion would be reflected in the lengths and orientations on the different axes, and define the polygon. Hence, if the screening process were seeking indicators that could serve several uses, different regions of the plotting space could be taken as reflecting the ability of an indicator to meet the criteria associated most closely with each use. The differential weight of evidence could be reflected in the density of the filled space of the plot. Such a graphic approach would allow visual assessment of the degree to which an indicator corresponded to the properties desired by different user groups, and indicators

falling short on important criteria would be readily apparent. However, only very few competing indicators could be superimposed on a single set of axes, and comparative evaluations among many indicators would be problematic. Although really superior indicators would stand out clearly, the graphic displays become too complex to interpret when large numbers of indicators are being evaluated, or if performance is good and poor on different criteria.

Other proposed methods of data reduction include clustering algorithms for grouping sets of indicators with similar performance on the criteria (CSAS, 2001), and ordination methods for a spatial display of the relative positions of the indicators in spaces of lower dimensionality than the number of criteria (Link *et al.*, 2001; Pitcher and Preikhost, 2001). Either approach has potential flaws. Ordinal scores are likely to be poorly calibrated across criteria, and possibly even across different types of indicators. Information on strengths of evidence is hard (but not impossible) to preserve in such analyses. When ordination methods reduce dimensionality among criteria, they seek overall patterns of covariation. Hence, they relegate the subtle distinctions among criteria such as sensitivity, specificity, and responsiveness to later axes, often considered noise or “stress”, so disguising features important for decision-making.

The types of data used in the selection of candidate indicators have much in common with those used in psychometrics: multiple criteria that overlap in information content, but vary in importance for different uses. At best, they are ordinal scores of cases on the criteria, and varying strength of evidence. Psychometrics addresses these analytical problems, particularly in the field of personality and aptitude testing, by developing multi-dimensional response profiles using the test scores directly (Dorfman and Hersen, 2001; Murphy and Davidshofer, 2001). The response profiles are interpreted relative to normative samples – scores of hundreds to thousands of subjects, whose performance traits are known accurately on exactly the properties that the test is intended to measure. For many tests, different norms must be provided for applications in different contexts.

Two important messages come from this work. First, the information in inherently multi-dimensional traits (like indicator value) should not be collapsed into misleadingly simple aggregate scores. Second, it will be necessary to build-up normative scores of indicator suites for ecosystems perturbed in various known ways before it would be legitimate to interpret the values of indicators with various properties in management contexts. There is much to learn about how to approach this complex task.

Step 6 – decide how many indicators are needed

This step requires strong interaction among the ultimate users. For the reasons discussed in Steps 1 and 2, it is simultaneously desirable to have the fewest possible

number of indicators to serve all uses, while having all key system components featuring in the objectives covered by trustworthy indicators. This is where the information on other threats must be taken into consideration, together with knowledge of how they may affect different candidate indicators.

Decisions on the number of indicators to retain are aided by effective profiling of how the candidate indicators score on the evaluation criteria. Effective profiling (graphical or otherwise) should show whether there are a few clusters of indicators with similar attributes, or a diverse array of indicators, each with a distinctive set of performance characteristics. In the former case, the number to retain would be a small multiple of the number of clusters. The actual multiple would increase with the number of operational objectives decided upon, as well as the number of different system components addressed by the objectives. In the latter case, it would probably be inappropriate to set a fixed number of indicators separately from a discussion of the number and types of threats. Identification of multiple threats should result in selecting even larger numbers of indicators. Although this issue has not been studied formally, we expect the multiplier effect to depend on how the ecosystem effects of other threats resemble the ecosystem effects of fishing. The more similar the effects of multiple forcers, the greater the number of indicators needed to differentiate between the contributions of each forcer to the status and trends in the indicators. Without such differentiation, it would be less likely meaningful to use the indicators in selecting effective management actions. This is also where the need for normative profiles becomes clear, as is routine in psychometrics. Indicator-based decisions have to be tested retrospectively for different conditions (governance systems, combinations of threats), so objective information can be accumulated about what combinations of properties of indicators are the best guides for decision-making.

Step 7 – make final selection

When working directly with the matrices of scores by criteria, and criteria by stakeholder weights, selection should strive to find suites of indicators that perform well on all criteria important for each expected use, as well as to cover the entire spectrum of ecological, social, and economic objectives. If no candidate indicators perform well on all the important criteria for a given use, then the suite should try to balance strengths and weaknesses. In other words, some indicators in the suite should perform well on each important criterion, and members of a suite should not all perform poorly on the same criteria. When the suite is intended to serve multiple purposes, it should be more effective to select indicators matched well to each intended use, rather than to derive a compromise among uses, not performing particularly well for any of them.

In this step, the reasons for selection should be well documented and retained. When indicators with known shortcomings are retained because they have unique strengths as well, users need to keep them in mind when interpreting their values and making decisions. Also, tolerances for weaknesses or strengths might change over time for different reasons. Time-series data expand continuously (as does knowledge of an ecosystem and the effects of fishing), new forcings (natural or anthropogenic) might become important, and societal values could change. All these factors would be the causes to reconsider which indicators to use, or how they are interpreted in practice. Retaining the evaluation matrices and the reasons for the selection of indicators allows choices or uses to be adapted without repeating the entire exercise, thus enhancing consistency.

Step 8 – report on the suite of indicators

Given the final suite of indicators, it is necessary to present the annual (or other periodic) values effectively. We found many different presentation methods, each with advantages and drawbacks (Table 3). Many integrating methods require some standardization and weighting, and have the potential to reintroduce all problems in selecting the indicators (Steps 3–6) to begin with.

Several reporting aspects are often entangled. Indicators may be used to report: (i) the current state; (ii) the dynamics of the state; (iii) value judgements about the state (well or poor); or (iv) value judgements about the dynamics (improving or worsening). To avoid confusion, each aspect should be taken in a separate step. For example, under certain conditions, a set of state indicators may be aggregated, and the aggregated index compared relative to an objective for the aggregate value. Likewise, some methods advocate scoring the dynamics of a set of state indicators, and then aggregating or presenting the set of these scores (Bellail *et al.*, 2003; EEA, 2003), because temporal changes in the aggregate value, and by inference, in the ecosystem, can be tracked easily. However, aggregation methods have a risk of concealing the nature of what is being perturbed. Moreover, even if the indicators being aggregated cover the properties of the ecosystem well, perturbations such as fishing may affect some state indicators in one direction, and others in the opposite direction (the “eclipse” effect; Andreasen *et al.*, 2001). Users usually are aware of obvious conflicts in the directional response of different indicators to fishing, but the expected patterns are not always founded on good theory (e.g. state indicators of diversity). Hence, aggregated trends should always be used with caution.

Many methods in Table 3 apply weightings, where, again, there is ample opportunity to present misleading information, because they do not differentiate between weighting for methodological reasons (redundancy, unequal

uncertainty among indicators), and weighting for policy reasons. Again this makes changes in the weighted value of the aggregate score difficult to interpret without returning to the patterns observed in the individual indicators.

Thus, we are faced again with the trade-off between the complexity of trying to interpret large quantities of information, and the risks inherent in collapsing information in apparently simple ways. The solution must lie in developing reference profiles for interpreting each indicator individually. Unfortunately, this solution is a long-term one for marine application.

Use of indicators in decision support

If a large suite of indicators is to be used in a formal decision-support system, the number of inputs to the system will be correspondingly large. Formal guidance as to how they should be treated is largely lacking. For example, Annex III-B of the *Bergen Declaration (2002)* includes five indicators of eutrophication, accompanied by a footnote that the ecological quality objectives for each of these represent “an integrated set and cannot be considered in isolation”, but no guidance is provided on how that is to be achieved. The precautionary approach (FAO, 1996a, b) could be interpreted as requiring management action to be matched to the indicator with highest risk of being at or outside its conservation reference point. However, analytical risk management approaches indicate that an overall risk profile should be built up across all indicators, and it is that risk, not each component, which should be managed. This would present a major challenge in practice, but it is the intent implied by agencies adopting both indicator- and risk-based management principles.

In both cases, when a suite of indicators is retained because each member complements some deficiency of the others, the question remains how to carry that information into the overall decision-making process. Consider the comparatively simple case of management using single decision rules for each indicator, with each rule individually tailored to the strengths and weaknesses of the indicator, as reflected in the evaluation matrices. Unless all of the decision rules associated with each indicator require exactly the same management response, a family of “meta-rules” would have to be developed to determine which management response is appropriate.

The full management problem is even more difficult. Not only are there multiple indicators supporting dialogue or decision-making in relation to a given objective, but keeping the ecosystem effects of fishing within sustainable bounds requires multiple operational objectives as well (ICES, 2001, 2003; FAO, 2002). Many management actions may affect the probability of achieving several objectives at once. At the same time, they may cause new problems. For example, closed areas cause redistribution of effort. What may be gained for some species may be totally wrong for others (Dinmore *et al.*, 2003; Hilborn *et al.*, 2004).

Table 3. Review with pros (+) and cons (–) of three categories of methods from the literature, corresponding to the three steps in combining indicators (IND): (i) standardization, (ii) weighting, (iii) combining.

A. Methods for standardizing IND

- A1a. **Scoring:** Convert IND values into scores (discrete variation with limited number of classes)
 +: easy for qualitative variation
 –: usually arbitrary for quantitative variation; no explicit scoring method available; huge scope for subjectivity (Rochet and Rice, 2005)
- A1b. **Fuzzy scoring:** Convert to qualitative variation, with limited number of classes; score each observation from “no” (0) to “high” (5) affinity with each modality
 +: allows uncertainty and limited knowledge
 –: not much experience available; complex to explain
- A2a. **Linear interpolation between observed extreme values:** Scale all IND on a common range (e.g. [0, 1]), assuming linear variation between minimum and maximum values
 +: simple
 –: IND may not show linear variation; sensitive to history of data series
- A2b. **Linear interpolation between reference values:** Similar to A2a, but uses predefined reference values
 +: simple
 –: linear variation not always relevant; reference values often difficult to define
- A3. **Multivariate methods:** Usually performed on normalized variance, hence IND standardized by their s.d.
 +: accounts for uncertainty and variability
 –: sample dependent

B. Weighting methods

- B1. **Multivariate methods:** Projections on maximum inertia axes, so giving lower weight to correlated IND
 +: objective way of reducing redundancy without eliminating potentially useful IND
 –: management objectives not taken into account
- B2. **Analytical Hierarchy Process (AHP; Tran *et al.*, 2002):** Breakdown of problem into smaller constituent parts at different levels in hierarchy, followed by series of pairwise comparison judgements at each level
 +: user-defined weighting
 –: number of comparisons increases exponentially with number of IND and potential values

C. Methods for combining IND

(Graphic displays)

- C1. **Kites (Garcia and Staples, 2000):** One standardized IND per edge: outer rim = “good”, centre = “bad”; scores linked and resulting area possibly shaded
 +: quick and easy; not too many data manipulations; easy to understand
 –: polygon influenced by order of presentation; misleading (equal weight suggested for all IND); potential redundancy
- C2. **Pie slices (Andreasen *et al.*, 2001):** One standardized IND per slice; circumference = “degraded” reference condition; IND value shaded +/–: same as C1, but better (shaded area equal whatever the order)
- C3. **Amoeba (Collie *et al.*, 2001):** Circle = reference; arrow lengths = values; arrow directions = correlations between IND; shape influenced by relative variances
 +: takes account of redundancy, because based on IND correlation
 –: hard to display multiple IND

[Indices]

- C4a. **Weighted average (Andreasen *et al.*, 2001):** Standardize indicators, define weights and average
 +: simple
 –: outcome determined by standardization and weights; hard to test weighting validity; prone to eclipsing (good traits may obscure bad ones)
- C4b. **Weighted geometric average:** Multiplying weighted IND rather than summing to increase influence of “bad” scores
 +/–: same as C4a
- C5. **Indices of Biotic Integrity (IBI; Hughes *et al.*, 1998; McCormick *et al.*, 2001):** Define reference condition, based on minimally disturbed sites, historical data, or models; score continuously by linear interpolation between reference values; IBI = sum of scores/number of IND; eliminate redundant and inconsistent IND based on correlations; measure variability in IND and IBI using multiple sampling at each site and estimate power of IBI

(continued)

Table 3. (continued)

	+	scoring methods may be improved and weights introduced; specified rules for combining scores
	-	eclipsing and redundancy can distort scores, but may be reduced by additional rules to eliminate some IND
C6.		Fuzzy numbers (Tran <i>et al.</i> , 2002): Normalize IND with 0 (= ideal) and 1 (= undesirable) by linear interpolation; each normalized IND with its observed minimum and maximum in a given site make a fuzzy number; compute fuzzy distance of each IND to 0 and 1, and weight and aggregate the distances
	+	appealing because some way to transfer uncertainty towards aggregated levels
	-	sampling distribution must be specified, generally without <i>a priori</i> basis; sensitive to assumed distribution
C7.		Framework for ecologically sustainable development (Chesson and Clayton, 1998): Define hierarchical structure of assessment; standardize IND (e.g. by linear interpolation); weight and sum at desired level, using prior-chosen weights; examine trends
	+	hierarchical structure allows examination at different levels; recognition that process is subjective; dynamic approach; possible to explore use in pressure and impact studies
	-	no account of uncertainty in data
[Multivariate ordination methods]		
C8a.		MDS of scored IND (Pitcher and Preikhost, 2001): Choose attributes that are easily and objectively scored with obvious "good" and "bad" extremes; ordinate set of fisheries or trajectory of a fishery in time; MDS (first axis supposed to represent sustainability); construct fixed reference points (extreme scores for each attribute) and randomization test
	+	general advantages of multivariate methods
	-	scores are arbitrary; reference points misleading, because no fishery can simultaneously exhibit all indicators at extreme values
C8b.		PCA and canonical correlation analysis (Link <i>et al.</i> , 2001): Gather metrics of community and abiotic and human factors; PCA; interpret axes in terms of exploitation; canonical correlation analysis of community vs. factors
	+	general advantages of multivariate methods
	-	interpretation not always obvious (but possibly improved by CCA); not easy to understand
C8c.		Multivariate analysis (Charvet <i>et al.</i> , 2000): Measure IND in a set of communities; fuzzy scoring and correspondence analysis; hierarchical clustering; for each group, profiles of IND (frequency distributions of mean scores); reference point possibly given by extreme situations
	+/-	same as C8b

Conclusions

Indicator-based decision-making can give managers structured insight into the likely effects of alternative actions, which is essential in integrated management approaches. However, this is only true if the performance characteristics of the indicators are understood, and if their trends and current values relative to reference points can be interpreted correctly. This is a particularly compelling reason to attempt a formal screening of the performances of candidate ecosystem indicators, as outlined in the framework presented here, even if the actual choices are to be made by partisan political processes rather than scientific ones.

We have tried to design and implement a test of the framework outlined (Rochet and Rice, 2005), but much remains to be done to establish its validity. Even when the complete framework is tested in interactive settings with managers, stakeholders, and scientists each fulfilling their normal roles, making improvements as needed, we expect selection to continue usually by consensus and dialogue. Nevertheless, the important function of the framework lies in its potential to structure the dialogue. If all steps are included in the dialogue leading to the selection of the final suite of indicators, the most important stumbling blocks

should have been addressed. This would be an improvement over a haphazard or manipulative approach, and a step towards the rigour and transparency required and justified, given their importance in subsequent management.

Acknowledgements

We thank John Field and Niels Daan for careful reviews and numerous suggestions that allowed us to clarify our reasoning and to shorten the presentation substantially. This document is based on work partially supported by the U.S. National Science Foundation under Grant No. 0003700.

References

- Andreasen, J. K., O'Neill, R. V., Noss, R., and Slosser, N. C. 2001. Considerations for the development of a terrestrial index of biological integrity. *Ecological Indicators*, 1: 21–35.
- Belfiore, S. 2003. The growth of integrated coastal management and the role of indicators in integrated coastal management: introduction to the special issue. *Ocean and Coastal Management*, 46: 225–243.
- Bellail, R., Bertrand, J., Le Pape, O., Mahé, J.-C., Morin, J., Poulard, J.-C., Rochet, M.-J., Schlaich, I., Souplet, A., and

- Trenkel, V. 2003. A multispecies dynamic indicator-based approach to the assessment of the impact of fishing on fish communities. ICES Document, CM 2003/V: 02. 12 pp.
- Bergen Declaration. 2002. Ministerial Declaration. Fifth International Conference on the Protection of the North Sea, 2001, Bergen, Norway. <http://odin.dep.no/archive/mdvedlegg/01/11/Engel069.pdf>.
- Bowen, R. E., and Riley, C. 2003. Socio-economic indicators and integrated coastal management. *Ocean and Coastal Management*, 46: 229–312.
- Charles, A. T. 2001. Sustainable Fishery Systems. Fish and Aquatic Resources Series, 5. Blackwell Science, Oxford. 370 pp.
- Charvet, S., Statzner, B., Usseglio-Polatera, P., and Dumont, B. 2000. Traits of benthic macroinvertebrates in semi-natural French streams: an initial application to biomonitoring in Europe. *Freshwater Biology*, 43: 277–296.
- Chesson, J., and Clayton, H. 1998. A framework for assessing fisheries with respect to ecologically sustainable development. Bureau of Rural Sciences, Australia.
- Collie, J., Gislason, H., and Vinther, M. 2001. Using AMOEBAs to integrate multispecies, multifleet fisheries advice. ICES Document, CM 2001/T: 01.
- CSAS. 2001. Proceedings of the National Workshop on Objectives and Indicators for Ecosystem-based Management. DFO Canadian Science Advisory Secretariat Proceedings, 2001/09. 140 pp.
- Cury, P., Shannon, L., and Shin, Y. J. 2003. The functioning of marine ecosystems: a fisheries perspective. In *Responsible Fisheries in the Marine Ecosystem*, pp. 103–123. Ed. by M. Sinclair, and G. Valdimarsson. FAO/CAB International, Rome, Italy/Wallingford, UK.
- Dinmore, T. A., Duplisea, D. E., Rackham, B. D., Maxwell, D. L., and Jennings, S. 2003. Impact of a large-scale area closure on patterns of fishing disturbance and the consequences for benthic communities. *ICES Journal of Marine Science*, 60: 371–380.
- Dorfman, W. I., and Hersen, M. 2001. Understanding Psychological Assessment. Kluwer Academic, New York. 377 pp.
- EC. 2003. Communication from the Commission to the Council and the European Parliament: “towards a strategy to protect and conserve the marine environment”. Council Conclusion, March 7, 2003.
- EEA. 2003. Environmental Performance Indicators for the European Union. http://themes.eea.eu.int/indicators/all_indicators_box.
- FAO. 1996a. Precautionary approach to fisheries. 1. Guidelines on the precautionary approach to capture fisheries and species introductions. FAO Fisheries Technical Paper, 350/1. 52 pp.
- FAO. 1996b. Precautionary approach to fisheries. 2. Scientific papers. FAO Fisheries Technical Paper 350/2. 210 pp.
- FAO. 2002. Indicators for sustainable development of fisheries. <http://www.fao.org/docrep/W4745E/w4745e0f.htm>.
- FAO. 2003. The ecosystem approach to fisheries. FAO Technical Guidelines for Responsible Fisheries, 4(Suppl 2). 112 pp. <ftp://ftp.fao.org/docrep/fao/005/y4470e/y4470e00.pdf>.
- Garcia, S. 1996. Indicators for Sustainable Development of Fisheries. Proceedings of the 2nd World Fisheries Congress. Workshop on Fisheries Sustainability Indicators, Brisbane, Australia.
- Garcia, S. M., and Staples, D. J. 2000. Sustainability reference systems and indicators for responsible marine capture fisheries: a review of concepts and elements for a set of guidelines. *Marine and Freshwater Research*, 51: 385–426.
- Gottret, M. A., and White, D. 2001. Assessing the impact of integrated natural resource management: challenges and experience. *Conservation Ecology*, 5: 17–29.
- Helstrom, C. W. 1968. *Statistical Theory of Signal Detection*. Pergamon Press. 470 pp.
- Hilborn, R., Stokes, K., Maguire, J.-J., Smith, T., Botsford, L. W., Mangel, M., Orensanz, J., Parma, A., Rice, J., Bell, J., Cochrane, K. L., Garcia, S., Hall, S. J., Kirkwood, G. P., Sainsbury, K., Stefansson, G., and Walters, C. 2004. When can marine reserves improve fisheries management? *Ocean and Coastal Management*, 47: 197–205.
- Hughes, R. M., Kaufman, P. R., Herlihy, A. T., Kincaid, T. M., Reynolds, L., and Larsen, D. P. 1998. A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences*, 55: 1618–1631.
- ICES. 2001. Report of the Working Group on Ecosystem Effects of Fishing. ICES Document, CM 2001/ACME: 09.
- ICES. 2002. Report of the Advisory Committee on Ecosystems. ICES Cooperative Research Report, 254. 131 pp.
- ICES. 2003. Report of the Advisory Committee on Fisheries Management. ICES Cooperative Research Report, 264 (3 volumes).
- Jennings, S., Greenstreet, S. P. R., and Reynolds, J. D. 1999. Structural change in an exploited fish community: a consequence of differential fishing effects on species with contrasting life histories. *Journal of Animal Ecology*, 68: 617–627.
- Link, J. S., Brodziak, J. K. T., Edwards, S. F., Overholtz, W. J., Mountain, D., Jossi, J. W., Smith, T. D., and Fogarty, M. J. 2001. Ecosystem status in the Northeast United States continental shelf ecosystem: integration, synthesis, trends and meaning of ecosystem metrics. Or getting the brass tacks of ecosystem based fishery management. ICES Document, CM 2001/T: 10. 41 pp.
- McCormick, F. H., Hughes, R. M., Kaufmann, P. R., Peck, D. V., Stoddard, J. L., and Herlihy, A. T. 2001. Development of an Index of Biotic Integrity for the Mid-Atlantic Highlands region. *Transactions of the American Fisheries Society*, 130: 857–877.
- MSC. 2004. MSC Principles and Criteria for Sustainable Fisheries. Marine Stewardship Council. <http://www.msc.org/html/>.
- Murphy, K. R., and Davidshofer, C. O. 2001. *Psychological Testing: Principles and Applications*, 5th edn. Prentice Hall, NJ. 544 pp.
- NRC. 1994. Restoring and Protecting Marine Habitat: the Role of Engineering and Technology. National Research Council, National Academy Press, Washington, DC.
- OECD. 1993. OECD Core Set of Indicators for Environmental Performance Reviews. Organisation for Economic Co-operation and Development, Paris.
- OECD. 1998. Towards Sustainable Development: Environmental Indicators. Organisation for Economic Co-operation and Development, Paris.
- Olsen, S. B. 2003. Frameworks and indicators for assessing progress in integrated coastal management initiatives. *Ocean and Coastal Management*, 46: 347–361.
- Piet, G. J., and Rice, J. C. 2004. Performance of precautionary reference points in providing management advice on North Sea fish stocks. *ICES Journal of Marine Science*, 61: 1305–1312.
- Pitcher, T., and Preikhost, D. 2001. RAPFISH: a rapid appraisal technique to evaluate the sustainability status of fisheries. *Fisheries Research*, 49: 255–270.
- Rice, J. 1995. Food web theory, marine food webs, and what climate change may do to northern marine fish populations. In *Climate Change and Northern Fish Populations*, pp. 561–568. Ed. by R. J. Beamish. Canadian Special Publications on Fisheries and Aquatic Sciences, 121.
- Rice, J. C. 2003. Environmental health indicators. *Ocean and Coastal Management*, 46: 235–259.
- Rochet, M.-J., and Rice, J. C. 2005. Do explicit criteria help in selecting indicators for ecosystem-based fisheries management? *ICES Journal of Marine Science*, 62: 528–539.
- Smith, A. D. M., Sainsbury, K. J., and Stevens, R. A. 1999. Implementing effective fisheries-management systems – management

- strategy evaluation and the Australian partnership approach. *ICES Journal of Marine Science*, 56: 967–979.
- Talaue-McManus, L., Smith, S. V., and Buddemeier, R. W. 2003. Bio-physical and socio-economic assessments of the coastal zone: the LOICZ approach. *Ocean and Coastal Management*, 46: 323–333.
- Tran, L. T., Knight, C. G., O'Neill, R. V., Smith, E. R., Riitters, K. H., and Wickham, J. 2002. Fuzzy decision analysis for integrated environmental vulnerability assessment of the mid-Atlantic region. *Environmental Management*, 29: 845–859.
- UNCSD. 2001. *Indicators of Sustainable Development: Guidelines and Methodologies*. United Nations Commission on Sustainable Development, Washington, DC.
- Walker, B. S., Carpenter, S., Anderies, N., Able, N., Cumming, G. S., Janssen, M., Lebel, L., Norberg, G., Peterson, D., and Richard, R. 2002. Resilience management in socio-economic systems: a working hypothesis for a participatory approach. *Conservation Ecology*, 6: 14–17.
- World Bank. 2002. *Environmental Performance Indicators, 2002*. <http://Worldbank.org/html/opr/pmi/envindi2.htm>.