Independent Multidisciplinary Science Team (IMST)

State of Oregon

Neil W. Christensen
Michael J. Harte
Robert M. Hughes
Victor Kaczynski
Nancy Molina
Carl Schreck
Carlton Yee

c/o
Oregon State University
Department of Forest Science
321 Richardson Hall
Corvallis OR 97331-5752

September 10, 2007

Tom Byler
Executive Director
Oregon Watershed Enhancement Board
775 Summer St. NE Ste 360
Salem, OR 97301-1290

Dear Tom:

Enclosed is the Independent Multidisciplinary Science Team's (IMST) report titled *Considerations for the Use of Ecological Indicators in Restoration Effectiveness Evaluation*.

This report was written in response to the Oregon Watershed Enhancement Board's (OWEB) request that the IMST provide a review of the scientific basis for the use of high-level ecological indicators in determining the effectiveness of Oregon's efforts to restore aquatic habitats and watersheds under the Oregon Plan for Salmon and Watersheds (Oregon Plan).

To meet this request, the IMST reviewed a large amount of literature, including scientific articles, governmental agency reports, formal reviews, and web-based information. In addition, we took an in-depth look at four case studies of large-scale restoration programs that use ecological indicators: Chesapeake Bay Program, South Florida Ecosystem Restoration Program, CALFED Bay-Delta Program, and Ohio Environmental Protection Agency's water quality improvement program. We also looked at indicator development in Oregon, namely the Oregon Department of Forestry's program, *Oregon Indicators of Sustainable Forest Management*, and the Institute for Natural Resource's *Environmental Indicators for the Oregon Plan for Salmon and Watersheds*.

Our report does not contain a critique of indicator programs *per se*. Rather, we provide a synthesis of the best available science and "lessons learned" gleaned from all of the sources we reviewed. We have attempted to balance discussion of the benefits of the use of ecological indicators with discussion of some of the associated technical and analytical challenges. Our reference sections contain numerous articles that further discuss points made in our report, and recommend consultation with these original sources for additional information. We solicited two outside peer reviews for our report, one from Oregon, and the other from Florida. The final report was adopted with full consensus of the Team on July 31, 2007.

We do not make any formal recommendations to the State in this report, but we do summarize key findings. They are:

- The use of ecological indicators is a promising approach to the assessment and monitoring of broad-scale ecosystem conditions. However, there are few examples of long-term indicator employment, especially for effectiveness monitoring, from which to evaluate success. One exception being the use of indices of biological integrity (IBIs) in water quality monitoring.
- If the goal of monitoring is to evaluate management effectiveness, the basis for cause-and-effect determination must be built into sampling design and analysis protocols. An essential starting point is a conceptual model that transparently depicts assumptions about significant factors that affect the indicator, and is supported by relevant research. Attempting to make effectiveness determinations from status-and-trend monitoring data without such a model can be highly problematic.
- It is highly desirable to supplement effectiveness monitoring programs with research plans that address knowledge gaps about ecosystem responses to changing conditions.
- Key questions to address in indicator monitoring framework design (including sampling design and analytical protocols) are:
    - Is the goal of monitoring simply to track status and trends of ecosystem features (descriptive), or is it to evaluate ecosystem responses to particular management actions or stressors (adaptive)?
    - If the goal of monitoring is adaptive, will monitoring results be judged against threshold values (for example, TMDLs), or some suite of desirable environmental states (for example, habitat distribution for rare species)?
- Significant technical and analytical issues that require attention during indicator framework development include:
    - The aggregation of factors into multi-metric or multivariate indicators,
    - Determining appropriate sampling design, and statistical methods for analysis of monitoring data,
    - Assessing response variables across temporal and spatial scales,
    - Temporal and spatial variation of indicators,
    - Aggregation of data from disparate data sets.
- For each indicator used, the following information is desirable:
    - How the indicator relates to ecosystem attributes of interest, and how reliable it is for reflecting those attributes,
    - The range of values the indicator can take, and the significance of those values,

- How the indicator value varies with natural and anthropogenic influences, and the temporal and spatial scales over which those changes occur,
- Historical trends in indicator values,
- The spatial extent of the indicator, and the ecological types for which it is applicable,
- Information about indicator cost-effectiveness, including alternative or emerging measurement technologies, as appropriate.

- In the absence of reference points against which to compare monitoring results, the magnitude of change and significance of trends cannot be evaluated. Reference points are often chosen to depict "natural" conditions, but can also include minimally disturbed, best attainable or most degraded (worst-case) sites, depending on the nature of the monitoring questions.
- Probabilistic random sampling designs appear more efficient at quantifying changes in ecological conditions due to anthropogenic activities than non-random designs. At the same time, in most real-world applications, random designs have tended to better characterize high frequency/low severity situations than low frequency/high severity events, due to the expense associated with a dense sampling network.
- A key facet of an indicator monitoring program is how information is communicated to policy-makers and the public. Clear presentation of transparent and meaningful information to intended audiences is crucial, and there are several examples presented as case studies (Appendix A of this report) that attempt to accomplish these goals.

We hope the information in this report will be helpful to you and the other Oregon Plan agencies as you move forward with consideration of ecological indicators. Please do not hesitate to contact IMST if you have any questions regarding this report.

Sincerely,

Nancy Molina by
Carl B. Schreck

Carl B. Schreck

Nancy Molina
IMST Co-Chair

Carl Schreck
IMST Co-Chair

cc: Suzanne Knapp, GNRO
Greg Sielglitz, OWEB
IMST

This page left blank intentionally for double sided printing

Considerations for the Use of Ecological Indicators in Restoration Effectiveness Evaluation

IMST Technical Report 2007-1

Released September 10, 2007

Independent Multidisciplinary Science Team
Oregon Plan for Salmon and Watersheds
http://www.fsl.orst.edu/imst

Members:
Neil W. Christensen
Michael Harte
Robert M. Hughes
Vic Kaczynski
Nancy Molina
Carl Schreck
Carlton Yee

Citation: Independent Multidisciplinary Science Team. 2007. Considerations for the Use of Ecological Indicators in Restoration Effectiveness Evaluation. IMST Technical Report 2007-1 Oregon Watershed Enhancement Board, Salem, Oregon.

Report preparation: This IMST report is based on initial draft documents prepared by an IMST subcommittee that included Nancy Molina, Neil Christensen, and Kathy Maas-Hebner (OSU Sr. Faculty Research Assistant). Significant contributions were made to the technical and analytical section by IMST members Michael Harte and Bob Hughes. One draft was technically reviewed by Don Stevens (Department of Statistics, Oregon State University) and Bob Doren (South Florida Ecosystem Restoration Task Force). Susie Dunham (OSU Faculty Research Assistant) assisted in editing and producing the final document. Draft outlines and content were discussed at the October 16 and September 27, 2006, and January 18 and May 17, 2007 IMST public meetings. Draft reports were discussed at IMST public meetings on February 23 and June 28, 2007. The final draft was discussed and unanimously adopted at the July 31, 2007 IMST public meeting.

Table of Contents

Acronyms and Abbreviations

ANOVA	analysis of variance
CADDIS	Causal Analysis/Diagnosis Decision Information System
CALFED	CALFED Bay-Delta Program
CBP	Chesapeake Bay Program
EMAP	Environmental Monitoring and Assessment Program
EPA	Environmental Protection Agency
FTE	full-time equivalent
GAO	United States Government Accountability Office
HRV	historical range of variation
IBI	Index of Biological Integrity
IBIs	indices of biological integrity
IMST	Independent Multidisciplinary Science Team
INR	Institute for Natural Resources
MASC	Monitoring and Analysis Subcommittee
NRC	National Research Council
ODF	Oregon Department of Forestry
ODFW	Oregon Department of Fish and Wildlife
Oregon Plan	Oregon Plan for Salmon and Watersheds
OWEB	Oregon Watershed Enhancement Board
SCG	Science Coordination Group
SFSCG	South Florida Science Coordination Group
SFER	South Florida Ecosystem Restoration (Task Force)
South Florida	South Florida Ecosystem Restoration
USEPA	United Stated Environmental Protection Agency

Key Findings

- The use of ecological indicators is a promising approach to the assessment and monitoring of broad-scale ecosystem conditions. However, there are few examples of long-term indicator use, especially for effectiveness monitoring, from which to evaluate success. One exception is the use of indices of biological integrity (IBIs) in water quality monitoring.

- If the goal of monitoring is to evaluate management effectiveness, the basis for cause-and-effect determination must be built into sampling design and analysis protocols. An essential starting point is a conceptual model that transparently depicts assumptions about significant factors that affect the indicator, and is supported by relevant research. Attempting to make effectiveness determinations from status-and-trend monitoring data without such a model can be highly problematic.

- It is highly desirable to supplement effectiveness monitoring programs with research plans that address knowledge gaps about ecosystem responses to changing conditions.

- Key questions to address in indicator monitoring framework design (including sampling design and analytical protocols) are:
  - Is the goal of monitoring simply to track status and trends of ecosystem features (descriptive), or is it to evaluate ecosystem responses to particular management actions or stressors (adaptive)?
  - If the goal of monitoring is adaptive, will monitoring results be judged against threshold values (e.g., Total Maximum Daily Loads), or some suite of desirable environmental states (for example, habitat distribution for rare species)?

- Significant technical and analytical issues that require attention during indicator framework development include:
  - The aggregation of factors into multi-metric or multivariate indicators,
  - Determining appropriate sampling design, and statistical methods for analysis of monitoring data,
  - Assessing response variables across temporal and spatial scales,
  - Temporal and spatial variation of indicators,
  - Aggregation of data from disparate data sets.

- For each indicator used, the following information is desirable:
  - How the indicator relates to ecosystem attributes of interest, and how reliable it is for reflecting those attributes,

  - The range of values the indicator can take, and the significance of those values,
  - How the indicator value varies with natural and anthropogenic influences, and the temporal and spatial scales over which those changes occur,
  - Historical trends in indicator values,
  - The spatial extent of the indicator, and the ecological types for which it is applicable,
  - Information about indicator cost-effectiveness, including alternative or emerging measurement technologies, as appropriate.

- In the absence of reference points against which to compare monitoring results, the magnitude of change and significance of trends cannot be evaluated. Reference points are often chosen to depict "natural" conditions, but can also include minimally disturbed, best attainable or most degraded (worst-case) sites, depending on the nature of the monitoring questions.

- Probabilistic random sampling designs appear more efficient at quantifying changes in ecological conditions due to anthropogenic activities than non-random designs. At the same time, in most real-world applications, random designs have tended to better characterize high frequency/low severity situations than low frequency/high severity events, due to the expense associated with a dense sampling network.

- A key facet of an indicator monitoring program is how information is communicated to policy-makers and the public. Clear presentation of transparent and meaningful information to intended audiences is crucial, and there are several examples presented as case studies (Appendix A of this report) that attempt to accomplish these goals.

Introduction

The concept of ecological indicators has been used extensively for a century to assess water quality in Europe and North America (Kolkwitz & Marsson 1909; Forbes & Richardson 1913; Keup *et al*. 1967; Wilm 1975). Since the 1995 publication of the Montréal Process Working Group's *Criteria and Indicators for the Conservation and Sustainable Management of Boreal and Temperate Forests*[1], interest in using ecological indicators to monitor ecosystem health has increased worldwide. Many indicator monitoring frameworks have been instituted, with varying purposes and degrees of rigor. Practitioners have grappled with basic questions, such as: What is an indicator? What exactly is to be indicated? What is the appropriate role of indicators in monitoring ecosystem conditions and the effectiveness of management policies or actions? How are indicators related to one another and to information on other factors that may be influencing the monitoring results?

At the conceptual level, the use of indicators for monitoring ecosystem conditions offers a promising approach. There are examples where indicators have been used successfully in effectiveness monitoring, especially where water quality is the parameter of interest. At the same time, a number of methodological challenges have arisen, and practitioners have often resolved these problems in intuitive ways with limited support from theoretical research (Shogren & Nowell 1992). In contrast to economic indicators (e.g., stock market indices), those using ecological indicators do not have a long track record of experience to draw from (NRC 2000). Ecological indicators have been developed and used in the US for a relatively short time to see meaningful results for some ecosystem attributes (e.g., recovery of forest structures in cutover watersheds), but sufficient time for others (e.g., changes in aquatic biota assemblages following curtailment of point source and diffuse pollution). Similarly, restoration effectiveness monitoring has often been inconsistent or sporadic, and success has been difficult to determine (Bernhardt *et al.* 2007; Hassett *et al.* 2007; Katz *et al.* 2007; Kondolf *et al.* 2007; Palmer *et al.* 2007; Rumps *et al.* 2007; Shah *et al.* 2007; Tompkins and Kondolf 2007). That said, a wide variety of ecological indicator applications have been undertaken with a large measure of thoughtful development and scientific oversight. The "lessons learned" from these efforts provide useful insights for the role of indicators in Oregon's salmonid recovery efforts.

Purpose of this Report

This report was prepared by IMST following a joint workshop with Oregon Watershed Enhancement Board (OWEB) which identified the need for general scientific guidance for the use of indicators in the evaluation of watershed restoration and salmonid recovery efforts (IMST 2006). It is intended for use by Oregon Plan for Salmon and Watersheds (Oregon Plan) partner agencies in deliberations about the use of system-level or "high-level" ecological indicators in answering questions about the overall effectiveness of the Oregon Plan. This report focuses on

[1] More information on the Montréal Process can be viewed at http://www.rinya.maff.go.jp/mpci/rep-pub/1995/santiago_e.html. Accessed on-line September 6, 2007

scientific topics related to developing, analyzing, reporting, and using ecological indicators. Political, economic, and social indicators are not addressed.

For the purpose of this report, IMST defines ***ecological indicators*** as measurable characteristics of the composition, structure, and function (or process) of ecological systems (following Niemi & McDonald 2004). The term "***high-level indicator***" is not widely used in the ecological literature, but state agencies in Oregon and Washington currently use this term to refer to ecological indicators that convey information about watersheds, river basins, or sub-regions. In keeping with this intent, IMST defines ***"high-level" ecological indicators*** as measurable characteristics that provide information relevant at these larger spatial scales, for the purposes of this report.

This report provides a synthesis of information and "lessons learned" from research, regulatory entities, and other organizations using ecological indicators in large-scale restoration efforts. The report is structured into three parts plus appendices:

- **Part 1 – Conceptual Issues** - deals generally with using ecological indicators for monitoring and assessing restoration effectiveness.
- **Part 2 – Technical and Analytical Challenges** - addresses specific analytical and methodological issues in the development and use of ecological indicators.
- **Part 3 – Communicating Information About Ecological Indicators** - summarizes information about presenting indicator monitoring results.

Four programs (Chesapeake Bay Program, (Chesapeake Bay) South Florida Ecosystem Restoration (South Florida), CALFED Bay-Delta Program (CALFED), Ohio Environmental Protection Agency (Ohio EPA)) were chosen as case studies to illustrate key points in this report (see Appendix A for detailed summaries of these programs). In its review, IMST considered independent critical reviews of the Chesapeake Bay (GAO 2005), South Florida (GAO 2003, 2007), and CALFED (Little Hoover Commission 2005) programs. In the majority of these programs, ecological indicators were part of an overall adaptive management strategy to improve the effectiveness of restorative management regimes, and reduce uncertainty about how systems respond to aggregated treatments. Used in this manner, indicator information can support the development of protocols for collecting and interpreting new information specifically geared toward improving plans and accounting for uncertainty (SFSCG 2006). Comparison of the four case studies highlighted a few key characteristics that distinguish programs with successful indicators (Table 1), including a strong monitoring framework, and quantifiable targets for both indicator values (especially biological criteria) and management actions.

Table 1. Comparison of the aquatic components of four restoration programs. *(Note: This comparison was developed by IMST from sources cited in Appendix A of this report, and may not reflect subsequent changes*)

Variable	Restoration Program Chesapeake Bay	South Florida	CALFED	Ohio EPA
Biological Indicators				
Assemblages	5	4	0	2
Multimetric indices	2	0	0	3
Metrics	29	33	0	28
Target Species	6	7	2	0
Biocriteria	0	3	0	3
Environmental indicators				
Flow Alteration	Yes	Yes	yes	yes
Physical Habitat Alt.	Yes	No	no	yes
Water Quality	Yes	Yes	yes	yes
Pollution Concentration	Yes	No	no	yes
Effluent Loadings	Yes	No	no	yes
Land Use	Yes	No	no	yes
Management Indicators				
Plans	Yes	Yes	yes	yes
Permits	Yes	Yes	no	yes
Grants	No	Yes	yes	yes
Enforcement	No	No	no	yes
Best Management Practices	Yes	Yes	yes	yes
Treatments	Yes	Yes	yes	yes
Major Disturbances				
Waste water	Yes	No	yes	yes
Agriculture	Yes	Yes	yes	yes
Hydrological Modification	Yes	Yes	yes	yes
Monitoring				
Duration (yrs)	24	11	13	28
Sites monitored/yr	unknown*	unknown**	unknown***	100s
Annual FTE/Millions $	unknown/2.1*	unknown/53.0**	unknown/11.2***	16/1.6
Study Design	ad hoc*	ad hoc	ad hoc	ad hoc
Annual Reporting	unknown	unknown	unknown	yes

*Maryland Department of Natural Resources conducts biological monitoring on 100s of sites/year with a $1.5M budget and uses a probability design. It is unknown whether this state-level information is integrated into the Chesapeake Bay Program.
**Florida Department of Environmental Protection conducts biological monitoring on 100s sites/year for 15 years, no agency response to budget request. It is unknown whether this state-level information is integrated into the South Florida Program.
***California Department of Fish and Game conducts biological monitoring on 100s of sites/year with a $0.9 million budget and 9 FTE. It is unknown whether this state-level information is integrated into the CALFED Program.

Ecological Indicator Development in Oregon

In 2000, the Governor's Natural Resource Office and Oregon Sea Grant sponsored a graduate thesis focused on setting measurable goals for the Oregon Plan. In this thesis O'Mealy (2000) presented a comprehensive summary of three large scale restoration programs that are similar to the Oregon Plan (South Florida, CALFED, and Chesapeake Bay). For this report, IMST reviewed more recent information pertaining to the programs evaluated by O'Mealy and found that many conclusions drawn in O'Mealy (2000) remain relevant, especially with regard to the broader social and governmental context within which indicator monitoring occurs.

In 2003, OWEB partnered with the Institute for Natural Resources (INR) to develop a system to track a small set of environmental indicators throughout Oregon (Dent *et al.* 2005). OWEB and INR hosted a workshop and brought together a diverse group of participants to discuss and recommend a set of environmental indicators. Although this effort appears to be a good first step for Oregon, the indicators monitoring scheme has not been fully implemented, and therefore IMST has not evaluated the indicators as part of this report.

The Oregon Department of Forestry (ODF) also has taken significant steps toward developing high-level indicators under a program entitled "Oregon Indicators of Sustainable Forest Management" (ODF 2007). The ODF program includes underpinnings for soil and water indicators derived from Dent *et al.* (2005). Because this program has not been fully implemented, IMST did not perform an in-depth review of it for this report. It does incorporate many of the key features recommended throughout the scientific literature on high-level indicators. IMST believes the ODF initiative provides an important foundation for high-level indicator use in Oregon, and supports ongoing efforts by ODF and its partner agencies in the Oregon Plan to assess needs and flesh out details within the general framework it presents. Because it is a new program, many details require refinement, especially with regard to analysis and interpretation of data. IMST hopes this report will facilitate further developments by summarizing the many lessons learned from other high-level indicator programs.

PART 1 - CONCEPTUAL ISSUES

Part 1 of this report is designed to provide general knowledge of ecological indicators and their potential uses in restoration effectiveness evaluation.

Definitions

This IMST report adopts the definition of ecological indicators as measures of ecosystem composition, structure, or function (Niemi & McDonald 2004) that provide needed information on environmental conditions, with “high-level” indicators representing broader-scale phenomena, as stated above. ***Composition*** refers to the identity and variety of elements in the system(s) studied. The organization or pattern of a system forms its ***structure***. ***Function*** alludes to processes and interactions that occur within and between economic, ecological and social systems, as well as their evolutionary development. The composition of an ecosystem can be measured using indicators of species identity, relative abundance, richness or diversity. The relative proportions of endangered, threatened, indigenous, or exotic species could also be measured with indicators. Habitat complexity, population age structure, distribution of physical features, and biogeoclimatic attributes are examples of indicators that may be used to describe the structural organization of ecological systems. Variables that can be expressed as rates, such as energy flows, biological productivity, succession rates, nutrient cycling, and hydrological processes, are examples of functional indicators (Noss 1990; Niemi & McDonald 2004).

An indicator may be comprised of a single parameter, such as road density to represent watershed health (i.e., a ***single-metric indicator***). Or an indicator may integrate several factors into an “index”, such as combining land cover type, water temperature, and road density to create a watershed health index (i.e., a ***multi-metric indicator***). To further complicate matters, high-level indicators may reflect intrinsically broad-scale information (for example, the proportion of the area of a watershed that is urbanized), or they may be comprised of finer-scale information that is “rolled up” or extrapolated from individual sampling points (for example, salmon abundance in a river basin, inferred from a probabilistic sampling design). Regardless of the geographic scale represented by an indicator, the issues concerning its selection and use are similar.

Indicators and Monitoring

Ecological indicators are useful insofar as they are linked to rigorous monitoring frameworks, produce results that are scientifically defensible, and can be easily understood and accepted by policy makers and the public (Jackson *et al.* 2000). Indicators have been shown to be helpful in accomplishing the following tasks (Dale & Beyeler 2001; Kurtz *et al.* 2001):

- Assessing and describing ecological conditions,
- Monitoring trends in phenomena of interest to society,
- Providing an early warning signal of deterioration in key ecosystem attributes,

- Diagnosing the causes of observed problems,
- Quantifying:
    - The magnitude of stresses occurring in the system,
    - The degree of exposure to stress, and
    - The ecological responses to exposure to stress, and
- Establishing thresholds for management action.

The growing body of literature on ecological indicators describes several ways to group and characterize indicators into “schemes” reflecting the underlying monitoring approaches. For example, the US Government Accountability Office (GAO 2005) and South Florida Science Coordination Group (SFSCG 2006) both describe a scheme consisting of three types of indicators: (1) those that measure programmatic or administrative progress, (2) those that measure the states of stressors or drivers (controlling factors or actions that cause change), and (3) those that assess outcomes in terms of ecological conditions. The NRC (2000) identified two types of indicators (1) those that represent the status or condition of a system, and (2) those that seek to identify cause-and-effect relationships. The distinction made in NRC (2000) is critical when ecological indicators are used to evaluate restoration effectiveness. While this is not a comprehensive list of schemes for categorizing indicators, it does illustrate variation among practitioners and types of monitoring for which indicators are used.

Indicators are frequently used to monitor ecological change but are designed and used differently depending on monitoring goals. ***Status and trend*** monitoring documents changes in conditions, stressors, or responses over time, without necessarily determining causes of observed results. ***Compliance*** (or ***implementation***) monitoring seeks to measure the extent to which the steps of a plan have been carried out, or how rigorously regulations are being adhered to. ***Effectiveness monitoring*** seeks to answer questions about the effectiveness of management activities or regimes to achieve target or predicted results over time. ***Validation monitoring*** involves testing assumptions about cause-and-effect relationships among conditions, stressors, and management activities, and is commonly considered equivalent to research (hypothesis testing). With the proper sampling and analytical framework, status/trend and effectiveness monitoring can be combined to maximize efficiency and use of information.

Conceptual Models as Part of Indicator Frameworks

A key component in all four restoration programs reviewed by IMST was the use of conceptual models to explicitly depict user assumptions about relationships among indicators. A ***conceptual model*** is “a visual and/or narrative explanation of how a system works or is expected to respond” (CALFED 2006a). In its deliberations to identify indicators for the US, the NRC (2000, page 29) indicated that “useful ecological indicators are based on clear conceptual models of the structure and functioning of the ecosystems to which they apply”. Conceptual models can be empirical or theoretical, quantitative or qualitative, but in any case some type of model is needed to identify known and assumed relationships between indicators, and how each indicator is expected to respond to ecosystem changes (NRC 2000). Ideally, conceptual models detail the assumed relationships between composition, structures, patterns, and processes relevant to the ecosystem being monitored (e.g., Figure 1, nutrient flow diagrams, food webs, Bayesian Belief Networks, Markovian matrices of landscape change).

To be an effective monitoring tool, a conceptual model must help identify and explain the causes of observed indicator trends and include all significant factors that affect the status of the indicator. Ideally, conceptual models include transparent use of quantitative information where available, listing of information sources, and relative weights of model variables (e.g., Figure 2; Table 2). One widely used system, Pressure-State-Response model, portrays relationships among a controlling factor (pressure), a condition upon which is it acting (state), and the outcome (response). Similarly, Driver-Linkage-Outcome models display relationships for several or many controlling factors (CALFED 2006a). The Pressure-State-Response model is widely applied but has also been criticized by Rapport & Singh (2006) who observe that, by isolating pressures, states and responses, the Pressure-State-Response model tends to be static and ignores the dynamic links between the components (the Driver-Linkage-Outcome model is one attempt to overcome this problem).

Conceptual models should facilitate interpretation of indicator response trends within the context of inherent environmental variation (Landres 1992; Wiersma 2005), including seasonal fluctuations, natural disturbance regimes, and other external stressors. To this end, social factors and anthropogenic ecosystem stressors can be important conceptual model elements (Harwell *et al.* 1999), along with global-scale drivers like climate change or population growth (GAO 2005) that may confound the interpretation of indicators.

The scientific underpinnings for conceptual models can come from a variety of qualitative and quantitative sources, including professional judgment, natural history, paleoecology, designed experiments, analytical predictions, computer simulations, and remote sensing time series (NRC 2000). One of the most significant drawbacks to developing conceptual models is a relative lack of ecological research information on interactions between ecosystem components. The South Florida Ecosystem Restoration program found significant variation in levels of research, and consequent uncertainty, about how indicators related to ecological features, conditions, and limiting factors (SFSCG 2006). This problem is particularly acute where indicators are used to monitor cause and effect relationships (e.g., effectiveness of restoration activities), in contrast to simply characterizing ecological conditions. Rigorous conceptual models can provide a framework for generating research hypotheses, which can subsequently lead to results that aid model refinement.

In practice, conceptual models typically include a combination of qualitative and quantitative components and consequently vary in format and degree of analytical rigor. Both quantitative and qualitative aspects of conceptual models are often developed through the use of expert panels. Key tasks for expert panels developing conceptual models are summarizing research findings for model input, identifying techniques to minimize subjectivity and bias that might arise in the process of weighting, prioritizing, or otherwise defining relationships among model variables, and openly discussing areas of inherent uncertainties and unpredictability. The benefits of transparently discussing uncertainties, and the effect those uncertainties have on conclusions, cannot be over-emphasized.

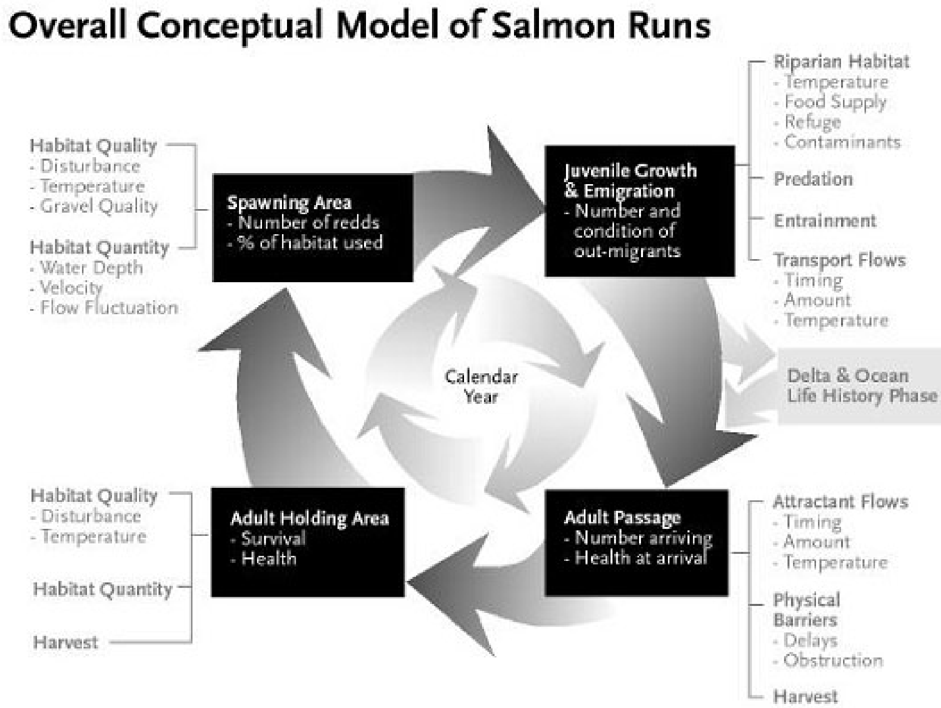

Figure 1. Example of a qualitative conceptual model from CALFED Bay-Delta program (http://science.calwater.ca.gov/sci_tools/salmon_model.shtml; Accessed on July 11, 2007). This approach illustrates the potential linkages among model components but does not provide many details. This type of figure does not include information on the source of knowledge behind the linkages, including whether there is documentation from research.

2A

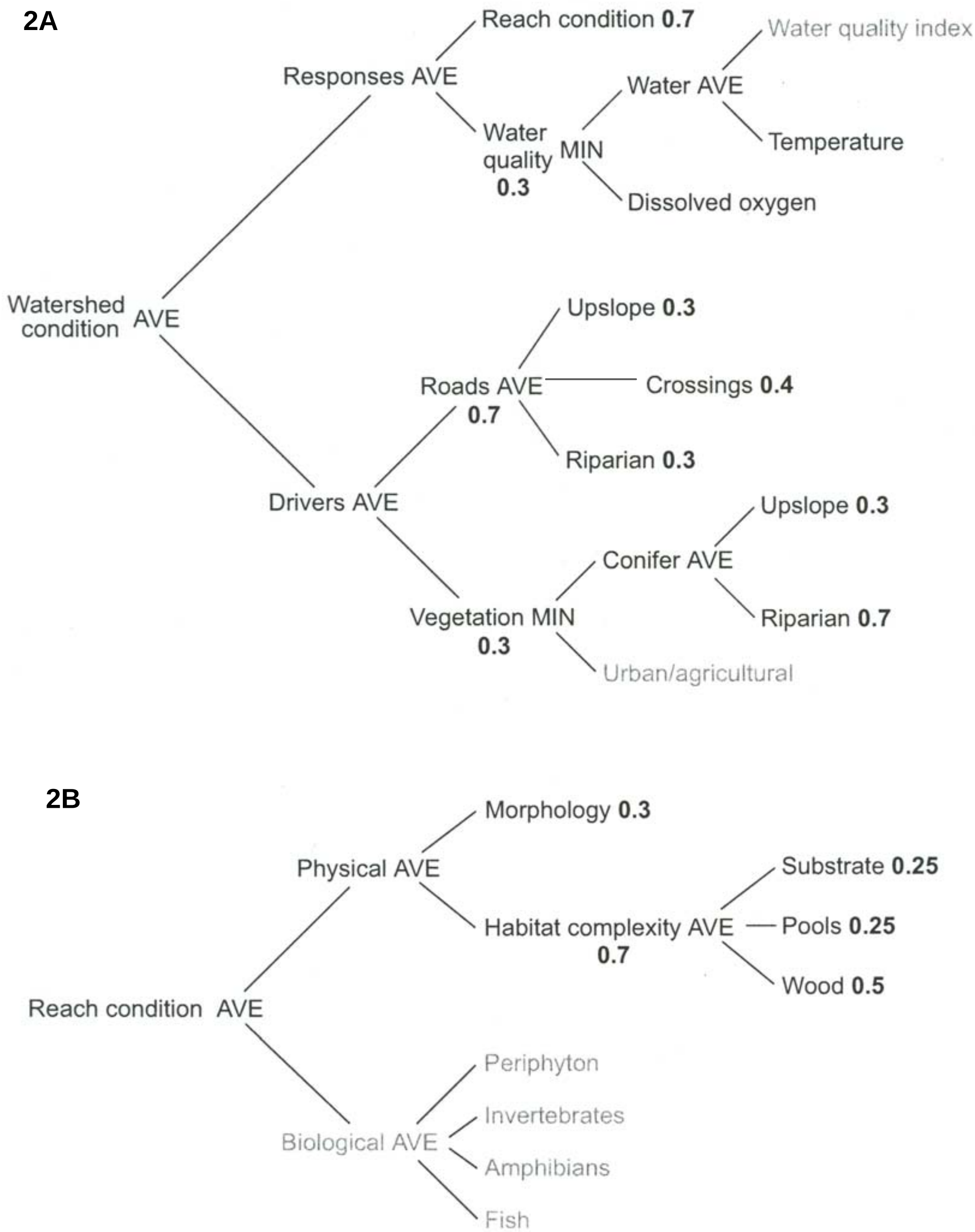

Figure 2. Watershed (Fig. 2A) and reach-scale (Fig. 2B) decision-support models used in Gallo *et al.* (2005; Figures 65 and 66). The models hierarchically aggregate a number of attributes into broader indices of reach and watershed condition. This aggregation of scores helps to determine whether a situation type is a “limiting factor” (worst condition score determines the combined score), “partially compensatory” (scores are counted equally), or “fully compensatory” (best condition scores determines the combined scores).

Table 2. Example of watershed evaluation criteria used in a decision-support model. “Data value” and “Evaluation score” columns show how the raw data from a watershed corresponds to the evaluated attribute scores (See figures 2A and 2B). The “Curve shape” column gives a graphical depiction of the relationship between the data value and evaluation score. The “Source” column gives the basis on which the curve was constructed. Table was reproduced from Gallo *et al.* (2005; Table 13).

Attribute and measure	Data value and node x-value	Evaluation score and node y-value	Curve shape	Source of criteria
Watershed model attributes				
High-slope road density (mi road/mi^2 watershed)	0 0.5	1 -1		AREMP[a] workshop 05/22/03
Upslope road density (mi road/mi^2 watershed)	0 4	1 -1		Dose and Roper 1994 AREMP workshop 05/22/03
Riparian road density (mi road/mi stream)—164-ft buffer	0 0.1	1 -1		AREMP workshop 07/01/04
Road crossing frequency (number of crossings/mi stream)	0 1.75	1 -1		AREMP workshop 07/01/04 Willamette NF data
Upslope vegetation—small-conifer cover (percentage of area with conifers <10 in d.b.h.)	10 40	1 -1		AREMP workshop 05/22/03
Riparian vegetation, large-conifer cover—197-ft buffer (percentage of area with conifers ≥20 in d.b.h.)	60 100	-1 1		AREMP workshop 07/01/04 d.b.h. from wildlife handbook Dose and Roper 1994 harvest roads vs. condition >30 percent of watershed impacted
Water temperature (maximum 7-day average, °C) Bull trout present	16 18 23	1 0 -1		AREMP workshop 05/22/03
Dissolved oxygen				
Percentage of saturation	<50 ≥50	-1 1		AREMP workshop 05/22/03
Mg/L	<4 ≥4	-1 1		
Reach model attributes				
Morphology				
Slope Entrenchment ratio Sinuosity Bankfull width:depth	Use to determine Rosgen stream type If D, F, G channel then -1, otherwise +1			Professional judgment
Pool frequency (number of bankfull widths per pool)				

Determining Monitoring Questions

Composing well-articulated monitoring questions and goals is a significant but often difficult task. The consequences of fuzzy goals in monitoring programs can range from disagreement about what data mean to monitoring systems that fail to provide information needed for resolution of management and policy issues. Both the Northwest Forest Plan monitoring effort (Haynes *et al.* 2006) and CALFED program (CALFED 2006a) suffered in part from lack of clear, quantifiable criteria, progress benchmarks, and consequent uncertainty about the causes and significance of observed trends.

A systematic and thoughtful approach to determining questions and goals begins with agreeing on the basic intent of monitoring. Two fundamental questions that *must* be answered at the outset are:

1. Is the monitoring intended to be merely descriptive, or is it necessary to compare monitoring results to some benchmark or target?
2. If comparisons are to be made, will they be against a single threshold value, or a suite of desirable states that may change over time?

The first question addresses a dichotomy between indicator frameworks that 1) use indicators in a *descriptive* or *predictive* way to track the status and trends of ecosystem features and 2) those that use indicators in an *adaptive* way to measure ecosystem structure, composition, or functional responses to management actions or stressors and compare them to stated goals or objectives (NRC 2000). In practice, indicators are often used for both purposes (Yoder *et al.* 2005). However, the distinction between these two uses is significant and should not be obscured in monitoring reports (GAO 2005). For effectiveness monitoring, the adaptive model is most likely to be used, since the fundamental questions revolve around the likely effects of restoration activities/programs, and the causes and implications of the observed changes.

The second question addresses a dichotomy in the use of benchmarks or targets against which the indicator results are judged. Wiersma (2005) differentiated between environmental and ecosystem benchmarks. ***Environmental benchmarks*** are threshold values for particular environmental phenomena which cause adverse effects when exceeded (e.g., Total Maximum Daily Loads for water quality parameters). In contrast, ***ecosystem benchmarks*** are desired conditions or states, often estimated from undisturbed reference sites to represent ‘healthy’ ecosystems (e.g., ecosystem characteristics that foster persistence of at-risk species). Environmental benchmarks are used most commonly in ecological risk assessments, while ecosystem benchmarks are more often found in indicator monitoring (Suter 2001). There are significant conceptual and analytical differences between assessing indicators against a threshold, versus against a suite of target conditions that may not be easy to define with a high degree of precision. Both types of benchmarks contribute to the effectiveness of monitoring and recovery programs.

Selecting Indicators

A scientifically credible indicator framework starts with a well-documented and defined protocol for selection. Appendix B of this report synthesizes guidelines for indicator selection from several sources. The US Environmental Protection Agency (USEPA 2000) published 15 guidelines which have been used extensively in the design of indicator frameworks across the US. They are grouped into four functional "phases" based on the following key criteria:

- Conceptual relevance - Is the indicator relevant to the monitoring question, and to the ecological resource or function being restored?
- Feasibility - Are the methods for sampling and measuring the environmental variables technically feasible, appropriate, and efficient for use in an effectiveness monitoring program?
- Response variability - Are errors of measurement and natural variability over time and space understood and documented?
- Interpretation and usefulness - Will the indicator convey information on ecological condition that is meaningful to monitoring restoration effectiveness?

In practice, USEPA's Environmental Monitoring and Assessment Program (EMAP) found that all aquatic biota assemblages it evaluated with the above criteria had very similar values. EMAP ended up making decisions on perceived societal value, responsiveness to disturbance, signal/noise ratio, information/cost ratio, and familiarity with a particular assemblage, sampling method, and analytical approach (Hughes 1993). Although those evaluations focused on aquatic biota assemblages, the same criteria were germane for selecting quantitative indicators.

Various approaches may be taken in choosing the initial suite of phenomena that are to be captured by indicators. Among the examples reviewed by IMST, a common approach to indicator selection was to first define ecosystem "health" or "integrity" in such a way that: (1) the relationship between the indicators and the parameters of the definition is clear, and (2) the endpoints (e.g., target conditions) which indicator monitoring can be evaluated against are quantifiable. Satisfactorily meeting these two criteria proved difficult in many of the cases reviewed by IMST (exceptions are Ohio EPA and USEPA's EMAP); nearly all authors identify the critical need for research to fill information gaps that plague this phase of indicator development. Dent *et al.* (2005) represents an attempt to resolve these issues for aquatic components in the Oregon Plan, and the USEPA has done so for streams of the conterminous US, using least disturbed regional reference sites, predictive models, and indices of biotic integrity (IBIs; See Appendix C of this report for a detailed discussion of IBIs; USEPA 2000; Stoddard *et al.* 2005; USEPA 2006).

Once selected, critical information related to each indicator needs to be well documented. Appendix A includes a good example of indicator documentation from the South Florida Ecosystem Restoration program. Comprehensive documentation would include (NRC 2000; Niemi & McDonald 2004; SFSCG 2006):

- How the indicator relates to and integrates the ecosystem attributes of interest,
- To what extent natural variation in indicator values can be distinguished from that caused by anthropogenic influences,

- The conceptual ecological model that underlies the indicator, including driver/stressor effects and historical impacts,
- The range of values the indicator can take, and the significance of those values,
- Goals, objectives, or desired future conditions for the indicator,
- The spatial extent over which the indicator occurs, and the ecological types for which it is applicable,
- Summary of relevant research about the indicator,
- Temporal and spatial scales over which the indicator is likely to change,
- Probable effects of new technologies on making the required measurements, and how soon significant technological changes are likely,
- Reliability of the indicator, and
- Cost-effectiveness.

Inferring Cause and Effect from Ecological Indicator Monitoring

One of the primary purposes of this IMST report is to assist OWEB and other Oregon Plan partners in making decisions about the use of indicators and other monitoring information in evaluating the effectiveness of restoration efforts. This involves making inferences about cause and effect when a high degree of uncertainty exists about the influence of uncontrolled and unmeasured factors that might influence indicator behavior. This emerged as one of the most problematic dimensions of indicator monitoring in IMST's review of the scientific literature. It is rooted in the fact that monitoring designs are inherently observational rather than controlled manipulative experiments. In observational studies many factors not included in the sampling design may affect the observed results.

Woolsey *et al.* (2007) illustrate a common approach to the treatment of the question of effectiveness; while they report significant correlation between restoration activities and desired outcomes in the Thur River system (Switzerland), there is little discussion of actual causality. Making the connection between restoration action and restoration success may be warranted in this example, but evidence of actual causation is not presented, resulting in a degree of uncertainty about why the observed results were obtained. This is often the case in effectiveness monitoring examples. Additionally, Cooperman *et al.* (2007) present a case study highlighting shortcomings that can result when monitoring designs are inadequate for determining restoration effectiveness.

It is not the intent of IMST to imply that indicator monitoring cannot be useful in evaluating restoration effectiveness. However, achieving this outcome requires detailed attention to both the design of monitoring programs and analysis of resulting data. Early and frequent consultation with statistical experts is a critical component of monitoring program design, particularly if restoration effectiveness will be assessed. This will help clarify what the monitoring results can and cannot imply about effectiveness.

One way to make effectiveness determinations is to start with sampling and analytical approaches that will address the necessary questions (e.g., Stoddard *et al.* 2005; Yoder *et al.* 2005; USEPA 2006). In this process it is essential to use a "weight of evidence" approach. For example, when the USEPA sought to determine causes, sources, and effects of acidic deposition across the US it used a probabilistic survey design coupled with selected mechanistic studies and multiple lines of evidence, then followed up by regulating sulfur dioxide emissions from the nation's power plants and incinerators, and measuring the results (Baker *et al.* 1991; Kaufmann *et al.* 1992). Similarly, Stoddard *et al.* (2005) and USEPA (2006) used a probability sampling design and a risk assessment model to associate major stressors with poor quality biological assemblages in streams of the western US and the conterminous US. Similar approaches have more recently been used to determine causes, sources and effects of global climate change. All include rigorously examining other likely stressors, causes, sources, and mechanisms. The USEPA has developed the Causal Analysis/Diagnosis Decision Information System (CADDIS)[2] for evaluating, weighting, and summarizing the types and quality of evidence for making assessments leading to regulatory and best management decisions. The types of evidence[3] considered and analyzed include spatial/temporal co-occurrence; evidence of exposure; biological mechanism; mechanistically plausible causes; causal pathways; stressor-response relationships field or laboratory studies and/or ecological model simulations; results from exposure manipulation at the site and other sites; verified predictions; analogous stressors; and the consistency and explanatory power of the evidence.

The CALFED Bay-Delta and Ohio EPA programs attempted to explicitly determine the effectiveness of management actions from indicator monitoring results. The expressed intent in the CALFED project is to use conceptual models, supplemented by extensive research, to explain driver/outcome relationships (CALFED 2006b). Other programs reviewed by IMST (including Chesapeake Bay and South Florida) appear to have implicit intentions of inferring effectiveness of restoration activities, but do not explicitly state this as a goal or propose analytical methods to determine cause-and-effect. However, water quality agencies in several states have developed indicators, biological criteria, rigorous monitoring methods, and sampling designs capable of assessing the results of water quality management actions, especially point-source and diffuse pollution controls (e.g., Stark 1985; Moss *et al.* 1987; Courtemanch 1995; Wright 1995; Yoder 1995; Klauda *et al.* 1998; Roth *et al.* 1998; Davies *et al.* 1999; Langdon 2001; Linham *et al.* 2002; Oberdorff *et al.* 2002; Wilton 2004; Ode *et al.* 2005; Yoder *et al.* 2005; Stark & Maxted 2007).

Indicator Frameworks

The term ***indicator framework*** refers to an organized hierarchy of ecological indicators used to evaluate a particular suite of monitoring questions. The sources reviewed by IMST for this report commonly identified four important features of indicator frameworks (Wefering *et al.* 2000; Dale & Beyeler 2001; GAO 2005; SFSCG 2006):

- Integration of indicators into a hierarchical system,

[2] Available at http://cfpub.epa.gov/caddis/index.cfm. Accessed on-line on September 6, 2007.
[3] Definitions and more information is available at http://cfpub.epa.gov/caddis/step.cfm?step=16 . Accessed on-line on September 6, 2007.

- Statistical independence among indicators (to avoid "double counting"),
- Spatially "binding" indicators by choosing those with a broad spatial extent,
- Choice of a suite of indicators that specifically address the complexities and interrelationships within the environment under study, and cover the key environmental gradients (e.g., soils, vegetation types, landforms) within the study area.

Development of an indicator framework that achieves these objectives is no small task, and likely requires building the framework from the ground up, rather than piecing together a variety of existing individual monitoring efforts. As discussed above, using conceptual models is also recommended. Several articles reviewed by IMST commented on the need for integration of the various *aspects* of indicator monitoring, in addition to linking the indicators themselves. For example, Bolgrien *et al.* (2005) identified the need for assessment frameworks that explicitly align issues of scale and indicator selection with underlying questions, and link components of opportunistic site monitoring, probabilistic sampling, and modeling.

Conclusions

Despite the many challenges discussed above, a number of institutions have developed and used quantitative ecological indicators, especially those related to aquatic biota assemblages. These range from coho salmon abundance, distribution and productivity used in western Oregon (ODFW 2007) to IBIs and predictive models used by USEPA (USEPA 2000; Stoddard *et al.* 2005; USEPA 2006; Whittier *et al.* 2007a; Appendix C of this report), the European Union (Pont *et al.* 2006), Canada (Steedman 1988; Bailey *et al.* 1998), New Zealand (Stark 1985; Joy & Death 2002; Stark & Maxted 2007), Great Britain (Moss *et al.* 1987), France (Oberdorff *et al.* 2002), Australia (Wright 1995), and several USA states including Ohio (Yoder *et al.* 2005), Maryland (Roth *et al.* 1998), California (Moyle & Randall 1998; Ode *et al.* 2005), Montana (Bramblett *et al.* 2005), Idaho (Mebane *et al.* 2003), Vermont (Langdon 2001), Texas (Linham 2002), Iowa (Wilton 2004), Wisconsin (Lyons *et al.* 1996, 2001), Minnesota (Mundahl & Simon 1999; Niemela *et al.* 1999); and New England states (Halliwell *et al.* 1999). In most of the above cases, the stressors resulted from anthropogenic disturbances, both point (industries, municipalities) and diffuse (agriculture, mining) sources. When these stressors were mitigated, indicators demonstrated clear changes in the one or two assemblages evaluated. The indicators used in the programs listed above may be less responsive to small-scale and local restoration of physical habitat structure. However, Hughes *et al.* (2004) and Kaufmann & Hughes (2006) reported that an aquatic vertebrate IBI was responsive to stream bed stability, instream cover, water temperature, nutrients, riparian cover and complexity, fine substrates, road density, and riparian human disturbance in the Oregon and Washington Coast Range.

PART 2 - TECHNICAL AND ANALYTICAL CHALLENGES

Many technical and analytical challenges face those using indicators to assess restoration effectiveness. Left unaddressed, these issues can lead to procedural and statistical errors with undesirable consequences for the usefulness of the indicator program. The NRC (2000) found the following to be common problems inherent in the indicator approach:

- Indicators generally need to be expressed numerically, due to analytical requirements. When nominal or categorical attributes are used, ambiguities in scoring, and lack of clarity about what the indicator value (or comparisons between values) actually means, can result. When categorical systems (such as “red-yellow-green” or “high-medium-low”) are used, thresholds and breakpoints between the categories need to be carefully thought out and justified,
- Multi-metric (or integrative) indicators may combine unrelated measures; this can impair the ability to determine linkages and cause/effect relationships between single drivers/stressors and outcomes, and
- Indicators can be difficult to understand and interpret if their component parts are not explained.

In addition, Harwell *et al.* (1999) found many of the indicator projects they reviewed were lacking:

- A systematic framework derived from ecological principles and ecological risk management,
- Characterization of ecosystem integrity across spatial and temporal scales, organizational hierarchies, and ecosystem types, and
- Transferability to other systems.

In this section, IMST focuses on topics requiring careful consideration prior to development of an indicator framework. This discussion is not exhaustive, and readers may benefit from reviewing the literature cited in this report. Topics covered in this section include:

- Single versus multi-metric or multivariate indicators,
- Determining appropriate survey and plot-scale sampling designs,
- Temporal and spatial scale considerations,
- Spatial and temporal variability, and
- Use of statistical methods in ecological indicator analysis.

Single versus Multi-metric or Multivariate Indicators

Depending on the design and goals of a monitoring program, single-metric, multi-metric, and multivariate indicators present various tradeoffs in efficiency of use, ease of interpretation, and the understanding that can be gained about an ecosystem and its potential for management.

Single-metric, multi-metric, and multivariate indicators all generally accomplish the same goals, namely to (Niemi & McDonald 2004):

1. Reflect the biotic or abiotic state of the environment,
2. Reveal evidence for the impacts of environmental change, and
3. Indicate the diversity or condition of species, taxa, or assemblages within an area.

Single-metric Indicators

Single-metric indictors are specific abiotic (e.g., stream temperature) or biotic (e.g., the abundance of a species) measures used to represent ecosystem characteristics of interest (Hellawell 1986; Boothroyd 1999; Niemi & McDonald 2004). Their advantages (Boothroyd 1999; Niemi & McDonald 2004) include being relatively easy to identify and sample, easy to analyze and interpret as trend data, and readily comparable to regulatory thresholds (e.g., Total Maximum Daily Loads).

Multi-metric and Multivariate Indicators

Multi-metric indicators consist of aggregated information from multiple variables assumed to represent ecological features of interest (Karr 1981; Karr *et al.* 1986; Resh *et al.* 1995). Because they are integrative, these indices may mask variation in the component variables, while simultaneously revealing cumulative effects of multiple stressors (Boothroyd 1999; Niemi & McDonald 2004). The latter is particularly important because assemblages rarely experience a single stressor. Multi-metric indicators are popular because they integrate ecological information into a single score that is compared against a reference score. One of the most widely known multi-metric indices is the IBI first proposed by Karr (1981). IBI (more appropriately considered a family of indices that vary regionally) typically is used to compare aquatic community data from a site of monitoring interest with data from a reference site. The index itself is expressed as the deviation of the monitored site from the reference site (Niemi & McDonald 2004), with larger IBI values reflecting higher similarity between sites. Appendix C of this report contains additional information about IBIs.

Problems with multi-metric indicators can arise where there is dependence among the component variables, with consequent redundancies and compounding errors (Reynoldson *et al.* 1997; Boothroyd 1999). The redundancy concern can be resolved by eliminating candidate variables that exceed a predetermined correlation threshold (Hughes *et al.* 1998; McCormick *et al.* 2001; Mebane *et al.* 2003; Hughes *et al.* 2004; Bramblett *et al.* 2005; Ode *et al.* 2005; Whittier *et al.* 2007a). Suter (2001) points out the benefits of multi-metric indicators that can be deconstructed, especially in risk assessment, where understanding the contributions of individual components is essential.

Multivariate indicators (Boothroyd 1999) differ from multi-metric indicators in the way the index is derived. Multi-metric indices are produced by aggregating data across component variables, while multivariate indices are produced by subjecting data to a variety of ordination or classification techniques, such as principal components analysis (O'Connor *et al.* 2000), canonical correlation (Kingston *et al.* 1992), detrended correspondence analysis, or combinations of these techniques (Dufrene & Legendre 1997). Multivariate methods are focused on determining "expected" index scores (e.g., for IBI), based on observed environmental variables. The expected scores are obtained by sampling both predictor and response variables across broad gradients and using multivariate techniques to discern relationships in the data. Multivariate indicators are used more commonly in Europe, but are beginning to be used in North America as well (Reynoldson *et al.*1997; Hawkins *et al.* 2000; Bailey *et al.* 2004). Recently, multivariate multi-metric (IBI) models have been used (Oberdorff *et al.* 2001, 2002; Pont *et al.* 2006; Tejerina-Garro *et al.* 2006; Pont *et al.* In Review).

Multivariate indicators are attractive for restoration effectiveness evaluation because they require no prior assumptions either in creating groups out of reference sites (other than whether those reference sites are appropriate) or in comparing test sites with reference groups; they simply reveal patterns, gradients, or trends in data (Reynoldson *et al.* 1997). Nevertheless, multivariate methods also have limitations (Gerritsen, 1995; Norris 1995; Boothroyd 1999; Stockwell & Peterson 2002). These include:

- They are not easily understood, interpreted and applied by non-specialists,
- A confusing range of available statistical techniques,
- A lack of consensus on the most reliable approaches in different circumstances,
- An assumption that relevant environmental variables are being measured and that the relationships measured are indeed causal,
- An intolerance of missing data, and
- A requirement for a large set of reference sites.

Multi-metric indices share many of these limitations, but perhaps to a lesser degree. Boothroyd (1999) compared the advantages and disadvantages of multi-metric and multivariate indicators (Table 3).

Table 3. Comparison of multi-metric versus multivariate indicators (adapted from Boothroyd 1999; pp 80–81)

Property	Multivariate Predictive Model	Multi-metric
Widely used throughout the world	+	+
Based on qualitative conceptual model	+	+
Requires consistent sampling & processing	+	+
Requires accurate & precise taxonomy	+	+
Reference sites provide baseline data	+	+
Differs by stream type or region	+	+
Synthesizes complex data into single number	+	+
Classifies & ranks sites	+	+
Detects temporal trends & spatial pattern	+	+
Based on statistical properties of data	+	+
Used in statistical significance testing	+	+
Understandable to non-biologists	+	+
Predicts taxa from environment	+	«
Retains useful information in several metrics		+
Metrics may be used to assess specific stressors		+
Can be developed from small data sets		+

«Recent predictive IBI models predict metric scores from environmental variables

Determining Appropriate Survey and Plot-scale Sampling Designs

Because the ability to detect trends in data is strongly influenced by the characteristics of the sampling scheme, monitoring restoration effectiveness requires careful consideration of sampling design. Larsen *et al.* (2004) give examples relevant to Pacific salmonid recovery efforts on this topic. There are four basic areas where critical sampling questions must be addressed:

- Establishing control/reference sites against which to evaluate results from treatment sites,
- Choice of sampling approach,
- Choosing an appropriate sample size, and
- Ensuring sample independence.

Reference Sites

Reference conditions against which monitoring results can be compared facilitate determining the magnitude of change and significance of trends in indicator data (NRC 2000). Reference site selection can be problematic and involve significant subjectivity. In the programs reviewed by IMST, reference sites that reflect natural conditions were viewed as a desirable source of baseline information. It is not our intent to present an exhaustive review of the literature on selection of reference sites here, but a few principles are worth highlighting.

First, reference sites should be equivalent to treated sites (e.g., disturbed or restored) with regard to basic environmental characteristics and the potential to produce particular biotic assemblages. However, this can be difficult to achieve within any monitoring design (e.g., Cooperman et al. 2007). Second, the suite of reference sites should represent not just current conditions, but also the natural ecological variation inherent in the ecotype, including succession (Harwell *et al.*1999; NRC 2000). Third, it is virtually impossible to perfectly match reference sites to sites where it is desirable to monitor indicator status and trends. The uncertainties that arise from the inevitable imperfect pairings of reference and monitored sites need to be factored into analyses, interpretation, and presentation of results, to avoid reaching misleading or false conclusions.

For high-level indicator monitoring, which typically occurs at larger scales, reference conditions may be characterized over much larger areas; individual reference *sites* give way to sets of reference sites, reference *watersheds* or *landscapes*. The problems described in the preceding paragraph are compounded when indicator monitoring occurs at larger scales. Aquatic ecologists often use eco-regional reference sites of differing slopes and sizes and assume that a range of naturally varying stream systems are represented (Hughes 1995; Stoddard *et al.* 2006; Whittier *et al.* 2006). Many practitioners advocate the use of retrospective studies that allow placement of current conditions within the context of historical or natural conditions. Consequently, comparisons are made within the temporal dimension, rather than spatial. Key assumptions of this approach may include that "historical" and "natural" are roughly equivalent, that there is some characteristic range of variation that is intrinsic to an ecological type, that adequate historical data are available, and that ecological succession will progress in the same way as in the past. Determining the validity of these assumptions is an important step in deciding how useful this approach may be. Confusion may result from the use of historical range of variation (HRV) as reference conditions if it is also inferred that "getting back to" HRV is the goal of restoration. Thus, the relationship between HRV and desired goals (e.g., "ecological integrity") needs to be carefully stated. An instructive example of this is found in Woolsey *et al.* (2007), where a "guiding image" was developed from a combination of historical data and conceptual modeling.

Many persons assume that reference sites represent only relatively undisturbed or natural conditions. Other kinds of "references" can also be very useful, especially where pristine reference sites are not available (which is frequently the case given global climate change and air pollution), for example, minimally disturbed, historical, least disturbed, and best attainable conditions (Stoddard *et al.* 2006). Andreasen *et al.* (2001) suggest "degraded sites" be used to define worst-case, socially unacceptable conditions, and these have been used in developing recent IBIs (Stoddard *et al.* 2005; Whittier *et al.* 2007b; Pont *et al.* In Review). "Desired future conditions" is a concept commonly used to define a desired endpoint when pristine reference sites are unavailable, or pristine conditions are not the goal of management.

Choice of Sampling Approaches

There is widespread agreement in the literature that probabilistic sampling is the best way to obtain unbiased assessments of ecological conditions in indicator monitoring. Systematic random designs can produce accurate, precise and unbiased information and allow calculation of confidence limits that cannot be determined from non-random designs. Moreover, probabilistic designs appear more efficient at quantifying changes in ecological conditions due to anthropogenic activities than non-random designs (Hughes *et al.* 2000). In a comparison of random versus non-random sample designs in surveys of spawning escapement of coastal coho in Oregon, Jacobs & Cooney (1995) found that use of non-random survey sites led to a three- to five-fold overestimation of coho abundance. Urquhart *et al.* (1998) and Hughes *et al.* (2000) call for the use of probabilistic or systematic random sampling such as that used in the USEPA's EMAP program. The same authors warn that focusing only on highly disturbed sites versus undisturbed sites provides biased assessments of ecological condition, and subjective site selection hinders evaluation of characteristic conditions (Urquhart *et al.* 1998; Urquhart & Kincaid 1999; Hughes *et al.* 2000).

Suter (2001) elaborates on some of the complexities of sample design in monitoring relevant to the points above. He points out that random designs tend to characterize high frequency/low severity situations well, but are prone to overlook low frequency/high severity events – i.e., they are unbiased relative to space (work well with spatially distributed phenomena like land use), but biased relative to risk (work less well with rare point-source phenomena). He suggests two solutions: 1) determine the distribution of effects of various stressors and sample accordingly, or 2) combine probabilistic and purposive sampling. The latter approach was used in tracking large numbers of rare or difficult to detect terrestrial species (e.g., fungi, lichens, and bryophytes) under the Northwest Forest Plan (Cutler *et al*. 2002; Molina *et al.* 2003).

Additionally, two other key aspects of trend monitoring are consistency and longevity, that is, the extent to which monitoring can be implemented in a consistent manner over multiple decades. Thus, issues related to funding, institutional commitment, replication of protocols, data management, and training of data collectors need to be addressed.

Temporal and Spatial Scale Considerations

If indicators are to be used effectively, it is necessary to understand the temporal and spatial scale at which ecological characteristics of interest exhibit variation. This will facilitate matching the scale of management actions and monitoring efforts to the scale of processes and outcomes of concern (Niemi & McDonald 2004; Niemi *et al.* 2004). Without this understanding of spatial and temporal scales, it is difficult to differentiate measurement error from changing ecological condition, or signals from unmeasured variables driving processes at different scales.

The direction and magnitude in which aquatic habitat and biotic assemblages respond to physical and biological processes depends on the spatial and temporal scales (reviewed in Rieman *et al.* 2006) over which processes operate. For example, intense episodic disturbances (e.g., floods, landslides) can cause dramatic fine-to-mid-scale habitat alterations within a watershed. However, at the overall watershed scale, the effects of such localized events may be more difficult to detect or seem less significant. Conversely, the significance of stressors operating at larger scales (e.g., regional droughts driven by climate cycles) may only be detected or understood by looking at broader scales. The timeframes over which ecosystems might be expected to recover from disturbance events that vary in geographic extent and intensity also differ.

The spatial and temporal scales of biological responses to environmental stressors also vary considerably depending on the biological level of organization (e.g., individuals, populations, assemblages), spatial requirements (e.g., home ranges) of species involved, and the time needed for recovery (Rieman *et al.* 2006). For example, linking site-scale habitat restoration effects to population-level responses of anadromous fish requires understanding both the scale of physical processes affected by restoration and the timeframe required for a population-level response.

Fausch *et al.* (1994) and Dunham *et al.* (1999) found that salmonid fish distributions were best explained by climatic gradients at larger scales while biotic interactions (e.g., competition, predation) were more strongly related to distributions at smaller scales. However, even at small scales, strong natural gradients in elevation, channel gradient, or geology, must be accounted for before one can detect the effects of management (Cooperman *et al*. 2007). Where there are strong human disturbance gradients, land use is often more important in explaining assemblage condition than are local habitat conditions. However, if an entire region is highly disturbed there may be no clear land use gradient, and local factors may appear more important, when in fact this is not the case. Wang *et al.* (2006) contains a review of research on how spatial scale affects stream habitat and aquatic biota, and concludes differences in observed biotic responses are primarily a function of scale, disturbance intensity, and statistical methods.

Given the diversity of environmental stressors, restoration methods, and ecological responses, addressing issues of scale in effectiveness monitoring requires at least two steps. First, relevant processes and their scale of operation must be clearly defined (this can be facilitated by the use of conceptual models). Second, monitoring efforts must incorporate sampling designs capable of detecting spatial or temporal patterning across the range of scales believed to be important (Rieman *et al.* 2006).

Spatial and Temporal Variability

All monitoring involves making repeated measurements at different times and places so that treatments can be compared, and trends be discerned. To minimize erroneous conclusions, sources of variability (including within-season variation for biotic indicators, as well as longer-term variation; USEPA 2000) as they pertain to sampling frequency in time and space, must be addressed. Random variability in characteristics like survival, population size, habitat selection, or habitat condition is inherent in all ecological systems. The minimum sampling extent and interval (both spatial and temporal) required to both, characterize an ecological indicator, and detect among-site differences depend on the level of random variability inherent in the system of interest.

Andreasen *et al.* (2001) encourage the use of pilot studies to determine:

- Sensitivity to changing conditions (levels of stressors or management activities), because indicators ideally reveal statistically significant differences between treatments,
- Natural variation in indicator values, because high inherent variability in indicators may mask responses to factors of interest, and
- Statistical independence, because if two measures are positively correlated they measure the same thing or, if measures are negatively correlated, positive change in one variable masks negative change in the other.

To control for spatial variability, indicator responses to various environmental conditions must be consistent across the monitoring region (USEPA 2000). If spatial variability occurs due to regional differences in habitat, it is necessary to normalize the indicator across the region, or to divide the reporting area into more homogeneous units. Similarly, with regard to within-season temporal variability, indicators should be sampled within a particular season, time of day, or other window of opportunity when their signals are determined to be strong, stable, and reliable, or when stressor influences are expected to be greatest (USEPA 2000). To control for temporal variability across years, monitoring should proceed for several years at sites known to have remained in the same ecological condition. Death (1999) suggests that calculations of both autocorrelation (dependence of a variable on its past values) and cross-correlation (dependence of a variable on the past values of another variable) can reveal non-linear patterns (e.g., seasonal), or a lack of them.

The NRC (2000) describes methods to account for the natural variation in oscillating systems (e.g., floodplain inundations and ocean conditions) in trend detection monitoring. They identify three basic tasks:

- Identify surrogates to characterize oscillations,
- Determine optimal sampling frequency from these surrogates, and
- Detect changes from data.

Simulation models and long-term paleoecology records (e.g., dendrochronology, pollen and charcoal analyses) can help approximate oscillations for current systems, in the absence of actual data, although this presumes that the past is a good indicator of the future. NRC (2000) also describes an approach for determining sampling frequency using such methods.

Independence of Samples

Generally the more replicates (sites at which independent application of the same treatment occur) used, the greater the statistical precision of the effectiveness monitoring. Sample design, including replication of treatment and control monitoring, can be time consuming and costly, and it is desirable to establish the least number of samples that will give the desired amount of precision and statistical power, given the variation in the parameter of interest.

Lack of sample independence can lead to ***psuedoreplication*** (Hurlbert 1984; Heffner *et al.* 1996; Death 1999; Miller & Anderson 2004), a common error in ecological studies (Hurlbert 1984; Heffner *et al.* 1996). When samples are pseudoreplicated, the natural random variation exhibited by a variable is not properly quantified (Miller & Anderson 2004). For example, repeated sampling of fish abundance from the same stream reach does not reflect the variation inherent in the stream system as a whole. Randomly drawing samples from different stream reaches experiencing the same stressor intensity or treatment would allow more accurate estimation of variability in the fish abundance response, although there is debate in the literature about whether this truly eliminates pseudoreplication (McGarvey & Hughes In Press). Random selection of sites along a reach (e.g., Omernik & Griffith 1991), versus ad hoc site selection can reduce pseudoreplication, but all sites are still not truly independent.

Pseudoreplicated samples appear larger in size than they truly are, giving the illusion of statistical power where little exists. In this situation, inferential statistics must be used with great care because many tests are designed for samples of independent observations. Inaccuracies are typically manifested in biased standard errors that misrepresent (typically underestimate) variation in the data and artificially inflate the significance of statistical comparisons.

Use of Statistical Methods in Ecological Indicator Analysis

A wide range of statistical methods are available for indicator data analysis. This section touches briefly on some primary considerations, and is not intended to be an exhaustive treatment of this subject. Consultation with a statistician, early and often, in the development and use of indicators is highly recommended.

Data Aggregation and Integration

Many regional indicator monitoring initiatives are based on data collected by multiple entities using different sampling designs and methods. If not coordinated and executed with great care, these efforts can reduce indicator data accuracy and precision as well as reduce how well the data represent the ecological characteristic of interest (Hughes *et al.* 2000). Often indicator monitoring is done by public agencies with varying mandates and information needs, along with organizations of volunteers or part-time data collectors. In restoration effectiveness monitoring, data are commonly collected project-by-project using different sampling designs, a mix of qualitative and quantitative variables, different measurement periods, varying or absent quality assurance/control practices, etc. Even in the case of monitoring for condition indicators at larger scales, inconsistent sampling and data management are the rule. Therefore, a significant challenge to indicator monitoring is the need to aggregate, integrate and combine data sets from diverse sources.

Data aggregation can require several steps (Rice & Rochet 2005):

- Screen data against a set of bottom-line acceptability criteria,
- Standardize data to bring them into comparable scales prior to aggregation,
- Weight data so that their significance, for example based on quality and/or relevance, is reflected in the aggregate data score, or
- Combine weighted standardized indicators.

Methods for integrating diverse metrics include use of the arithmetic mean, weighted average, graphic displays, and multivariate statistics (Andreasen *et al.* 2001; O'Connor *et al*. 2004). Gates (2002) discusses the use of meta-analysis techniques in aggregating data from different ecological studies. Paquette *et al.* (2006) is an example of a meta-analysis that included international forest research to underplant tree seedlings in managed forests. Rice & Rochet (2005) review a wide range of approaches to data aggregation and integration for indicator development (See Appendix D of this report).

Astin (2006) describes a framework used to merge water quality data from multiple sources in the Potomac River basin (eastern US). Issues addressed in that article include:

- Resolution of data formats and file types into a common system,

- Combinations of study designs, including ad hoc vs. probabilistic sampling frameworks,
- Resolution of diverse measurement periods,
- Differences in quality assurance among sampling entities,
- Variations in subsample sizes and data collection protocols,
- Use of varying taxonomic levels, and
- Resolution of different methods for measuring abundance (e.g., abundance classes with different ranges vs. actual counts).

In general, approaches of this type involve identifying commonalities in data sets and protocols (including protocols for sampling), and selecting a set of core attributes for analysis. By necessity, this results in significant generalization of data to the least specific level, and exclusion of sometimes significant quantities of data that do not fall within the "overlap". Significant gaps in spatial coverage may result from this process. Although Astin (2006) asserts a high degree of rigor for this methodology, the process involved making a number of assumptions about data definitions and quality that did not appear to have been sufficiently evaluated (again, by necessity; information was not available to do so). Combining data sets can create substantial statistical problems, and the oversight of a statistician is necessary if such efforts are undertaken. The alternative approach of building a comprehensive, hierarchical, statistically rigorous indicator sampling system, such as USEPA's EMAP has been shown to be more cost-effective in some cases (Larsen 1995; Hughes *et al.* 2000).

The CALFED Bay-Delta Science Consortium has developed a framework for managing natural resources data collected by a variety of entities involved in the CALFED indicators effort[4]. This approach is built around the following principles:

- A configuration that allows reliable replication and updating,
- Flexibility of data platforms and formats,
- Ability to accommodate users with a variety of system capacities,
- Accessibility for user groups (data providers, aggregators, and users),
- Version control by those closest to the data,
- Adequate documentation (including extensive metadata and peer-review)
- Use of common or translatable vocabularies, and
- Known validation levels.

Discerning Significant Differences in Indicator Data

Parametric (e.g., t-test or analysis of variance) or nonparametric tests (e.g., Kruskall-Wallis or Mann-Whitney U tests) can be used to assess differences between control and treatment sites assuming that an appropriate number of control and treatment replicates have been measured (Ramsey & Schafer 2002). Parametric tests perform better when sample sizes are similar and

[4] The CALFED Data Management framework can be viewed at http://science.calwater.ca.gov/pdf/bdsc_data_system.pdf. Accessed on-line September 10, 2007.

assumptions about the data distribution are met. Non-parametric tests are generally preferable when assumptions about data distribution are not met because they typically involve recoding data in ways that rectify distribution problems. However, nonparametric tests typically reduce the ability to detect significant differences and their use requires careful consideration of the consequences of Type I and Type II statistical errors (see below). These techniques are widely described in the literature dealing with statistical techniques in ecology. Note that both parametric and non-parametric tests assume data collected via independent random sampling.

Despite the wealth of literature and statistical packages, ecological researchers still make basic errors in statistical technique selection and use. Two common errors are (Death 1999):

- Use of a Model I Analysis of Variance (ANOVA) test when a Model II or Model III ANOVA should be used.
- Choosing a probability level without regard to Type I or Type II statistical errors.

A Model I or fixed-effects model in ANOVA should be applied in situations where the experimenter assumes all possible occurrences of a treatment are being considered in the analysis (e.g., all years in a flow record). A Model II (random effects) ANOVA is used when occurrences and/or treatments are randomly selected from a larger population. When both fixed and random effects are present a Model III ANOVA may be most appropriate. Use of these alternate models can have quite different statistical consequences, and it is important that the right statistical test is selected for a given situation (Death 1999; USEPA 2000; Ramsey & Schafer 2002).

At the heart of effectiveness monitoring is the hope that statistical analysis of the monitoring indicators will demonstrate whether or not the restoration intervention had a positive effect. In most cases this means that there is a statistically significant difference between the control and treated sites. A Type I error (false positive) occurs when the analyst concludes the restoration method is working, when in fact treatment differences should be attributed to chance. A Type II error (false negative) occurs when the analyst concludes the restoration treatment is not effective, when in fact it is. In laboratory and large agricultural field studies that can be repeated relatively easily and where research results do not produce winners and losers, scientists are most concerned with avoiding Type I errors. In contrast, restoration professionals may be more interested in avoiding Type II errors. Type II errors (i.e., determining “no significant impact” when in fact there is one) can be extremely serious in indicator monitoring that is assessing potentially detrimental impacts to the environment (Andreasen *et al.* 2001). Likewise, determining no significant improvement from mitigation can be a serious Type II error, if such hypothesized improvements show no positive response because of flawed study designs or indicators (Cooperman *et al*. 2007), and the mitigation is discontinued. When future restoration funding may be contingent on demonstrating success, the choice of probability level can be significant, and should be thought through carefully. Increasing the sample size will reduce the probability of a Type II error. McGarvey (2007) also suggested reducing the probability of Type II error by increasing the probability of Type I error (e.g., by increasing error tolerance (α) from 0.05 to 0.10), increasing sample size, choosing the most responsive indicators, and using equivalence tests in which the burden of proof is switched by making the effect of concern the null hypothesis and the “no effect” result the alternative hypothesis.

Limiting Ecological Factors and Statistical Analysis

To put monitoring results within the context of ecological reality, statistical analyses must take into account the influence of limiting factors (Scharf *et al.* 1998; Cade *et al.* 1999). A limiting factor can be viewed as the essential environmental factor (e.g., water, temperature, light) among all those affecting growth, reproduction and survival of an organism that is most constrained, and thus controls these important biological processes (Cade *et al.* 1999; Dunham *et al.* 2002). Because limiting factors constrain organisms on many levels, they can exert tight controls over the possible range of biotic responses that can be measured using indicators (Cade & Noon 2003). Consequently, interpretation of monitoring data requires determining which limiting factors are controlling the distribution of the indicator response.

Most statistical regression and correlation methods estimate changes in the mean of a response variable in relation to one or several predictor variables (e.g., environmental stressors, restoration treatments, etc.). This approach has limited usefulness for estimating, detecting or testing relationships involving multiple limiting factors (Scharf *et al.* 1998; Cade *et al.* 1999; Hiddink & Kaiser 2005) and can mislead data interpretations when limiting factors constrain the data distribution (Cade & Noon 2003). Relationships between limiting factors and ecological responses may result in data patterns with distinct upper or lower constraints (Dunham *et al.* 2002). To address this attribute of ecological data, a range of alternative methods has been proposed. For example, Scharf *et al.* (1998) and Cade *et al.* (1999) found that quantile regression techniques based on absolute least value models were better at detecting relationships at upper and lower bounds than the more commonly used ordinary least squares method.

Hiddink & Kaiser (2005) in a review of the practical implications of limiting factors note that limiting factors affect variation at the upper boundary of a sample distribution, while below this upper boundary variation can be influenced by many other factors. From this they draw two significant conclusions:

1. Ecological indicators such as abundance, biomass, and biodiversity can identify sites in good condition but are far less useful in detecting sites affected by environmental degradation because these sites can have both high and low abundances.

2. Using ecological indicators to examine spatial patterns in the impact of an environmental factor is problematic.

Hiddink & Kaiser (2005) suggest, but do not provide supporting evidence, that the variation in limiting factors is greater in space than in time. They conclude that ecological indicators are better suited for repeat sampling or monitoring at a specific site than for measuring spatial variability. Yoder *et al.* (2005) provide an example of these concepts applied in the state of Ohio. The USEPA’s EMAP program has dealt with some of these issues by focusing sampling on index periods (to reduce temporal variability), by using indicators not based solely on abundance, and by factoring in natural spatial differences (Stoddard *et al.* 2005).

PART 3 – COMMUNICATING INFORMATION ABOUT ECOLOGICAL INDICATORS AND STATUS AND TRENDS OF ECOSYSTEM HEALTH

Relaying the results of indicator monitoring to various audiences is a key feature of any monitoring effort. Different indicators may be needed to inform the scientific community, policy makers, and the general public, and these groups may differ on the amount of detail they expect about how indicators were collected and analyzed. All groups have a need to understand how indicators connect with predetermined program and/or monitoring goals and objectives.

Ideally, ecological indicators represent knowledge about the ecosystem and natural processes, are selected through best available procedures, are easily interpreted, and have the potential to promote 'competence' in public participation (Chess *et al.* 2005). Turnhout *et al.* (2007) described one perception of the relationship between science and policy as a transfer of knowledge. Within that perception "*...ecological indicators arrange the transfer of scientific knowledge by selecting, integrating, and translating scientific knowledge into usable knowledge for policy*" (Turnhout *et al.* 2007; p. 220). Indicators are also a way to benchmark policy performance and set a framework for reporting to a wider stakeholder community on the benefits and costs of a policy (Hertin *et al.* 2001). This section focuses on communicating indicator status and trends to decision-makers, natural resource professionals, and the public-at-large.

General Principles of Effective Communication

The selection of indicators depends on technical and communicative criteria. Gray & Wiedemann (1999) listed communicative criteria: truthfulness, informativeness, clarity, relevance, and resonance. Based on four of these criteria (clarity, truthfulness, informativeness, and relevance), Chess *et al.* (2005) conducted a qualitative study in New Jersey to determine which types of indicators effectively communicated to members of intermediary groups (those who disseminate government information to lay people, e.g., journalists, legislative staff, environmental advocacy groups) and the general public. Descriptions of the criteria (which are not completely independent of one another), findings from Chess *et al.* (2005), and examples follow:

1. <u>Clarity</u> encompasses the visual display of data, data presentation in numerical or qualitative terms, extent of complexity, and representation of uncertainty (Chess *et al.* 2005). Members of the intermediary groups included in Chess *et al.*'s study indicated that graphics and tables should be accompanied by explanatory text to avoid possible multiple interpretations when presentations are perceived as vague. They also suggested that overly technical writing and use of jargon and acronyms obscures the information provided by the use of indicators.

   For example, environmental factors or policy related events can easily be overlaid onto data graphs for presentation to the public and decision makers. When this is done, the composite figure(s) need to be reviewed with the above principles in mind to determine if the new figures are relaying accurate information. One such example is adding a line to salmon abundance or survival rates (Figures 3A and B) that indicates when the Oregon Plan was implemented as was done in OWEB (2005):

3A 3B

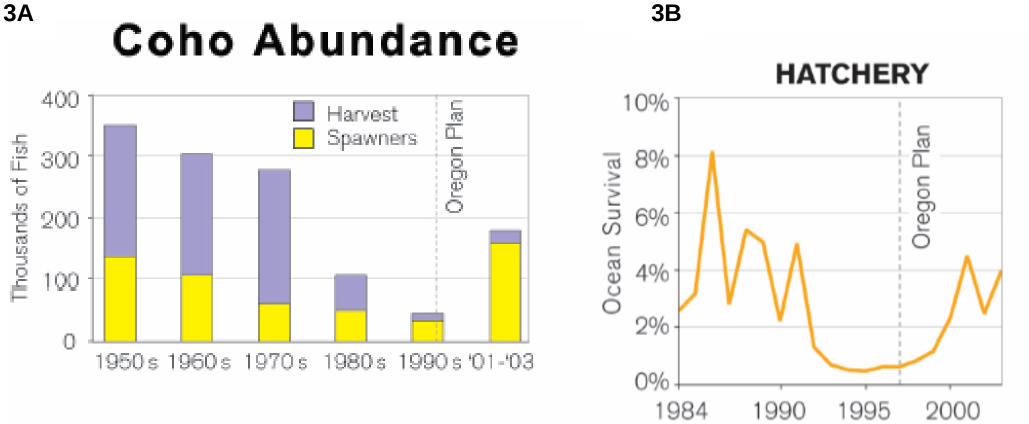

Figure 3. Example of composite figures. Figures were reproduced from OWEB (2005) with permission from the agency.

These figures could be misinterpreted to mean that the implementation of the Oregon Plan had a significant and almost immediate effect on coho abundance and hatchery fish survival in the ocean. Figure 3A, showing total coho abundance partitioned into commercially harvested fish and those that returned to spawn, suggests that the Oregon Plan had a strong influence on harvest rates and overall abundance. This figure does not include significant information about changes in major environmental (ocean conditions) or management (commercial harvest levels under Pacific Fishery Management Council) stressors that would allow a reader to accurately interpret the potential causes of abundance changes. Similarly, the published legend for Figure 3B (not shown) indicated that ocean conditions strongly influence hatchery fish survival, but did not include ocean condition information in the figure itself.

2. <u>Truthfulness</u> “implies that the indicator should not only be technically meaningful, but its presentation should be transparent and not manipulated in order to encourage ‘favorable’ perceptions” (Gary & Wiedemann 1999, p. 211). Intermediaries in Chess *et al.*’s (2005) study most often had criticisms related to this criterion of how indicator data and trends were displayed. The intermediaries noted that increasing transparency required information on data sources beyond simply citing the responsible agency. Indicating a specific individual, division or report as the source would make the information more ‘trustworthy’. In a GAO (2005) review of the Chesapeake Bay Program, panelists noted that the strength of the program depended on public perception of the Bay program’s integrity and if reports underwent an independent science review prior to publication, the public would have sufficient trust in the report so that other reports on the bay’s health would not be seen as needed. Chess *et al.* (2005) concluded that the truthfulness criterion presents particular difficulties because trust in the source of information can affect the perception of the information.

Intermediaries also noted that when there was a lack of clarity in indicators (including failure to explain data sources and providing contextual information) it was seen as a purposeful attempt to mislead and to portray environmental progress in overly optimistic terms. Intermediaries used terms such as 'spin' and 'propaganda" to characterize the use of indicators by government (Chess *et al.* 2005). One example from Chess *et al.* (2005), related the low percentage of severely impaired large rivers in New Jersey as indicated by macroinvertebrate indices. The low percentage (12%) could be a reflection of sampling design, not of improved water quality. Another intermediary pointed out that an increasing trend in apparent compliance with water quality laws by companies could reflect decreased oversight by the agency, not better compliance.

Likewise semantics such as simple word choice can also elicit misinterpretations of indicators and trends, such as the use of the word "restoration". The Chesapeake Bay program had this statement on its web site[5]:

"*Since 1996, Bay Program partners have been working to* <u>*restore*</u> [emphasis added] *riparian forest buffers throughout the watershed. Chesapeake 2000 set a goal of* <u>*restoring*</u> *2,010 miles of buffers by 2010. This goal was achieved eight years ahead of schedule in 2002*."

"Restoring" implies a return to a more natural state or function. In this case, what was actually achieved was 2,010 miles of buffers planted with trees and other vegetation. The goals of achieving functioning buffers that can shade surface waters, filter excess nutrients and sediments, and provide wildlife habitat will not be met for decades, therefore the use of "restored" could be seen as misleading or overly optimistic.

3. <u>Informativeness</u> relates to the extent to which the indicator meets the needs of the intended audience (Chess *et al.* 2005). Each major audience may require different information. For example, scientists may require sampling methodology, means with a measure of variation or confidence intervals, confidence limits, and scope of inference; journalists may need less detailed, technical information but more local information and how it may affect their readers; the general public may prefer knowing the bigger picture over time with a connection on how it may affect their health or recreation opportunities.

4. <u>Relevance</u> refers to the applicability of information to an individual, in contrast to informativeness that relates to a more generalized audience (Chess *et al.* 2005). An example of an indicator with relevance is the number of beach closures due to bacteria levels or shell fish closures resulting from domoic acid levels, rather than bar charts reporting bacterial or domoic acid levels. The indicators that may be most salient to public audiences may not provide the level of detail needed for scientific or intermediary audiences. but they still accurately portray trends and allow people to grasp the relevance to their own lives.

   From a cautionary standpoint, Chess *et al.* (2005) and Johnson & Chess (2006) concluded that public audiences may not always understand the limits on use of data. This can be

[5] http://www.chesapeakebay.net/forestbuff.htm. Accessed on February 7, 2007.

partially alleviated by describing how the information should and should not be used during communication with the public.

5. Resonance was defined by Gray & Wiedemann (1999) as the significance of an indicator for its intended audience. Resonant indicators represent those things most important to the public, such as being able to view a mountain on a regular basis, or the number of days sport fishing was allowed in the Columbia River for salmon. Qualitative resonant or motivating indicators often need to be balanced with quantitative, ecologically meaningful indicators.

   A well-referenced indicator that resonates with the general public is *Bernie Fowler's Sneaker Index*, a measure of water clarity in the Chesapeake Bay. Fowler, a former Maryland State Senator was quoted as saying "*...although this is not a scientific measure, it puts restoring the river on a human scale.*"[6] Compared to more technical data sources on turbidity levels (e.g., Secchi disk, nephelometer/turbidimeter), wading into the water to see when white shoes are no longer visible may be easier for members of the general audience at varying ages to understand and connect and therefore participate with local efforts.

Translating Technically Complex Indicators to a Common Language

Schiller *et al.* (2001) describe a process they used to translate indicators for ecological condition into common language to increase communication between agency staff and scientists with decision-makers and the public. This multi-step process includes moving away from describing what is measured by the indicators and how measurements are done, toward depicting the kinds of information that combinations of indicators provide about aspects of the environment valued by society. Using EMAP forest indicators as a case study, Schiller *et al.* (2001) found that respondents in their small group discussions (which did not include technical experts) preferred to be presented with information about the environment, and did not want or need descriptions of what was measured. They also found that there was often a mismatch between the details provided about individual indicators and the specificity with which the respondents were most comfortable.

Indicators and trends portrayed to general audiences need to be understandable, credible and useful, however, Johnson & Chess (2006) found that in general, people will accept indicator information they can comprehend, even though they don't necessarily agree with the results. Common language indicators provide a conceptual link between status and trends monitoring and formal ecological risk assessments by connecting the measured indicators with socially valued aspects of the environment. Schiller *et al.* (2001) concluded that without reporting mechanisms such as common language indicators, environmental information presented as discrete findings may be ignored by the general public and minimally used by decision makers regardless of scientific and environmental relevance.

---

[6] Quote from Chesapeake Bay Program's website at http://www.chesapeakebay.net/status/baycleanup.cfm?SUBJECTAREA=GET_INVOLVED, accessed on February 1, 2007.

Performance Measurements and Report Cards

Reporting the status and trends in ecosystem conditions and target populations can be done through various methods and can serve multiple roles. ”Report cards” are often used to describe progress made toward ecological goals, and to provide agency or program accountability via performance measures in reaching those goals. Report cards can summarize large amounts of complex information in a clear concise format. Performance measurements provide a means of communication that can help everyone involved in the process to think more strategically and to help citizens understand what value they receive for the money spent (Kettl 2001 as cited by Chess *et al.* 2005).

The format of report cards can vary based on the scope and objectives of a specific program. Effective reports contain information that is relevant, accurate, timely, consistent, thorough, precise, objective, transparent, and peer reviewed or verified (GAO 2005). The internet allows a unique opportunity to develop nested, hypertext reports that allow a reader to move from general information targeted to the general public to more detailed information on quantitative data, measurements, and analysis targeted to scientists (e.g., Canadian Environmental Sustainability Indicators 2006[7]; US and Canada's State of the Great Lakes 2005 Report[8]; Chesapeake Bay Trends and Indicators[9]). While various examples exist of how to display information, there are fewer examples of frameworks available of how to develop report cards that reflect the conceptual understanding of ecological principles and ecosystem integrity across temporal and spatial scales (Harwell *et al.* 1999).

Harwell *et al.* (1999) designed a fairly in-depth framework to develop ecosystem report cards that are linked to social values and scientific information. This is a two phase process; one working from top-down and the other working from bottom-up. The framework is based on five tiers that define the relationship between:

Tier 1. social <u>goals</u>,

Tier 2. <u>objectives</u> that disaggregate goals into more specific items but are still characterized in layman's terms,

Tier 3. <u>essential ecosystem characteristics</u> that capture relevant scientific information into a limited number of discreet characteristics that describe major ecosystem features,

Tier 4. <u>indicators</u> (endpoints) that are driven by scientific issues and social values that are defined by scientists and constitute the environmental attributes that need to be monitored to indicate status or trends and link back directly to one or more characteristics in Tier 3, and

Tier 5. <u>measures</u> determined by the scientists and used to collect data.

Harwell *et al.* (1999) also proposed several criteria for effective ecosystem health report cards. In essence a well design report card should:

---

[7] http://www.ec.gc.ca/environmentandresources/CESIFULL2006_e.cfm#5. Accessed February 5, 2007.
[8] http://binational.net/solec/sogl2005_e.html. Accessed February 5, 2007.
[9] http://www.chesapeakebay.net/indicators.htm. Accessed February 7, 2007.

1. be understandable to multiple audiences,
2. address differences in ecosystem responses across time (particularly in the context of natural variability),
3. show the status of the ecosystem,
4. characterize the selected ecosystem endpoints, and
5. transparently provide the scientific basis for the assigned grades (so that readers may define their own criteria and interpret trends).

Stressors (including measures to characterize the stressor, intensity, frequency, duration, and distribution) and ecological effects need to be monitored and evaluated in parallel to understand anthropogenic risks to the environment, and both should be included in the report card (Harwell *et al.* 1999). This allows monitoring, evaluation, and modification of management actions based on performance criteria.

REFERENCES

Andreasen JK, O'Neill RV, Noss R, Slosser NC (2001) Considerations for the development of a terrestrial index of biological integrity. *Ecological Indicators* 1: 21–35.

Astin LE (2006) Data synthesis and bioindicator development for nontidal streams in the interstate Potomac River basin, USA. *Ecological Indicators* 6: 664–685.

Bailey RC, Norris RH, Reynoldson TB (2004) *Bioassessment of Freshwater Ecosystems: Using the Reference Condition Approach*. Kluwer, Boston, MA.

Bailey RC, Kennedy MG, Dervish MZ, Taylor RM (1998) Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. *Freshwater Biology* 39: 765–774.

Baker LA, Herlihy AT, Kaufmann PR, Eilers JM (1991) Acidic lakes and streams in the United States: the role of acidic deposition. *Science* 252: 1151–1154.

Bernhardt ES, Sudduth EB, Palmer MA, Allan JD, Meyer JL, Alexander G, Follastad-Shah J, Hassett B, Jenkinson R, Lave R, Rumps J, Pagano L (2007) Restoring rivers one reach at a time: Results from a survey of US river restoration practitioners. *Restoration Ecology* 15(3): 482–493.

Bolgrien DW, Angradi RT, Schweiger EW, Kelly JR (2005) Contemplating the assessment of great river ecosystems. *Environmental Monitoring and Assessment* 103: 5–20.

Boothroyd IKG (1999) Macroinvertebrate monitoring: Alternative methods. In: *The Use of Macrovertebrates in Water Management: Recommendations of the New Zealand Macroinvertebrate Working Group* (ed Pyle E) New Zealand Ministry for the Environment, Wellington, New Zealand.

Bramblett RG, Johnson TR, Zale AV, Heggem DG (2005) Development and evaluation of a fish assemblage index of biotic integrity for northwestern Great Plains streams. *Transactions of the American Fisheries Society* 134: 624–640.

Bryce SA (2006) Development of a bird integrity index: Measuring avian response to disturbance in the Blue Mountains of Oregon, USA. *Environmental Management* 38: 470–486.

Bryce SA, Hughes RM, Kaufmann PR (2002) Development of a bird integrity index: using bird assemblages as indicators of riparian condition. *Environmental Management* 30: 294–310.

Cade BS, Noon BR (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and Environment* 1(8): 412–420.

Cade BS, Terrell JW, Schroeder RL (1999) Estimating effects of limiting factors with regression quantiles. *Ecology* 80: 311–323.

CALFED Bay-Delta Program (2006a) *Indicators and Performance Measures Phase I Report: Core Indicators and Plan (9/7/06 draft)*. CALFED Bay-Delta Program, http://science.calwater.ca.gov/pdf/monitoring/monitoring_attachment_1_phase_1_report_091906.pdf.

CALFED Bay-Delta Program (2006b) *Framework for indicators for science, management and adaptive management in the CALFED Bay-Delta Program (1/5/06 and 4/28/06 drafts)*. CALFED Bay-Delta Program, http://science.calwater.ca.gov/monitoring/monitoring_framework.shtml.

CALFED Bay-Delta Program (2006c) *Indicators and Performance Measures. Appendix to Phase 1 Report: Core Indicators and Plan.* CALFED Bay-Delta Program, http://science.calwater.ca.gov/pdf/monitoring/monitoring_attachment_2_phase_1_appendix_091906.pdf

Charvet S, Statzner B, Usseglio-Polatera P, Dumont B (2000) Traits of benthic macroinvertebrates in semi-natural French streams: an initial application to biomonitoring in Europe. *Freshwater Biology* 43: 277–296.

Chess C, Johnson BB, Gibson G (2005) Communicating about environmental indicators. *Journal of Risk Research* 8(1): 63–75.

Chesson J, Clayton H (1998) *A Framework for Assessing Fisheries with Respect to Ecologically Sustainable Development*. Report of the Bureau of Rural Sciences, Canberra.

Collie J, Gislason H, Vinther M (2001) Using AMOEBAs to integrate multispecies, multifleet fisheries advice. *ICES Document* CM 2001/T: 01.

Cooperman MS, Hinch SG, Bennett S, Branton MA, Galbraith RV, Quigley JT, Heise BA (2007) Streambank restoration effectiveness: lessons learned from a comparative study. *Fisheries* 32: 278–291.

Courtemanch DL (1995) Merging the science of biological monitoring with water resource management policy: criteria development. In: *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (eds Davis WS, Simon TP) Lewis Publishers, Boca Raton, FL.

Cutler R, Edwards TC Jr, Alegria J, McKenzie D (2002) *A Sample Design Framework for Survey and Manage Species Under the Northwest Forest Plan*. Proceedings of the Section on Statistics and Environment, 2001 Joint Statistical Meeting, American Statistical Association, Alexandria, VA.

Dale VH, Beyeler SC (2001) Challenges in the development and use of ecological indicators. *Ecological Indicators* 1: 3–10.

Davies SM, Tsomides L, DiFranco J, Courtemanch D (1999) *Biological Monitoring Retrospective: Fifteen Year Summary for Maine Rivers and Streams*. DEPLW1999-26. Maine Department of Environmental Protection, Augusta, ME.

Death RG (1999) Macroinvertebrate monitoring: statistical analysis. In: *The use of Macroinvertebrates in Water Management* (ed Winterbourn MJ) Ministry for the Environment, Wellington, New Zealand.

Dent L, Salwasser H, Achterman G (2005) *Environmental Indicators for the Oregon Plan for Salmon and Watersheds*. Institute for Natural Resources, Oregon State University, Corvallis, OR.

Dufrene M, Legendre P (1997) Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecological Monographs* 67(3): 345–366.

Dunham JB, Cade BS, Terrell JW (2002) Influences of spatial and temporal variation on fish–habitat relationships defined by regression quantiles. *Transactions of the American Fisheries Society* 131: 86–98.

Dunham J, Peacock M, Rieman B, Schroeter R, Vinyard G (1999) Local and geographic variability in the distribution of stream-living Lahontan Cutthroat Trout. *Transactions of the American Fisheries Society* 128: 875–889.

Fausch K, Nakano S, Ishigaki K (1994) Distribution of two congeneric charrs in streams of Hokkaido Island, Japan: Considering multiple factors across scales. *Oecologia* 100: 1–12.

Forbes SA, Richardson RE (1913) Studies on the biology of the upper Illinois River. *Bulletin of the Illinois Natural History Survey* 9: 481–574.

Fore LS (2003) Response of diatom assemblages to human disturbance: development and testing of a multi-metric index for the Mid-Atlantic Region (USA). In: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (ed Simon TP) CRC Press, Boca Raton, FL.

Frederick PC, Ogden JC (1997) Philopatry and nomadism: contrasting long-term movement behavior and population dynamics of white ibis and wood storks. *Colonial Waterbirds* 20: 316–323.

Frederick PC, Ogden JC (2001) Pulsed breeding of long-legged wading birds and the importance of infrequent severe drought conditions in the Florida Everglades. *Wetlands* 21: 484–491.

Frey DG (1977) Biological integrity of water-a historical approach. In: *The Integrity of Water* (eds Ballentine RK, Guarraia LJ) US Environmental Protection Agency, Washington, DC.

Gallo K, Lanigan SH, Eldred P, Gordon SN, Moyer C (2005) *Preliminary Assessment of the Condition of Watersheds*. General Technical Report PNW-GTR-647. USDA Forest Service, Pacific Northwest Research Station, Portland, OR.

Garcia SM, Staples DJ (2000) Sustainability reference systems and indicators for responsible marine capture fisheries: a review of concepts and elements for a set of guidelines. *Marine Freshwater Research* 51: 385–426.

Gates S (2002) Review of methodology of quantitative reviews using meta-analysis in ecology. *Journal of Animal Ecology* 71: 547–557.

Gawlik DE (2002) *South Florida Wading Bird Report*. Vol. 8. South Florida Water Management District, West Palm Beach, FL.

Gerritsen J (1995) Additive biological indices for resource management. *Journal of the North American Benthological Society* 14: 451–457.

Government Accountability Office (GAO) (2003) *South Florida Ecosystem Restoration: Task Force Needs to Improve Science Coordination to Increase the Likelihood of Success*. Report to Subcommittee on Interior and Related Agencies, Committee on Appropriations, House of Representatives. GAO-03-345. US Government Accountability Office, Washington, DC.

Government Accountability Office (GAO) (2005) *Chesapeake Bay Program: Improved Strategies are Needed to Better Assess, Report, and Manage Restoration Progress*. Report to Congressional requesters. GAO-06-96. US Government Accountability Office, Washington, DC.

Government Accountability Office (GAO) (2007) *South Florida Ecosystem: Restoration Is Moving Forward but is Facing Significant Delays, Implementation Challenges, and Rising Costs*. GAO-07-520. US Government Accountability Office, Washington, DC.

Gray PCR, Wiedemann PM (1999) Risk management and sustainable development: mutual lessons from approaches to the use of indicators. *Journal of Risk Research* 2(3): 201–218.

Halliwell DB, Langdon RW, Daniels RA, Kurtenbach JP, Jacobson RA (1999) Classification of freshwater fish species of the northeastern United States for use in the development of indices of biological integrity, with regional applications. In: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (ed Simon TP) CRC Press, Boca Raton, FL.

Harwell MA, Myers V, Young T, Bartuska A, Gassman N, Gentile JH, Harwell CC, Appelbaum S, Barko J, Causey B, Johonson C, McLean A, Smola R, Templet P, Tosini SA (1999) Framework for an ecosystem integrity report card. *BioScience* 49(7): 543–556.

Hassett BA, Palmer MA, Bernhardt ES (2007) Evaluating stream restoration in the Chesapeake Bay watershed through practitioner interviews. *Restoration Ecology* 15(3): 563–572.

Hawkins CP, Norris RH, Hogue JN, Feminella JW (2000) Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10: 1456–1477.

Haynes RW, Bormann BT, Lee DC, Martin JR (eds) (2006) *Northwest Forest Plan—the First 10 Years (1994–2003): Synthesis of Monitoring and Research Results*. General Technical Report PNW-GTR-651. US Department of Agriculture, Forest Service, Pacific Northwest Research Station Portland, OR.

Heffner RA, Butler MJ, Keelan RC (1996) Pseudoreplication revisited. *Ecology* 77: 2558–2562.

Hellawell JM (1986) *Biological Indicators of Freshwater Pollution and Environmental Management*. Elsevier, London.

Hertin J, Berkhout F, Moll S, Schepelmann P (2001) *Indicators for Monitoring Integration of Environment and Sustainable Development in Enterprise Policy*. Final Report, SPRU- Science and Technology Policy Research, University of Sussex, Brighton, UK.

Hiddink JG, Kaiser MJ (2005) Implications of Liebig's law of the minimum for the use of ecological indicators based on abundance. *Ecography* 18: 264–270.

Hill BH, Herlihy AT, Kaufmann PR, DeCelles SJ, Vanter Borgh MA (2003) Assessment of streams of the eastern United States using a periphyton indicator of biological integrity. *Ecological Indicators* 2: 325–338.

Hill BH, Herlihy AT, Kaufmann PR, Stevenson RJ, McCormick FH, Johnson CB (2000) The use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19: 50–67.

Hughes RM (1993) *Stream Indicator and Design Workshop*. EPA-600-R-93-138. US Environmental Protection Agency, Corvallis, OR.

Hughes RM (1995) Defining acceptable biological status by comparing with reference conditions. In: *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (eds Davis WS, Simon TP) Lewis Publishing, Boca Raton, FL.

Hughes RM, Gammon JR (1987) Longitudinal changes in fish assemblages and water quality in the Willamette River, Oregon. *Transactions of the American Fisheries Society* 116: 196–209.

Hughes RM, Oberdorff T (1999) Applications of IBI concepts and metrics to waters outside the United States and Canada. In: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (ed Simon TP) CRC Press, Boca Raton, FL.

Hughes RM, Howlin S, Kaufmann PR (2004) A biointegrity index for coldwater streams of western Oregon and Washington. *Transactions of the American Fisheries Society* 133: 1497–1515.

Hughes RM, Larsen DP, Omernik JM (1986) Regional reference sites: a method for assessing stream potentials. *Environmental Management* 10: 629–635.

Hughes RM, Paulsen SG, Stoddard JL (2000) EMAP-Surface Waters: a multiassemblage, probability survey of ecological integrity in the USA. *Hydrobiologia* 422/423: 429–443.

Hughes RM, Kaufmann PR, Herlihy AT, Kincaid TM, Reynolds L, Larsen DP (1998) A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences* 55: 1618–1631.

Hurlbert SH (1984) Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54: 187–211.

Independent Multidisciplinary Science Team (IMST) (2006) *Watershed and Aquatic Habitat Effectiveness Monitoring: A Synthesis of the Technical Workshop*. Technical Report 2006-1, Independent Multidisciplinary Science Team, Oregon Watershed Enhancement Board, Salem, OR.

Jackson LE, Kurtz JC, Fisher WS (2000) *Evaluation Guidelines for Ecological Indicators*. EPA/620/R-99/005. US Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

Jacobs SE, Cooney CX (1995) *Improvement of Methods Used to Estimate the Spawning Escapement of Oregon Coastal Natural Coho Salmon*. Oregon Department of Fish & Wildlife. Portland, OR.

Johnson BB, Chess C (2006). Evaluating public responses to environmental trend indicators. *Science Communication* 28(1): 64–92.

Joy MK, Death RG (2002) Predictive modelling of freshwater fish as a biomonitoring tool in New Zealand. *Freshwater Biology* 47: 2261–2275.

Karr JR (1981) Assessment of biotic integrity using fish communities. *Fisheries* 6(6): 21–27.

Karr JR, Chu EW (1999) *Restoring Life in Running Waters: Better Biological Monitoring*. Island Press, Covelo, CA.

Karr JR, Yoder CO (2004) Biological assessment and criteria improve total maximum daily load decision making. *Journal of Environmental Engineering* 130: 594–604.

Karr JR, Fausch KD, Angermeier PL, Yant PR, Schlosser IJ (1986) *Assessing Biological Integrity in Running Waters: A Method and its Rationale*. Illinois Natural History Survey. Special Publication 5.

Katz SL, Barnas K, Hicks R, Cowen J, Jenkinson R (2007) Freshwater habitat restoration actions in the Pacific Northwest: A decade's investment in habitat improvement. *Restoration Ecology* 15(3): 494–505.

Kaufmann PR, Hughes RM (2006) Geomorphic and anthropogenic influences on fish and amphibians in Pacific Northwest coastal streams. In: *Landscape Influences on Stream Habitat and Biological Assemblages* (eds Hughes RM, Wang L, Seelbach PW) Symposium 48, American Fisheries Society, Bethesda, MD.

Kaufmann PR, Herlihy AT, Baker LA (1992) Sources of acidity in lakes and streams of the United States. *Environmental Pollution* 77: 115–122.

Kerans BL, Karr JR (1994) A benthic index biotic integrity (B-IBI) for rivers of the TennesseeValley. *Ecological Applications* 4: 768–785.

Kettl DF (2001) Putting performance measurement to work in the federal government. *La Follette Policy Report* 12: 8–12.

Keup LE, Ingram WM, Mackenthun KM (1967) *Biology of Water Pollution*. Federal Water Pollution Control Administration, US Department of the Interior, Washington, DC.

Kingston JC, Birks HJB, Uutala AJ, Cumming BF, Smol JP (1992) Assessing trends in fishery resources and lake water aluminum from paleolimnological analyses of siliceous algae. *Canadian Journal of Fisheries and Aquatic Science* 49: 116–27.

Klauda R, Kazyak P, Stranko S, Southerland M, Roth N, Chaillou J (1998) The Maryland Biological Stream Survey: A state agency program to assess the impact of anthropogenic stresses on stream habitat quality and biota. *Environmental Monitoring and Assessment* 51: 299–316.

Klemm DJ, Blocksom KA, Fulk FA, Herlihy AT, Hughes RM (2003) Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. *Environmental Management* 31: 656–69.

Kolkwitz R, Marsson M (1908) Ökologie der pflanzlichen saprobien. *Berichte der Deutschen Botanischen Gesellschaft* 26a: 505–519.

Kondolf GM, Anderson S, Lave R, Pagano L, Merenlender A, Bernhardt ES (2007) Two decades of river restoration in California: What can we learn? *Restoration Ecology* 15(3): 516–523.

Kurtz JC, Jackson LE, Fisher WS (2001) Strategies for evaluating indicators based on guidelines from the Environmental Protection Agency's Office of Research and Development. *Ecological Indicators* 1: 49–60.

Landres PB (1992) Ecological indicators: panacea or liability? In: *Ecological Indicators*, Vol. 2. (eds McKenzie DH, Hyatt DE, McDonald VM) Elsevier Applied Science, London.

Langdon RW (2001) A preliminary index of biological integrity for fish assemblages of small coldwater streams in Vermont. *Northeastern Naturalist* 8: 219–232.

Larsen DP (1995) The role of ecological sample surveys in the implementation of biocriteria. In: *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (eds Davis WS, Simon TP) Lewis Publishing, Boca Raton, FL.

Larsen DP, Kaufmann PR, Kincaid TM, Urquhart NS (2004) Detecting persistent change in the habitat of salmon-bearing streams in the Pacific Northwest. *Canadian Journal of Fisheries and Aquatic Sciences.* 61:283–291.

Leonard PM, Orth DJ (1986) Application and testing of an index of biotic integrity in small, coolwater streams. *Transactions of the American Fisheries Society* 115: 401–415.

Linham GW, Kleinsasser LJ, Mayes KB (2002) *Regionalization of the Index of Biotic Integrity for Texas streams*. Texas Parks and Wildlife Department, Austin, TX.

Link JS, Brodziak JKT, Edwards SF, Overholtz WJ, Mountain D, Jossi JW, Smith TD, Fogarty MJ (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.

Little Hoover Commission (2005) *Still Imperiled, Still Important: The Little Hoover Commission's Review of the CALFED Bay-Delta Program.* Little Hoover Commission, CA. 112 p. http://www.lhc.ca.gov/lhcdir/183/report183.pdf.

Lyons J, Piette RR, Niermeyer KW (2001) Development, validation, and application of a fish-based index of biotic integrity for Wisconsin's large warm water rivers. *Transactions of the American Fisheries Society* 130: 1077–1094.

Lyons J, Wang L, Simonson TD (1996) Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. *North American Journal of Fisheries Management* 16: 241–256.

McCormick FH, Hughes RM, Kaufmann PR, Peck DV, Stoddard JL, Herlihy AT (2001) Development of an index of biotic integrity for the Mid-Atlantic Highlands region. *Transactions of the American Fisheries Society* 130: 857–877.

McGarvey DJ (2007) Merging precaution with sound science under the Endangered Species Act. *BioScience* 57: 65–70.

McGarvey DJ, Hughes RM (In Press) Longitudinal zonation of Pacific Northwest (USA) fish assemblages and the species-discharge relationship. *Copeia*.

Mebane CA, Maret TR, Hughes RM (2003) An index of biological integrity (IBI) for Pacific Northwest rivers. *Transactions of the American Fisheries Society* 132: 239–261.

Miller DL, Leonard PM, Hughes RM, Karr JR, Moyle PB, Schrader LH, Thompson BA, Daniels RA, Fausch KD, Fitzhugh GA, Gammon JR, Halliwell DB, Angermeier PL, Orth DJ (1988) Regional applications of an index of biotic integrity for use in water resource management. *Fisheries* 13(5):

12–20.

Miller RB, Anderson MJ (2004) Remedies for pseudoreplication. *Fisheries Research* 70: 397–407.

Molina R, McKenzie D, Lesher R, Ford J, Alegria J, Cutler R (2003) *Strategic survey framework for the Northwest Forest Plan Survey and Manage Program*. General Technical Report PNW-GTR-573. U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Moss D, Furse MT, Wright JF, Armitage PD (1987) The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwater Biology* 17: 41–52.

Moyle PB, Randle PJ (1998) Evaluating the biotic integrity of watersheds in the Sierra Nevada, California. *Conservation Biology* 12: 1318–1326.

Mundahl ND, Simon TP (1999). Development and application of an index of biotic integrity for coldwater streams of the upper Midwestern United States. In: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (ed Simon TP) CRC Press, Boca Raton, FL.

Niemela S, Pearson E, Simon TP, Goldstein RM, Bailey PA (1999) Development of an index of biotic integrity for the species depauperate Lake Agassiz Plain ecoregion, North Dakota and Minnesota. In: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (ed Simon TP) CRC Press, Boca Raton, FL.

Niemi GJ, McDonald ME (2004) Application of ecological indicators. *Annual Review of Ecology, Evolution, and Systematics* 35: 89–111.

Niemi G, Wardrop D, Brooks R, Anderson S, Brady V, Paerl H, Rakocinski C, Brouwer M, Levinson B, McDonald M (2004) Rationale for a new generation of indicators for coastal waters. *Environmental Health Perspectives* 112: 979–98.

Norris RH (1995) Biological monitoring: the dilemma of data analysis. *Journal of the North American Benthological Society* 14: 440–450.

Noss RF (1990) Indicators for monitoring biodiversity: A hierarchical approach. *Conservation Biology* 4: 355–364.

NRC (2000) *Ecological Indicators for the Nation*. National Research Council, National Academy of Sciences, Washington, DC.

Oberdorff T, Pont D, Hugueny B, Chessel D (2001) A probabilistic model characterizing fish assemblages of French rivers: a framework for environmental assessment. *Freshwater Biology* 46: 399–415.

Oberdorff T, Pont D, Hugueny B, Porchers JP (2002) Development and validation of a fish-based index for the assessment of 'river health' in France. *Freshwater Biology* 47: 1720–1734.

O'Connor RJ, Walls TE, Hughes, RM (2000) Using multiple taxonomic groups to index the ecological condition of lakes. *Environmental Monitoring and Assessment* 61: 207–28.

Ode PR, Rehn AC, May JT (2005) A quantitative tool for assessing the integrity of southern coastal California streams. *Environmental Management* 35(4): 493–504.

Ogden JC (1994) A comparison of wading bird nesting colony dynamics (1931– 1946 and 1974–1989) as an indication of ecosystem conditions in the southern Everglades. In: *Everglades. The ecosystem and its restoration* (eds Davis SM, Ogden JC) CRC Press, Boca Raton, FL.

Ogden JC, Davis SM, Barnes TK, Jacobs KJ, Gentile JH (2005) Total system conceptual ecological model. *Wetlands* 25: 955–979.

O'Mealy MK (2000) *Setting Measurable Goals for Salmon and Watershed Recovery; Examples and Recommendations for the Oregon Plan for Salmon and Watersheds*. MS Thesis, Oregon State University, Corvallis, OR.

Omernik JM, Griffith GE (1991) Ecological regions versus hydrologic units: frameworks for managing water quality. *Journal of Soil and Water Conservation* 46: 334–340.

Oregon Department of Forestry (ODF) (2007) *Draft Oregon Indicators of Sustainable Forest Management*. Oregon Department of Forestry, Salem, OR. 108 p. Available at: http://egov.oregon.gov/ODF/RESOURCE_PLANNING/Sustainable_Forest_Indicators_Project.shtml#Sustainable_Forestry_Indicator_Package. Accessed July 11, 2007.

Oregon Department of Fish and Wildlife (ODFW) (2007) *Appendix 2. Desired Status: Measurable Criteria for the Oregon Coast Coho Conservation Plan for the State of Oregon*. Oregon Department of Fish and Wildlife, Salem, OR. http://www.oregon.gov/OPSW/cohoproject/PDFs/Appendix2_final.pdf.

Oregon Watershed Enhancement Board (OWEB) (2005) *2003–2005 Oregon Plan Biennial Report,* Volume 2. Oregon Watershed Enhancement Board, Salem, OR.

Paquette A, Bouchard A, Cogliastro, A (2006) Survival and growth of under-planted trees: A meta-analysis across for biomes. *Ecological Applications* 16(4): 1575–1589.

Pitcher T, Preikhost D (2001) RAPFISH: a rapid appraisal technique to evaluate the sustainability status of fisheries. *Fisheries Research* 49: 255–270.

Plafkin JL, Barbour MT, Porter KD, Gross SK, Hughes RM (1989) *Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish*. EPA-444-4-89-001. US Environmental Protection Agency, Washington, DC.

Pont D, Hughes RM, Whittier TR, Schmutz S (In Review) A single predictive IBI model for fish assemblages of all western USA streams and rivers: a first approximation for the USA. *Transactions of the American Fisheries Society*.

Pont D, Hugueny B, Beier U, Goffaux D, Melcher A, Noble R, Rogers C, Roset N, Schmutz S (2006) Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. *Journal of Applied Ecology* 43: 70–80.

Ramsey F, Schafer D (2002). *The Statistical Sleuth: A Course in Methods of Data Analysis*. Duxbury Press, Pacific Grove, CA.

Rapport DJ, Singh A (2006) An ecohealth-based framework for state of environment reporting. *Ecological Indicators* 6: 409–428.

RECOVER (2004) *CERP Monitoring and Assessment Plan. Part I: Monitoring and supporting research*. South Florida Water Management District and US Army Corps of Engineers.

Resh VH, Norris RH, Barbour MT (1995) Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Australian Journal of Ecology* 20: 108–121.

Reynoldson TB, Norris RH, Resh VH, Day KE, Rosenberg DM (1997) The reference condition: a comparison of multi-metric and multivariate approaches to assess water quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society* 16: 833–852.

Rice JC, Rochet MJ (2005) A framework for selecting a suite of indicators for fisheries management. *ICES Journal of Marine Science* 62: 516–527.

Rieman B, Dunham J, Clayton J (2006) Emerging concepts for management of river ecosystems and challenges to applied integration of physical and biological sciences in the Pacific Northwest, USA. *International Journal of River Basin Management* 4(2): 85–97.

Roth N, Southerland M, Chaillou J, Klauda R, Kazyak P, Stranko S, Weisberg S, Hall L Jr, Morgan R II (1998) Maryland Biological Stream Survey: development of a fish index of biotic integrity. *Environmental Monitoring and Assessment* 51: 89–106.

Rumps JM, Katz SL, Barnas K, Morehead MD, Jenkinson R, Clayton SR, Goodwin P (2007) Stream restoration in the Pacific Northwest: Analysis of interviews with project managers. *Restoration Ecology* 15(3): 506–515.

Scharf FS, Juanes F, Sutherland M (1998) Inferring ecological relationships from the edges of scatter diagrams: comparison of regression techniques. *Ecology* 79: 448–460.

Schiller A, Hunsaker CR, Kane MA, Wolfe AK, Dale VH, Suter GW, Russell CS, Pion G, Jensen MH, Konar VC (2001) Communicating ecological indicators to decision makers and the public. *Ecology and Society* 5(1): 19. [on-line] URL: http://www.consecol.org/vol5/iss1/art19/.

Shah JJF, Dahm CN, Gloss SP, Bernhardt ES (2007) River and riparian restoration in the Southwest: Results of the National River Restoration Science Synthesis Project. *Restoration Ecology* 15(3): 550–562.

Shogren JF, Nowell C (1992) Economics and ecology: A comparison of experimental methodologies and philosophies. *Ecological Economics* (5): 101–126.

Simon TP, Lyons J (1995) Application of the index of biotic integrity to evaluate water resource integrity in freshwater ecosystems. In: *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (eds Davis WS, Simon TP) Lewis Publishing, Boca Raton, FL.

South Florida Ecosystem Research (SFER) (2003) *Task Force Needs to Improve Science Coordination to Increase the Likelihood of Success*. GAO-03-345 Report. South Florida Ecosystem Restoration. http://www.gao.gov/new.items/d03345.pdf

South Florida Ecosystem Research (SFER) (2006a) *Report of the Review Panel Concerning Indicators for Restoration*. South Florida Ecosystem Restoration. http://www.sfrestore.org/scg/documents/indicators%20REPORT%20final%20from%20Jordan%20MAY%208%202006_w_line%20numbers.pdf.

South Florida Ecosystem Research (SFER) (2006b) *Indicators for Restoration*. South Florida Ecosystem Restoration. http://www.sfrestore.org/scg/documents/2006_system_wide_indicators_report.pdf

South Florida Ecosystem Restoration Task Force (SFSCG) (2006) *Indicators for Restoration: South Florida Ecosystem Restoration*. September 5, 2006 draft report to the South Florida Ecosystem Restoration Task Force. South Florida Science Coordination Group, Florida International University, Miami, FL.

Stark JD (1985) A macroinvertebrate community index of water quality for stony streams. *Water and Soil Miscellaneous Publications* 87. Wellington Ministry of Works and Development, New Zealand.

Stark JD, Maxted JR (2007) A biotic index for New Zealand's soft bottomed streams. *New Zealand Journal of Marine and Freshwater Research* 41: 43–61.

Steedman RJ (1988) Modification and assessment of an index of biotic integrity to quantify stream quality in southern Ontario. *Canadian Journal of Fisheries and Aquatic Sciences* 45: 492–501.

Stockwell DRB, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modeling* 148: 1–13.

Stoddard JL, Larsen DP, Hawkins CP, Johnson RK, Norris RH (2006) Setting expectations for the ecological condition of running waters: the concept of reference condition. *Ecological Applications* 16: 1267–1276.

Stoddard JL, Peck DV, Paulsen SG, Van Sickle J, Hawkins CP, Herlihy AT, Hughes RM, Kaufmann PR, Larsen DP, Lomnicky G, Olsen AR, Peterson SA, Ringold PL, Whittier TR (2005) *An Ecological Assessment of Western Streams and Rivers*. EPA-620-R-05-005. US Environmental Protection Agency, Washington, DC.

Suter GW II (2001) Applicability of indicators monitoring to ecological risk assessment. *Ecological Indicators* 1: 101–112.

Tejerina-Garro FL, de Merona B, Oberdorff T, Hugueny B (2006) A fish-based index of large river quality for French Guiana (South America): method and preliminary results. *Aquatic Living Resources* 19: 31–46.

Tompkins MR, Kondolf GM (2007) Systematic postproject appraisals to maximize lessons learned from river restoration projects: Case study of compound channel restoration projects in Northern California. *Restoration Ecology* 15(3): 524–537.

Tran LT, Knight CG, O'Neill RV, Smith ER, Riitters KH, Wickham J (2002) Fuzzy decision analysis for integrated environmental vulnerability assessment of the mid-Atlantic region. *Environmental Management* 29: 845–859.

Turnhout E, Hisschemöller M, Eijsackers H (2007) Ecological indicators: between the two fires of science and policy. *Ecological Indicators* 7: 215–228.

United States Environmental Protection Agency (USEPA) (2000) *Evaluation Guidelines for Ecological Indicators*. EPA-620-R-99-005. US Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

United States Environmental Protection Agency (USEPA) (2006) *Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams*. EPA-841-B-06-002. US Environmental Protection Agency, Washington, DC.

Urquhart NS, Kincaid TM (1999) Designs for detecting trend from repeated surveys of ecological resources. *Journal of Agricultural, Biological and Environmental Statistics* 4: 404–414.

Urquhart NS, Kincaid TM, Paulsen SG, Larsen DP (1998) Monitoring for policy-relevant regional trends over time. *Ecological Applications* 8: 246–257.

US Environmental Protection Agency (USEPA) (2000) *Evaluation Guidelines for Ecological Indicators*. EPA-620-R-99-005. US Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

US Environmental Protection Agency (USEPA) (2006) *Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams*. EPA-841-B-06-002. US Environmental Protection Agency, Washington, DC.

Wang L, Seelbach PW, Hughes RM (2006) Introduction to influences of landscape on stream habitat and biological assemblages. In: *Landscape Influences on Stream Habitat and Biological Assemblages* (eds Hughes RM, Wang L, Seelbach PW) Symposium 48, American Fisheries Society, Bethesda, MD.

Wefering FM, Danielson LE, White NM (2000) Using the AMOEBA approach to measure progress toward ecosystem sustainability within a shellfish restoration project in North Carolina. *Ecological Monitoring* 130: 157–166.

Whittier TR, Stoddard JL, Hughes RM, Lomnicky G (2006) Associations among catchment- and site-scale disturbance indicators and biological assemblages at least- and most-disturbed stream and river sites in the western USA. In: *Landscape Influences on Stream Habitat and Biological Assemblages* (eds Hughes RM, Wang L, Seelbach PW) Symposium 48, American Fisheries Society, Bethesda, MD.

Whittier TR, Hughes RM, Stoddard JL, Lomnicky GA, Peck DV, Herlihy AT (2007a) A structured approach to developing indices of biotic integrity: three examples from western USA streams and rivers. *Transactions of the American Fisheries Society* 136: 718–735.

Whittier TR, Hughes RM, Lomnicky GA, Peck DV (2007b) Fish and amphibian tolerance classifications, tolerance values, and an assemblage tolerance index for western USA streams and rivers. *Transactions of the American Fisheries Society* 136: 254–271.

Wiersma YF (2005) Environmental benchmarks vs. ecological benchmarks for assessment and monitoring in Canada: Is there a difference? *Environmental Monitoring and Assessment* 100: 1–9.

Wilhm JL (1975) Biological indicators of pollution. In: *River Ecology* (ed Whitton BA) University of California Press, Berkeley, CA.

Wilton T (2004) *Biological Assessment of Iowa's Wadeable Streams*. Iowa Department of Natural Resources, Des Moines, IA.

Woolsey S, Capelli F, Gonser T, Hoehn E, Hostmann M, Junker B, Paetzold A, Roulier C, Schweizer S, Tiegs S, Tockner K, Weber C, Peter A (2007) A strategy to assess river restoration success. *Freshwater Biology* 52: 752–769.

Wright JF (1995) Development and use of a system for predicting macroinvertebrates in flowing waters. *Australian Journal of Ecology* 20: 181–197.

Yoder CO (1995) Policy issues and management applications of biological criteria. In: *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (eds Davis WS, Simon TP) Lewis Publishers, Boca Raton, FL.

Yoder CO, Rankin ET, Smith MA, Alsdorf BC, Altfater DJ, Boucher CE, Miltner RJ, Mishne DE, Sanders RE, Thoma RF (2005) Changes in fish assemblage status in Ohio's nonwadeable rivers and streams over two decades. In: *Historical Changes in Large River Fish Assemblages of the Americas* (eds Rinne JN, Hughes RM, Calamusso B). Symposium 45, American Fisheries Society, Bethesda, MD.

Appendices

Appendix A. Restoration Program Case Studies

Chesapeake Bay Program

The Chesapeake Bay Program (CBP) was created in 1983 when Maryland, Pennsylvania, Virginia, the District of Columbia, the Chesapeake Bay Commission, and EPA agreed to establish a partnership to restore the Chesapeake Bay. Coordination and support of monitoring activities of the CBP is the responsibility of the Monitoring and Analysis Subcommittee (MASC[10]). Various governmental agencies, academic partnerships and private organizations perform the monitoring activities of the CBP. The MASC coordinates those efforts by providing a forum for internal communication regarding all CBP monitoring activities that include collection, management, integration and analysis of data from multiple scientific disciplines. Activities of the MASC are supported by five workgroups: (1) the Indicators Workgroup, (2) the Analytical methods and Quality Assurance Workgroup, (3) the Data Management and Acquisition Workgroup, (4) the Nontidal Water Quality Workgroup, and (5) the Tidal Monitoring and Analysis Workgroup.

The GAO (2005) concluded that improved strategies were needed to assess, report, and manage restoration progress. The GAO (2005) also noted that while the Bay Program had over 100 measures to assess progress toward meeting restoration commitments and providing information to guide management decisions, it had not yet developed an integrated approach that would allow it to translate these individual measures into an assessment of overall progress toward achieving broad restoration goals. In recognition of this need, a task force began working on an integrated approach in November 2004.

The GAO (2005) report was also critical of the *State of the Chesapeake Bay* reports that were the primary mechanism for reporting the current health status of the bay. The GAO (2005) concluded that these reports did not effectively communicate the bay's current conditions because they focused on the status of individual species or pollutants instead of providing information on a core set of ecosystem characteristics. The GAO (2005) indicated that the credibility of these reports was negatively impacted because various kinds of data such as monitoring data, results of program actions, and the results of its predictive model were commingled without clearly distinguishing among them. Moreover, the lack of independence in the Bay Program's reporting process led to negative trends being downplayed and a rosier picture of the bay's health being reported than may have been warranted. The program recognized that improvements were needed and is developing new reporting formats.

The Indicators Workgroup is charged with developing indicators that communicate progress in restoration of water quality and living resources in the Chesapeake Bay and its watershed (CBP Indicators Redesign Workgroup[11]). The workgroup consists of 16 members representing state and federal agencies, research institutions, and public interest groups. Their charge is to: (1) develop a framework that relates Bay Program indicators to each other in ways that provide explanations of the current conditions of the Bay and watershed aquatic ecosystem, and report

---

[10]Chesapeake Bay Program. Bay Trends and Indicators. Accessed online February 2, 2007. http://www.chesapeakebay.net/indicators.htm

[11]Chesapeake Bay Program. Indicators Redesign Workgroup. Accessed online February 2, 2007. http://www.chesapeakebay.net/irw.htm

progress in their restoration, (2) develop guidelines for the design of indicators to provide the most effective communication with the public, and (3) develop indices that combine multiple indicators to provide over arching measures of Bay and watershed ecosystem health, stressors, and progress in restoration.

Three kinds of indicators are used: (1) those that assess the health of the Chesapeake Bay and its tidal tributaries (Table A1), (2) those that measure and communicate specific actions aimed at improving the health of the Chesapeake Bay and its watershed (Table A2), and (3) those that measure conditions and factors that are altering the health of the Chesapeake Bay and its watershed (Table A3). For each indicator, an overview, a Status and Trends report, and data can be accessed online at http://www.chesapeakebay.net/status.cfm?sid=215. The Status and Trends reports provide an overview of the indicator, the restoration goal and percentage of restoration achieved, a list of related indicators, and data source. The online interface provides access to several types of data related to the Chesapeake Bay. CBP databases can be queried based upon user-defined inputs such as geographic region and date range, resulting in downloadable files that can be imported to any program (e.g., SAS statistical package, Microsoft Excel, Microsoft Access) for further analysis.

Table A1. Indicators that assess the health of the Chesapeake Bay and its tidal tributaries. Descriptions on each indicator can be found at: http://www.chesapeakebay.net/status.cfm?view=Bay Health&subjectarea=INDICATORS

Bay Health

Animals:

Fish:

- Shad Returning to the Susquehanna River
- Striped Bass (Juvenile Indices)
- Striped Bass Abundance (Spawning Female Biomass)

Shellfish:

- Blue Crab (Juveniles)
- Blue Crab Abundance (Spawning Female Index)
- Native Oyster Abundance (Biomass)
- Oyster Spat (James River)
- Oyster Spat (Maryland)

Habitat:

Underwater bay grasses:

- Bay Grass Abundance (Baywide)
- Bay Grass Abundance (Upper, Middle and Lower Bay Zones)
- Bay Grass Density

Plankton and Bottom Dwellers:

Benthos:

- Bottom Habitat (Benthic Index of Biotic Integrity)
- Bottom Habitat by Region (Benthic Index of Biotic Integrity)

Plankton:

- Phytoplankton (Index of Biotic Integrity)

Water Quality:

Chemical Contaminants:

- Chemical Contaminants

Chlorophyll a:

- Chlorophyll a: Annual Assessment
- Chlorophyll a: Three-Year Assessment (Guidance Achievement)

Dissolved Oxygen:

- Dissolved Oxygen: Annual Assessment
- Dissolved Oxygen: Three-Year Assessment (Standards Attainment)

Water Clarity:

- Water Clarity: Three-Year Assessment (SAV-Based Clarity Standards Attainment)
- Water Clarity: Annual Assessment (Mid-Channel)

Table A2. Indicators that measure conditions and factors that are altering the health of the Chesapeake Bay and its watershed. Descriptions of each indicator can be found at: http://www.chesapeakebay.net/status.cfm?view=Factors%20Impacting%20Bay%20Health&subjectarea=INDICATORS

Factors Impacting Bay Health

Fisheries Harvest:
- Blue Crab (Commercial Harvest and Fishing Mortality Rate)
- Oysters (Commercial Harvest)

Land Use:
- Chesapeake Bay Watershed Forests
- Chesapeake Bay Watershed Development Trends
- Chesapeake Bay Watershed Land Use
- Chesapeake Bay Watershed Riparian Forest Buffers

People:
- Chesapeake Bay Watershed Population

Pollutants:

Nitrogen:
- Nitrogen Loads Delivered to the Bay from Municipal and Industrial Wastewater
- Nitrogen Loads and River Flow to the Bay
- Nontidal Nitrogen Loads and River Flow to Chesapeake Bay
- Sources of Nitrogen Loads to the Bay

Phosphorus:
- Phosphorus Loads Delivered to Bay from Municipal and Industrial Wastewater
- Nontidal Phosphorus Loads and River Flow to Chesapeake Bay
- Phosphorus Loads and River Flow to the Bay
- Sources of Phosphorus Loads to the Bay

Sediment:
- Nontidal Sediment Loads and River Flow to Chesapeake Bay
- Sources of Sediment Loads to the Bay

River Flow:
- River Flow into Chesapeake Bay

Table A3. Indicators that measure and communicate specific actions aimed at improving the health of the Chesapeake Bay and its watershed. Descriptions of each indicator can be found at:
http://www.chesapeakebay.net/status.cfm?view=Restoration%20and%20Protection%20Efforts&subjectarea=INDICATORS

Restoration and Protection Efforts

Fostering Stewardship:
- Chesapeake Bay Partner Communities
- Public Access Points to the Chesapeake Bay and Its Tributaries
- Water Trails in the Chesapeake Bay Watershed

Managing Fisheries:
- Hatchery Reared American Shad Stocking
- Fisheries Management Effort Index (Blue Crab, Oyster, Striped Bass, Shad, Menhaden)

Managing Habitats:
- Opening Rivers to Migratory Fish
- Wetlands Restoration
- Bay Grasses Planted
- Riparian Forest Buffers Planted
- Wetlands Enhancement
- Wetlands Regulatory Programs

Managing Pollutants and Land Use:
- Brownfields Redevelopment in Chesapeake Bay Watershed
- Nitrogen in Rivers Entering the Bay: Flow Adjusted Concentration Trends
- Phosphorus in Rivers Entering the Bay: Flow Adjusted Concentration Trends
- Pollution Control Summary (Controlling Nitrogen, Phosphorus and Sediment)
- Watershed Land Preservation: by Entity
- Agricultural Pollution Controls
- Sediment in Rivers Entering the Bay: Flow Adjusted Concentration Trends
- Wastewater Pollution Controls
- Watershed Land Preservation
- Watershed Management Plans Developed
- Watershed Management Plans Developed (by State)

The CBP places environmental indicators in a hierarchy to assess environmental condition and/or change. The lowest levels in the hierarchy are indicators that characterize management actions such as steps taken by regulatory agencies and responses by regulated and nonregulated communities (Figure A4). In the example below, National Pollutant Discharge Elimination System permits represent Level 1 management actions while the number of acres under nutrient management plans represents a Level 2 response. Intermediate levels might represent changes in nitrogen or phosphorus discharge to the Bay (Level 3) and changes in water clarity (Level 4). Indicators such as those listed in Table A1 provide a measure of ecosystem health at the top of the hierarchy in Levels 5 and 6.

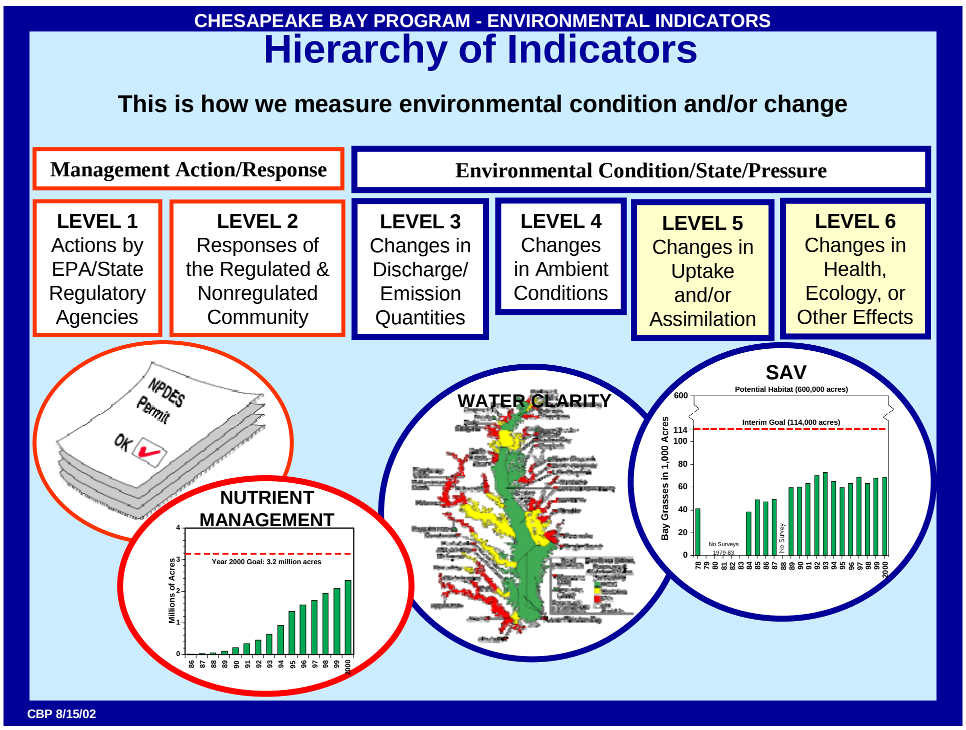

Figure A1. Schematic illustration of the use of hierarchal indicators in the Chesapeake Bay Program. Figure from: http://www.chesapeakebay.net/status.cfm?sid=88 (accessed online February 2, 2007)

CBP indicators are also placed in "Tracks" or areas of focus as well as being placed in a hierarchy representing actions to impacts. The way in which the hierarchy and tracks are used together to measure progress toward ecosystem restoration is illustrated in Figure A2. Examples of indicators that might be used in each track are:

Track 1. **Nutrients**

- Total phosphorus concentration
- Total nutrient concentration
- Acres under nutrient management plan

Track 2. **Living Resources**

- Acres of Bay grasses
- Striped Bass
- Oysters

Track 3. **Toxics**

- Industrial releases and transfers of chemical contaminants
- Acres under Integrated Pest Management

Track 4. **Cross-cutting indicators**

- Bald eagles
- Boat pump-out facilities
- Chesapeake Basin forests

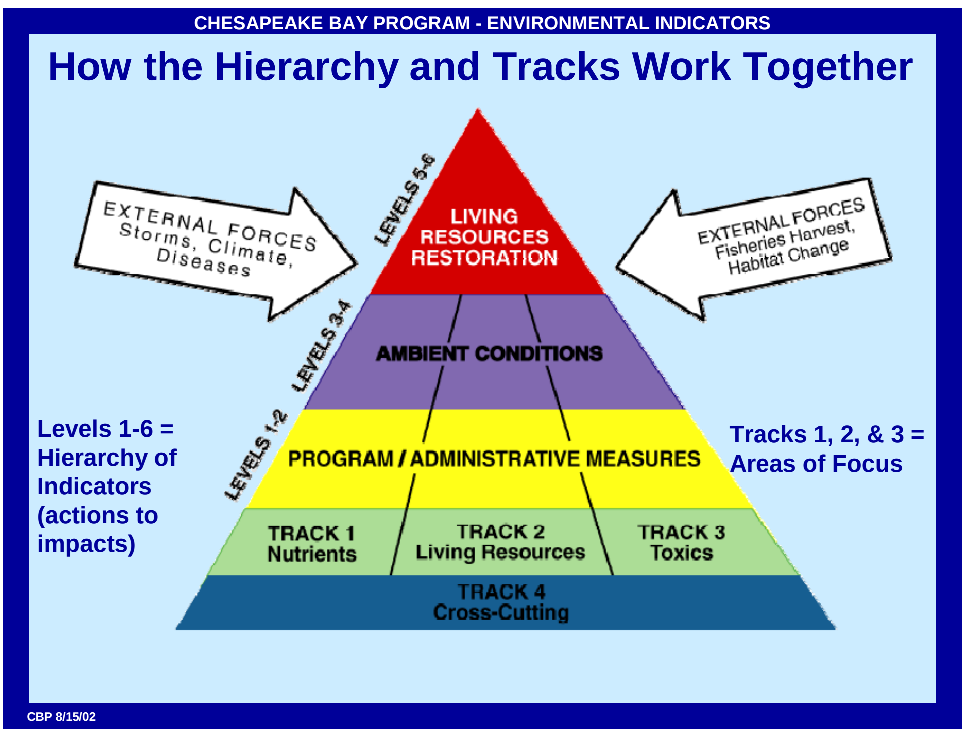

Figure A2. Schematic illustration of the way in which hierarchal indicators and indicator tracks are used together to measure progress toward Chesapeake Bay restoration. Figure reproduced from: http://www.chesapeakebay.net/status.cfm?sid=88 (accessed online February 2, 2007)

South Florida Ecosystem Restoration

The South Florida Ecosystem Restoration Task Force (SFER Task Force) was established in 1996 and includes seven federal, two tribal, and five state and local government representatives charged with overseeing restoration, preservation and protection of the South Florida ecosystem (SFER 2006a; SFER Task Force[12]). SFER Task Force duties include: (1) coordinate the development of consistent policies, strategies, plans, programs, projects, activities, and priorities; (2) exchange information regarding programs, projects and activities of the agencies and entities represented on the Task Force; (3) facilitate the resolution of interagency and intergovernmental conflicts; (4) coordinate scientific and other research, and (5) provide agencies and entities with assistance and support in carrying out restoration activities.

A GAO (2003) review based upon expenditures over Fiscal Years 1993–2002 concluded that the SFER Task Force needed to improve science coordination to increase the likelihood of success. The GAO (2003) report stated that, while scientific understanding of restoration issues had increased gaps in scientific information and adaptive management tools would soon hinder the success of restoration. Specifically, the GAO identified the need for information on the risks of contaminants to plants and animals in the ecosystem and the need to develop adaptive management tools—such as models and a comprehensive monitoring plan based on key indicators—that allow scientists to assess how the implementation of restoration projects and plans affect the ecosystem and whether this implementation is resulting in successful restoration. The GAO also concluded that the Task Force had not provided clear direction or adequate funding to ensure that scientific activities were being adequately coordinated by the Science Coordination Group (SCG).

A subsequent GAO (2007) review found that while many of the restoration effort's projects have been completed or are ongoing, a core set of projects that are critical to the success of the restoration are behind schedule or not yet started. The completed projects will provide improved water quality and additional habitat for wildlife, and the ongoing projects will also help restore wildlife habitat and improve water flow within the ecosystem. However, the projects most critical to the restoration's overall success are among those that are currently being designed, planned, or have not yet been started. Some of these projects are behind schedule by up to 6 years. In addition, the review noted that of the 27 primary mathematical models that guide the restoration effort, only 21 are able to interface with other models and provide a more comprehensive picture of the impact of restoration efforts on the ecosystem. Agency officials stated that additional interfaces are needed.

In 2004 the Task Force initiated the development of a *Plan for Coordinating Science* and in 2005 directed the SCG (see www.sfrestore.org) to develop system-wide indicators for restoration. The Task Force will use indicators to judge the performance of projects in achieving restoration goals outlined in the Comprehensive Everglades Restoration Plan and goals and projects included in the Task Force Strategic Plans. Indicators will also be used to evaluate ecological changes resulting from the implementation of restoration projects and to adapt and improve, add, replace or remove indicators as new scientific information and findings become available.

---

[12] South Florida Ecosystem Restoration Task Force. Accessed online January 31, 2007. http://www.sfrestore.org

The goal was to select a set of indicators that captured the “essence” or defining set of “features” of South Florida ecosystems to include characteristics distinctive of landscape, trophic constituents, biodiversity, and physical properties. The SCG developed an initial suite of South Florida system-wide indicators of restoration success using a 4-step process.

Step 1. *Evaluate various restoration efforts to identify possible indicators for inclusion in the suite of system-wide indicators,*

Step 2. *Use guidelines to select relevant indicators for Everglades Ecosystem applicability, evaluate the list of indicators for individual and collective value and coverage of Everglades' ecosystem regions, characteristics, trophic interactions, and functions,*

Step 3. *Identify “indicator gaps” and develop new indicators to fill identified gaps,*

Step 4. *Select final system-wide suite of indicators, develop indicator documentation and communication proposal, and identify “indicator gaps” to be filled by 2008 or beyond.*

In Step 2 above, two sets of guidelines were used to determine that the suite of indicators collectively provided sufficient “coverage” of the regions, characteristics, trophic interactions, properties and functions of the ecosystem. One set of guidelines was for ecological indictors and the other set for indicators of compatibility of the built-system elements of the Comprehensive Everglades Restoration Plan intended to provide water and maintain levels of flood protection of developed areas dominated by humans. Guidelines used are shown in Table A4.

Table A4. Restoration Indicator Guidelines developed by South Florida Ecosystem Restoration Task Force, Science Coordination Group (SFER Task Force 2006).

Ecological Indicator Guidelines	***Restoration Compatibility Guidelines***
1. Is the indicator relevant to the ecosystem and does it respond to variability at a scale that makes it applicable to the entire system or a large or important portion of it?	1. Does the indicator provide a measure of compatibility of the built system with ecological restoration?
2. Is the indicator feasible to implement (is someone collecting data already)?	2. Is the indicator feasible to implement (is someone collecting data already)?
3. Is the indicator sensitive to system drivers?	3. Is the indicator sensitive to system drivers (stressors, operations of water management)?
4. Is the indicator interpretable in a common language?	4. Is the indicator interpretable in a common language?
5. Are there situations where even an "optimistic" trend with regard to the indicator might suggest a "pessimistic" restoration trend?	5. Is the indicator scientifically defensible?
6. Are there situations where a "pessimistic" trend with regard to the indicator may be unrelated to restoration activities?	6. Are clear measurable targets established for the indicator to allow for assessments of success of affects of management actions and operations on ecological restoration?
7. Is the indicator scientifically defensible?	7. Does the indicator have specificity? Does it indicate a feature specific enough to result in management action or corrective action?
8. Are clear, measurable targets established for the indicator to allow for assessments of success of ecological restoration and effects of management actions?	
9. Does the indicator have specificity? Does it indicate a feature specific enough to result in management action or corrective action?	
10. What level of ecosystem process or structure does the indicator address?	

In Step 3 of the 4-Step process, indicators were evaluated to assess their relative coverage of different ecosystem characteristics, properties, spatial coverage and drivers/stressors. A color-coded system was used to visually illustrate adequacy of indicator coverage. A green color indicated either (a) research has been done establishing a direct statistical correlation or (b) the area or regions are directly monitored for this indicator. A yellow color was used to indicate either (a) a link identified by the Conceptual Ecological Model exists but may not be a research established statistical correlation, or (b) the region is only partially covered or represented by the indicator. An orange color indicated that either (a) an assumed ecological link suggesting the indicator integrates information about this feature of the ecosystem but that no research based links have been demonstrated or (b) the region is not well monitored for this indicator but the indicator could apply to this region with expanded monitoring. A black color was used to indicate either (a) that this feature is not being studied or monitored and the indicator is presumed not to include this ecological feature in the information it provides or (b) that this region is not included for this indicator in any monitoring. This process proved useful in identifying key gaps in indicator coverage, leading to inclusion of an exotic plant indicator. Additionally, the process helped identify potential indicators for which significant additional work is required to produce, develop and peer review conceptual ecological sub-models.

The SCG recommended a suite of 13 indicators for the 2006 biennial reporting period (Table 2). Collectively, these indicators will help the Task Force assess restoration goals and targets:

1. Fish & Macroinvertebrates
2. Wading Birds (Woodstork, White Ibis & Roseate Spoonbill)
3. Florida Bay Submerged Aquatic Vegetation
4. Florida Bay Algal Blooms
5. Crocodilians (Alligators & Crocodiles)
6. American Oysters
7. Periphyton-Epiphyton
8. Juvenile Pink Shrimp
9. Lake Okeechobee Littoral Zone
10. Invasive Exotic Plants
11. Water Volume
12. Salinity Intrusion in the Biscayne Aquifer
13. Flood Protection – C-111 Basin

Each of these indicators is described in detail in the *Indicators for Restoration* draft report (SFER Task Force 2006). Indicator descriptions range from six to 16 pages in length and follow a common format as illustrated by the description of the Wading Birds indicator as reprinted from the *Indicators for Restoration* draft report.

Indicator description reproduced from SFER (2006b).

WADING BIRDS (White Ibis and Wood Storks)

Author: John C. Ogden

What is this Indicator?
Extremely large numbers of colonial wading birds were one of the defining characteristics of the pre-drainage wetlands of south Florida (Ogden *et al.* 2005). Of particular relevance in understanding the population dynamics of wading birds in the pre-drainage system, are the combined features of large spatial extent and highly variable hydrological conditions that created and maintained a mosaic of wetland habitats. This combination is what made it possible for the region to support large nesting colonies of two species of wading birds with quite different foraging strategies and prey requirements, White Ibis and Wood Storks.

White Ibis forage for small fish and crayfish in very shallow water in wetlands that dry annually in most years. Ibis tend to forage close to nesting colonies (<20 km) and, therefore, relocate their colony sites and change the timing of nesting from year to tear in response to shifting locations of high densities of prey (Frederick & Ogden 1997). In contrast, Wood Storks tend to forage on the larger sizes of marsh fishes, often in deeper pools that do not dry annually. Storks routinely soar great distances from colony sites (25-75 km) and are able to reuse traditional colony sites for many years, irrespective of shifting locations of prey. Historically, ibis initiated nesting in most years in mid- to late dry seasons when water levels were low, while storks initiated nesting early in dry seasons when water levels were higher. With a comparatively short nesting cycle for ibis, and a much longer cycle for storks, both species fledged young in the late dry season when prey concentrations were generally highest.

The broad restoration goals for ibis and storks are about recovering the kind of ecosystem with the spatial and temporal variability to support large numbers of both of these behavioral and habitat specialists. The specific restoration goals for these two species include targets for numbers of nesting pairs, location of colonies, timing of nesting, and an increase in the size and frequency of the larger nesting assemblages (“super colonies”).

- An initial numerical goal for ibis and storks is to recover and/or sustain nesting populations of 50,000 and 5,000 birds, respectively. Long-term numerical goals have yet to be set.
- The restoration goal for location of stork colonies is a return of large nesting colonies in the southern, mainland estuaries of Everglades National Park, and a return to multiple colony sites in the Big Cypress basin.
- The restoration goal for timing of nesting by storks is a recovery of the historical pattern of colony formation in the early dry season months, November-January.
- An increase in the size of ibis super colonies, and an increase in the frequency to not less than two super colony events per 10 years.

CERP MAP Hypotheses related to Wading Bird Indicators (RECOVER 2004):

- Restoration of the density, seasonal concentrations, size structure, and taxonomic composition of marsh fishes and other aquatic fauna to levels that support sustainable breeding populations of higher vertebrates,
- Shift the distribution of high densities populations of marsh fishes and other aquatic fauna from artificially-pooled areas (WCAs) to the restored natural pools in the southern Everglades,
- Shift the foraging distribution of wading birds in response to expected trends in the density, distribution, and concentration of prey organisms,
- Re-establish wading bird nesting colonies in the coastal regions of the southern Everglades and an increase in the numbers of nesting pairs and colony sizes in response to desired trends in populations of prey organisms.

- Increase nesting success/survival rates of wading birds

What Has Happened To Affect The Indicator?

The drainage of extensive areas of short-hydroperiod wetlands, large-scaled alterations in water depth and distribution patterns due to compartmentalization of wetlands in the central Everglades, and the reduction of freshwater flows into the formerly more productive estuaries, are the anthropogenically-induced stressors that have substantially impacted ibis, storks and other wading birds in south Florida (Ogden 1994). Both ibis and storks have responded to these stressors by largely abandoning former nesting and roosting sites in the southern Everglades and Big Cypress basins, by delaying the initiation of nesting by several months (storks), and by rarely forming the "super-colonies" that once characterized the south Florida wetlands (ibis) (Frederick & Ogden 2001). The number of ibis nesting in south Florida has declined from an estimated 100,000 – 200,000 birds in the 1930s - 1940s (years of super colonies) to 20,000 – 60,000 birds since the late 1990s. The number of nesting storks has declined from 14,000 – 20,000 birds prior to 1960 to about 2,000 – 5,000 birds since the late 1990s (Ogden 1994). The loss of early-dry season foraging habitats has caused storks to delay the initiation of nesting by 2-3 months in many years, which has often resulted in young birds still being in nests when summer rains begin, and prey concentrations are lost. The disruption of natural hydrological patterns has substantially disrupted natural wet-dry patterns, thought to be of major importance in organizing the production pulses that supported super-colony formation.

What Areas of the Southern Florida Ecosystems Does This Indicator Cover?

White Ibis and Wood Storks, and other associated species of wading birds in south Florida (e.g., Great Egrets, Snowy Egrets, Tricolored Herons), are system-wide indicators for the south Florida wetlands. The areas used by these birds include the following RECOVER & SCG regional modules: Greater Everglades, Florida Bay and Southern Estuaries, Northern Estuaries, Big Cypress, Lake Okeechobee, and the Kissimmee River Basin. On seasonal, annual, and multi-year periods, these species of wading birds move about over large spatial scales in locating and utilizing good foraging habitats. The seasonal and annual variability in rainfall that characterizes south Florida means that the optimum foraging conditions for wading birds also vary both temporally and spatially. Wading birds are integrating information from many different regions in determining when and where they forage and form nesting colonies. In addition, individual wading birds may fly long distances daily, between roosts or nesting colonies and optimum foraging sites. The daily, seasonal, and annual patterns of movement by wading birds often occur at multi-landscape scales, and can cross among freshwater and estuarine communities.

Why Is This Indicator Important?

1. The Indicator is relevant to the Everglades ecosystem and responds to variability at a scale that makes it applicable to the entire ecosystem or large portions of the ecosystem:

- White Ibis and Wood Storks and other species of colonial-nesting wading birds are well adapted to be successful in a healthy Everglades-type ecosystem;
- These species are characteristic of the freshwater and estuarine greater Everglades system;
- Ibis and storks are top predators in Everglades aquatic food chains;
- The distribution and abundance of ibis, storks and other wading birds is determined by temporal and spatial scales of production and availability of aquatic prey;
- Ibis and storks and other species of wading birds move about over large spatial scales in response to variable seasonal and annual patterns in the quality of foraging habitat;
- The quality of good foraging habitat is directly linked to regional and system-wide hydrological patterns.

2. The indicator is feasible to implement and is scientifically defensible:

- Survey protocols for foraging and nesting patterns for ibis and storks and other species of colonial-nesting wading birds are well developed in south Florida;

- Major portions of the Everglades ecosystem are currently being surveyed for nesting colony patterns;
- Many surveys of nesting colonies and foraging patterns have previously been conducted, providing a strong record of past patterns;
- There is a strong body of research and published information for wading birds in the Everglades system, providing a solid, base-line understanding of the linkages between hydrological patterns and the ecology and biology of wading birds;
- Wading birds have already been established as indicators for CERP success, and are included in the RECOVER Monitoring and Assessment Plan, and as a recommended CERP Interim Goal.

3. The indicator is sensitive to system drivers (stressors):

- Wading birds show sensitivities to anthropogenically-induced altered hydropatterns in the Everglades by changing the location, timing and magnitude of nesting and foraging at system-wide scales;
- A strong set of working hypotheses have been developed to explain how and why wading birds have been adversely affected by drainage and management practices in the Everglades system, as a basis for predicting wading bird responses to restoration programs.

4. The indicator is integrative.

- The nesting and foraging patterns of ibis and storks and other species of wading birds is strongly influenced by patterns of abundance and availability of aquatic prey, which in turn are influenced by the production and density of prey, which are determined by past and current hydrological patterns;
- Ibis and storks feed on different prey, and have different foraging strategies, therefore the collective responses of these two species, and other species of wading birds, reveal broad system-wide conditions of aquatic production and availability;
- The high levels of mobility of wading birds, both in time and space, can reveal how wading birds are integrating information of foraging and nesting conditions over large temporal and spatial scales.

5. Goals and performance measures are established in the RECOVER MAP for the indicator and the following metrics are recommended for monitoring:

- Numbers of nesting colonies
- Locations of nesting colonies
- Timing of nesting
- Species composition of nesting colonies
- Frequency of occurrence of “super colonies”.

Discussion
Large numbers of showy wading birds were a conspicuous feature of the predrainage wetlands of south Florida. Single nesting colonies that contained an estimated 100,000 to 200,000 birds were reported in some early years. Although most of the early colonies were decimated by plume hunters in the late 19th Century, protective legislation and good remaining habitat conditions during the early 20th Century allowed most of the nesting species to fully recover by the 1930s. The huge “rookery” that was located along the extreme headwaters of Shark River was estimated in 1934 to have been a mile long and several hundred feet wide, and was so packed with nests and young birds that it was difficult to walk through the colony without pushing into nests (R.P.Allen, field notes). These bird cities were symbolic of the richness and abundance of the former south Florida wetlands, and they had largely disappeared by the end of the 1960s.

The location and size of colonies, the species composition, and the timing of nesting by wading birds in the pre-drainage south Florida wetlands were largely determined by the physical and ecological characteristics of these wetlands. It is predicted that the recovery of these historical nesting patterns will be a strong indicator that the mainland wetlands in south Florida have been successfully restored to an Everglades-type of ecosystem that much more closely resembles the pre-drainage system than do the current wetlands. Successful recovery of historical White Ibis and Wood Stork nesting patterns will be especially indicative of restoration success because of the special and contrasting behavioral and habitat characteristics between these two species. Recovery of wetland systems that can support large numbers of both of these two species will be an ultimate measure of Everglades' restoration success.

Longer-Term Science Needs

The White Ibis and Wood Stork indicators are based on patterns of nesting for these two species. For these patterns to be properly measured and evaluated over time, a comprehensive, system-wide program of monitoring nesting colonies is required (locations, species composition, numbers of nesting pairs, measures of success). Currently, no such system-wide survey of nesting colonies is in effect (Gawlik 2002). The regions of south Florida that are being systematically surveyed are the three WCAs, plus mainland Everglades National Park. Important regions that are not being systematically surveyed include Lake Okeechobee, the Big Cypress basin, and portions of the Caloosahatchee and St. Lucie estuaries. Much information on the basic biology, food habits, movement patterns for ibis and storks has been researched and reported. In the context of restoration, several key questions remain unaddressed. These include:

- A more complete understanding of the biology and ecology of the two species of freshwater crayfish, key prey species for the ibis. An especially important question pertains to the ecological conditions that supported the tremendous numbers of crayfish that were reported in the pre-drainage Everglades basin.
- The natural pattern of high water and drought that are hypothesized to have organized pulses of production in an otherwise oligotrophic system, and which may have supported the periodic formation of super colonies, is poorly understood. Key questions have to do with the role of multi-year droughts in nutrient and production dynamics in the greater Everglades.
- Although systematic surveys of wading bird foraging patterns have been conducted for many years, the relationships between wading bird abundance and foraging patterns, and the location, size and timing of nesting colonies is still poorly understood.

Literature Cited

Frederick, P.C. and J.C. Ogden. 1997. Philopatry and nomadism: contrsting long-term movement behavior and population dynamics of white ibis and wood storks. Colonial Waterbirds 20: 316-323.

Frederick, P.C. and J.C. Ogden. 2001. Pulsed breeding of long-legged wading birds and the importance of infrequent severe drought conditions in the Florida Everglades. Wetlands 21: 484-491.

Gawlik, D.E (ed.). 2002. South Florida Wading Bird Report. Vol. 8. South Florida Water Management District (issued annually 1995-2002; more recent issues through 2006 edited with several subsequent editors).

Ogden, J.C. 1994. A comparison of wading bird nesting colony dynamics (1931-1946 and 1974-1989) as an indication of ecosystem conditions in the southern Everglades. Pp. 531-570, in, Everglades. The ecosystem and its restoration. St. Lucie Press. Delray Beach, FL

Ogden, J.C., S.M. Davis, T.K. Barnes, K.J. Jacobs, and J.H. Gentile. 2005. Total System conceptual ecological model. Wetlands 25: 955-979.

RECOVER. 2004. CERP monitoring and assessment plan. Part I. Monitoring and supporting research. South Florida Water Management District and U.S. Army Corps of Engineers.

CALFED Bay-Delta Program

The CALFED Bay-Delta Program (CALFED) was initiated in 1994 and is the largest and most comprehensive water management and ecosystem restoration project in the nation (CALFED[13]; Little Hoover Commission 2005). The Bay-Delta Plan is primarily a comprehensive approach to reduce conflicts over limited water supplies. Four program objectives are addressed through 11 major program elements. Program objectives include; (1) water supply reliability, (2) levee system integrity, (3) water quality, and (4) ecosystem restoration. Program elements addressing these objectives include: (1) water management, (2) storage, (3) conveyance, (4) water use efficiency, (5) water transfers, (6) environmental water account, (7) drinking water quality, (8) watershed management, (9) levee system integrity, (10) ecosystem restoration, and (11) science. Details about these program elements can be obtained from online resources (CALFED[14]).

CALFED defines two types of indicators (CALFED 2006b).

> *"**Indicators** are a broad set of measurements used to evaluate the state of the system and provide better understanding about how the system is working"*, and

> *"**Performance measures** are indicators that are used to evaluate progress towards program goals".*

Beyond these definitions, CALFED characterizes three general levels of indicators:

1. *Administrative indicators* (also called "input measures" or "input indicators"). These describe what resources (funds, programs, projects) are being implemented (or plan to be implemented). Examples: Dollars spent, number of projects implemented
2. *Driver indicators* (also called "pressures," "management actions" and "other factors"). These describe the factors that may be influencing outcomes. There are two types of driver indicators:
   - *Outputs*, which are on-the-ground implementation of management actions, such as acres of habitat restored, and
   - *Uncontrollable factors*, which are often natural phenomena not caused by the management actions of the program, such as weather and hydrologic fluctuations
3. *Outcome indicators* (also called "response," "ecosystem status or state" or "results" indicators). These describe measurements related to the ultimate outcome of the drivers – and should be closely related to the goals and objectives of the program. Examples: For water quality, indicators may include measures of public health protection for tap water and cost of treatment. For water supply reliability, indicators may be related to the ability of supply to meet demand. For ecosystem restoration, indicators can be population level of key species, diversity indices, or other indicators of ecosystem status and processes. Quantitative models may provide predicted outcome indicators that can be used to evaluate future management options.

---

[13] CALFED Bay-Delta Program. Description. Accessed online 26 January, 2007. http://www.calwater.ca.gov/
[14] CALFED Bay-Delta Program. Program elements. Accessed online 26 January 2007b. http://www.calwater.ca.gov/Programs/Programs.shtml

CALFED uses a conceptual model to link these three types of indicators to policy, implementation, and cause and effect, as illustrated in Figure A3.

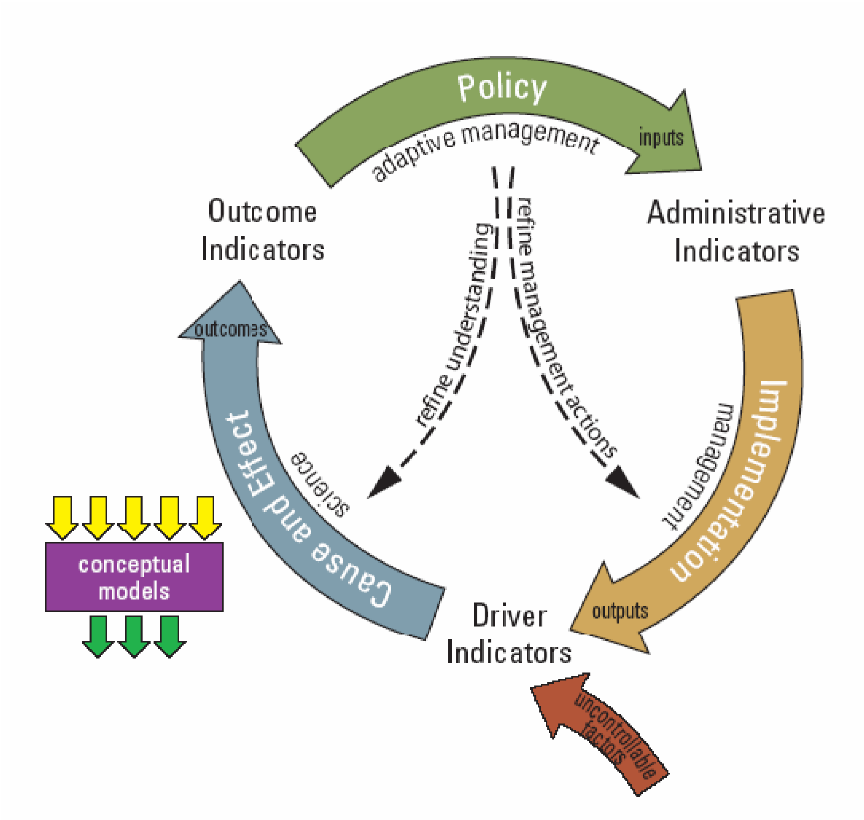

Figure A3. CALFED's conceptual model describing the relationship between three different levels of indicators and the activities of managing a complex system in the environment. Reproduced from: http://www.science.calwater.ca.gov/monitoring/includes/monitoring_indicators_112805.jpg (accessed on-line September 10, 2007.

An external review in 2005 concluded that the CALFED Bay-Delta Authority was unable to effectively coordinate activities, push agencies to perform, or provide rigorous oversight (Little Hoover Commission 2005). In their review, the Little Hoover Commission made four recommendations.

> **Recommendation 1**. *State and federal leaders need to refine the strategy for developing and implementing long-term and sustainable solutions to the Bay-Delta. That strategy should be integrated into a comprehensive water policy for California that encourages the best use of a scarce and essential resource.*
>
> **Recommendation 2**. *The California Bay-Delta Authority as a coordinating entity should be replaced by a leadership structure that has the authority to accomplish CALFED's mission.*
>
> **Recommendation 3**. *Implementation of CALFED must be strategic, performance-based, and accountable for outcomes.*

Recommendation 4. *The State must provide more meaningful opportunities for the public and stakeholders to participate in the CALFED process to raise awareness, increase transparency, reduce conflicts and provide accountability.*

In response to the Little Hoover Commission review, CALFED implementing agencies, with guidance from the CALFED Science Program, embarked on a 10-Year action plan to develop indicators and performance measures to be used to understand cause and effect relationships between actions and outcomes, track progress toward goals, inform decisions, and assess progress and performance. Agencies formed a subcommittee and subgroups to develop indicators for each of the four CALFED program objectives using a four-phase approach. Phase 1 of that approach included identifying primary program objectives, selecting core indicators, and determining the availability of comprehensive monitoring data and conceptual models (CALFED 2006a, 2006c). To guide their discussions through the process, each subgroup responded to the following list of questions and tasks.

Overall questions

- What strategic objectives were selected to work on for this phase and why?
- What other efforts do you need to coordinate with, including linkages to the other subgroup topics?
- Identify which indicators have linkages for environmental justice, working landscapes, watershed management.

Questions specific to each strategic objectives and outcome indicator

- What are the strategic goals and objectives, and the narrative or quantitative performance goals and targets in the program documents related to this indicator?
- Document any conceptual models and quantitative models that identify drivers related to the outcome indicator, and also if there are additional conceptual (& quantitative) models for the drivers.
- Document what monitoring data exist for the outcome indicator and the driver indicators, and any information available about the data quality.
- Identify the significant data and information gaps and provide and short-term ballpark estimate of resources needed to complete monitoring, evaluation and reporting of this performance measure.

As of September 2006, core outcome indicators had been chosen for only one of CALFED's primary program objectives, *Water Supply Reliability*. The three indicators selected were: (1) Acre-feet of water made available and dedicated for Bay-Delta water quality and fish restoration improvements, (2) Ten year moving average of acre-feet of water delivered, and (3) Yearly unanticipated and uncompensated reductions in scheduled water delivery. Indicators under consideration for the water quality program objective include water quality at intakes (organic carbon, salinity, bromide, nutrients and pathogens), water quality at tap (disinfection byproducts, salinity, taste, and odor), toxicities to aquatic organisms, and mercury concentrations in biosentinel species and fish consumed by humans. Indicators under consideration for the *Levee System Integrity* program objective include the quantity of material required to prevent levee overtopping and a measure of levee anomalies and potential weak spots. No core indicators had been determined for the Ecosystem Restoration program objective at the time the Phase 1 Report was issued (CALFED 2006a).

A limited number of performance measures had been identified earlier (CALFED Bay-Delta Science Program[15]). Performance measures dated August 2003 include two indicators related to the Water Quality program objective, Drinking Water Quality – Bromide, and Drinking Water Quality – Organic Carbon. Six performance measure indicators for the Ecosystem Restoration program objective were identified: (1) Sacramento River Processes, (2) System-wide Central Valley Chinook Salmon, (3) Delta Smelt, (4) Fall-Run Chinook Salmon in the Tuolumne River, (5) Winter-Run Chinook Salmon in the Sacramento River, and (6) Spring-run Chinook Salmon in Butte Creek. "Acres Flooded" was identified as the single performance measure associated with the Levee System Integrity program objective. Information about each performance indicators can be obtained by downloading files from the CALFED Science Program webpage (http://science.calwater.ca.gov/library.shtml).

Generally, performance measure descriptions follow a similar format as illustrated below for Delta Smelt:

Ecosystem Restoration: Delta Smelt

- *What Is This Indicator and Why Is It Important?*
- *What Has Happened To Affect the Indicator?*
- *What Do the Data Show?*
- *Discussion*
- *Summary Data*
- *Conceptual Model*

Technical Note: Delta Smelt

- *The Indicator*
  - Goal:
  - The Data:
- *Longer-Term Science Needs*
- *Literature Cited*

[15] CALFED Bay-Delta Science Program. Accessed online January 29, 2007. http://science.calwater.ca.gov/library.shtml

Ohio Environmental Protection Agency

For over two decades the Ohio EPA[16] has used biological response indicators to measure changes and assess progress in improving water quality (http://www.epa.state.oh.us/dsw/document_index/psdindx.html). Central to the Ohio approach is the recognition that among measurable indicators of progress, only biological response indicators focus on the outcome of ecosystem improvement (Figure A4). Ohio's use of indicators has been successful because biological response indicators, along with other indicators, have been monitored in a systematic, standardized manner over a sustained period of time.

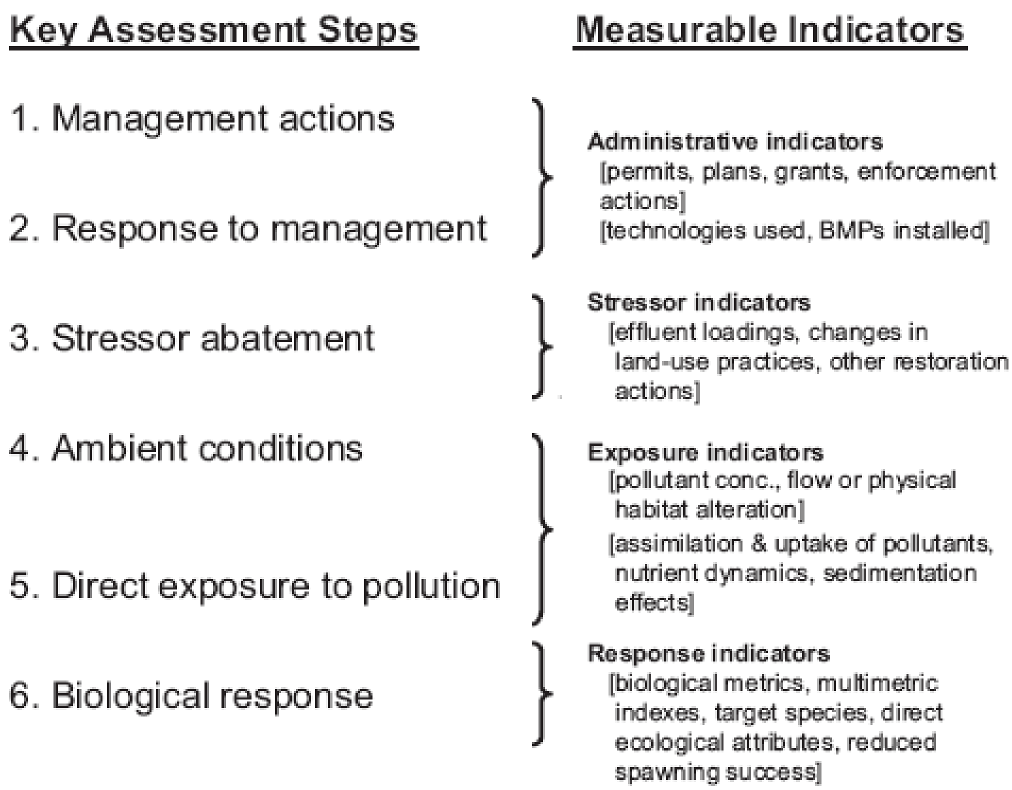

Figure A4. Hierarchy of monitoring and assessment indicators. All can be used to measure and manage environmental progress, but only biological responses focus on end outcome. Figure reproduced from Karr & Yoder (2004) with permission of the ASCE.

The effectiveness of this approach was demonstrated in a recent publication describing changes in fish assemblage status in Ohio's nonwadeable rivers in response to water pollution abatement and other water quality management programs (Yoder *et al.* 2005). Fish assemblage data collected at more than 8,000 sites in 1,750 rivers and streams since 1979 were used to calculate an IBI calibrated for Ohio rivers. IBI and river mileage data were used to calculate an Area of

---

[16] Ohio EPA. Division of Surface Water. Biological and Water Quality Report Index. Accessed online May 3, 2007. http://www.epa.state.oh.us/dsw/document_index/psdindx.html

Degradation Value (ADV) and an Area of Attainment Value (AAV) as quantitative indices of negative and positive deviations from target IBIs.

ADV and AAV were calculated as follows:

> ADV/AAV = Σ [(aIBIa + aIBIb) – (pIBIa + pIBIb)] *(RMa – RMb), for *a* = 1 to *n*, where:
>
> > aIBI*a* = actual IBI at river mile *a*,
> >
> > aIBI*b* = actual IBI at river mile *b*,
> >
> > pIBI*a* = IBI biocriterion at river mile *a*,
> >
> > pIBI*b* = IBI biocriterion at river mile *b*,
> >
> > RM*a* = upstream most river mile,
> >
> > RM*b* = downstream most river mile, and
> >
> > *n* = number of samples.

Calculated IBIs, ADVs and AAVs were used to quantitatively describe longitudinal and temporal changes in biological response to improved sewage treatment (Figure A5). The upper panel illustrates the measured and target IBI for the Scioto River downstream from Columbus, Ohio from 1979 to 1996. Box and whisker plots in the middle panel illustrate how IBIs have improved over time in response specific changes in sewage treatment. Temporal changes in ADV and AAV indices are shown in the lower panel.

Yoder *et al.* (2005) also documented the impact of changes in land management on biological response indicators (Figure A6). Upper and middle panels show how IBI and ADV/AAV for the Auglaize River improved over a sixteen year period with the adoption of conservation tillage or no tillage practices (lower panel).

In summarizing their research, Yoder *et al.* (2005) report that positive response in IBI and ADV/AAV were measured 4 to 5 years after implementing improved wastewater treatment, but that positive responses were less apparent in rivers influenced by complex industrial sources, agricultural nonpoint sources and extensive hydrologic modification.

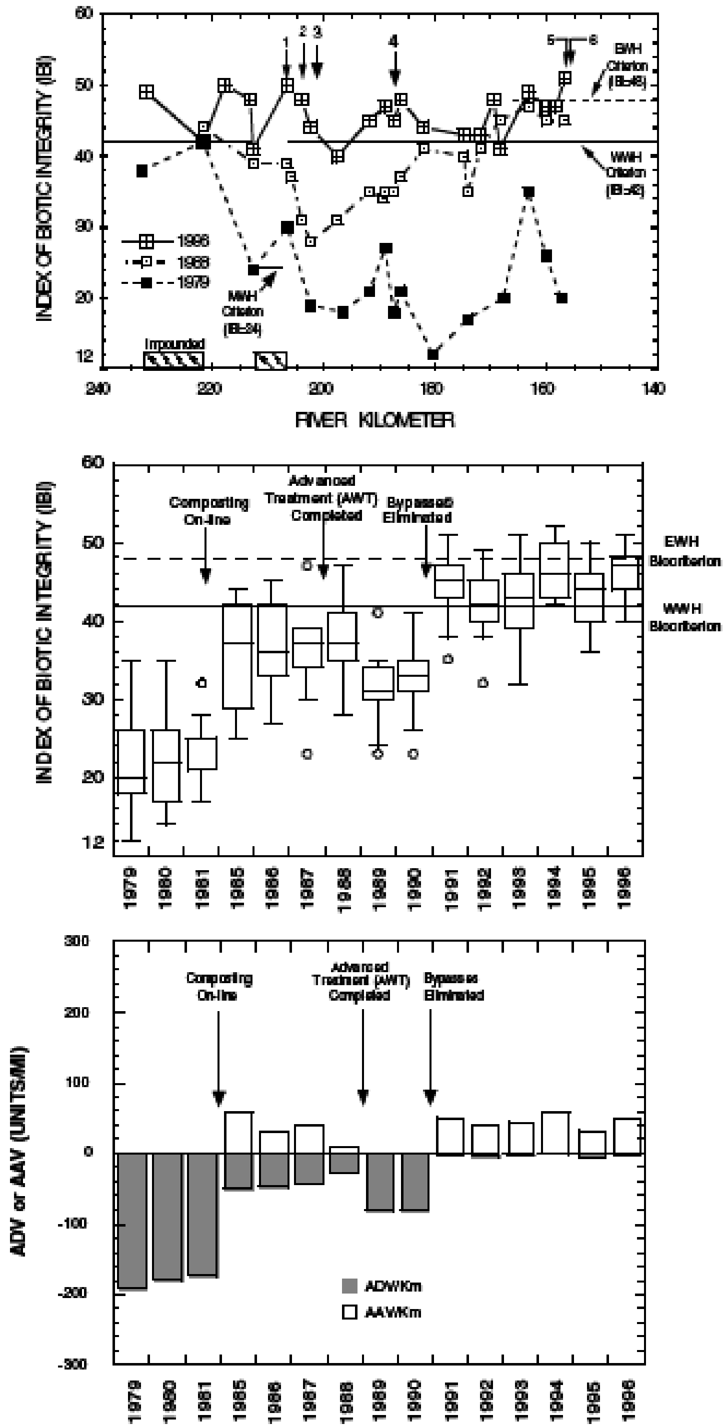

Figure A5. Longitudinal profile of IBI scores in the central Scioto River main stem in and downstream from Columbus, Ohio in 1979, 1988, and 1996 (upper panel). Annual IBI results from the central Scioto River main stem between 1979 and 1996 (middle panel), Area of Degradation Value (ADV), and Area of Attainment Value (AAV)/km during the same period (lower panel). Significant changes in the operation of the Columbus sewage treatment system are noted on each panel. (Figure reproduced from Yoder *et al.* 2005 with permission from the American Fisheries Society)

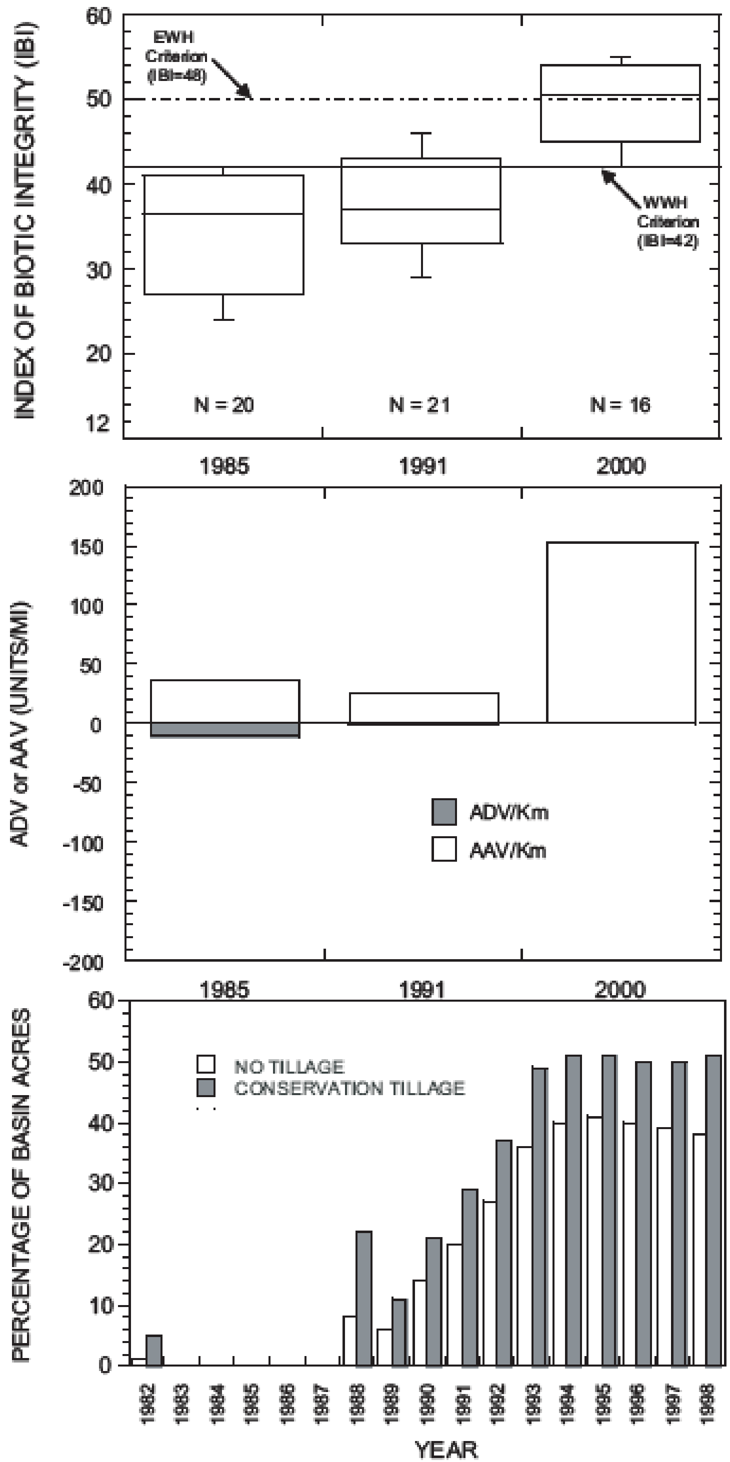

Figure A6. Box-and-whisker plot of IBI values in the Auglaize River main stem between Wapakoneta and Ft. Jennings, Ohio in 1985, 1991, and 2000 (upper panel; WWH = Warm water Habitat; EWH = Exceptional Warm water Habitat; N = number of samples). Area of Degradation Value (ADV) and Area of Attainment Value (AAV)/km for the same segment and years (middle panel). Percent of conservation tillage and no till acres in northwestern Ohio between 1982 and 1998 (lower panel). (Figure reproduced from Yoder *et al.* 2005 with permission from the American Fisheries Society).

References for Appendix A

CALFED Bay-Delta Program (2006a) *Indicators and Performance Measures Phase I Report: Core Indicators and Plan (9/7/06 draft)*. CALFED Bay-Delta Program, http://science.calwater.ca.gov/pdf/monitoring/monitoring_attachment_1_phase_1_report_091906.pdf.

CALFED Bay-Delta Program (2006b) *Framework for indicators for science, management and adaptive management in the CALFED Bay-Delta Program (1/5/06 and 4/28/06 drafts)*. CALFED Bay-Delta Program, http://science.calwater.ca.gov/monitoring/monitoring_framework.shtml.

CALFED Bay-Delta Program (2006c) *Indicators and Performance Measures. Appendix to Phase 1 Report: Core Indicators and Plan.* CALFED Bay-Delta Program, http://science.calwater.ca.gov/pdf/monitoring/monitoring_attachment_2_phase_1_appendix_091906.pdf

Government Accountability Office (GAO) (2003) *South Florida Ecosystem Restoration: Task Force Needs to Improve Science Coordination to Increase the Likelihood of Success*. Report to Subcommittee on Interior and Related Agencies, Committee on Appropriations, House of Representatives. GAO-03-345. US Government Accountability Office, Washington, DC.

Government Accountability Office (GAO) (2005) *Chesapeake Bay Program: Improved Strategies are Needed to Better Assess, Report, and Manage Restoration Progress*. Report to Congressional requesters. GAO-06-96. US Government Accountability Office, Washington, DC.

Government Accountability Office (GAO) (2007) *South Florida Ecosystem: Restoration Is Moving Forward but is Facing Significant Delays, Implementation Challenges, and Rising Costs*. GAO-07-520. US Government Accountability Office, Washington, DC.

Karr JR, Yoder CO (2004) Biological assessment and criteria improve total maximum daily load decision making. *Journal of Environmental Engineering* 130: 594–604.

Little Hoover Commission (2005) *Still Imperiled, Still Important: The Little Hoover Commission's Review of the CALFED Bay-Delta Program.* Little Hoover Commission, CA. 112 p. http://www.lhc.ca.gov/lhcdir/183/report183.pdf.

South Florida Ecosystem Restoration (SFER) (2006) *Report of the Review Panel Concerning Indicators for Restoration*. South Florida Ecosystem Restoration. http://www.sfrestore.org/scg/documents/indicators%20REPORT%20final%20from%20Jordan%20MAY%208%202006_w_line%20numbers.pdf.

South Florida Ecosystem Restoration (SFER)Task Force (2006) *Indicators for Restoration: South Florida Ecosystem Restoration*. September 5, 2006 draft report to the South Florida Ecosystem Restoration Task Force. South Florida Science Coordination Group, Florida International University, Miami, FL

Yoder CO, Rankin ET, Smith MA, Alsdorf BC, Altfater DJ, Boucher CE, Miltner RJ, Mishne DE, Sanders RE, Thomas RF (2005) Changes in fish assemblage status in Ohio's nonwadeable rivers and streams over two decades. In: *Historical Changes in Large River Fish Assemblages of the Americas* (eds Rinne JN, Hughes RM, Calamusso B). Symposium 45, American Fisheries Society, Bethesda, MD.

Appendix B. A Synthesis of indicator selection guidelines

INDICATOR GUIDELINE	SOURCE
Reflects the primary ecological and biochemical processes of the system	Harwell *et al.* 1999; NRC 2000; USEPA 2000; SFSCG 2006
Is applicable to most of the system	Harwell *et al.* 1999; NRC 2000; USEPA 2000; SFSCG 2006
Is based on an understood and accepted conceptual model (has a valid theoretical basis; information is sufficient to develop model) • Based on well-accepted scientific theory • Response to stressors or disturbance can be predicted • Signifies an impending change in key system characteristics • Predicts changes that can be averted by mgt. action	NRC 2000; Dale & Beyeler 2001; SFSCG 2006
Has a track record of past experience	NRC 2000; USEPA 2000; SFSCG 2006
Fits the applicable temporal and spatial scales, i.e., those scales at which indicator exhibits least stochastic variation, and weakest dependence on small changes in scale	NRC 2000; USEPA 2000
Is sufficiently sensitive to detect significant changes in variables of interest, including stressors; low variability in response; signal can be detected over system noise	Harwell *et al.* 1999; NRC 2000; USEPA 2000; Dale & Beyeler 2001; SFSCG 2006
Has reasonable data requirements; is detectable; is easily measured; variation can be estimated; feasible to implement; technologically able to be remeasured (repeatable)	NRC 2000; USEPA 2000; Dale & Beyeler 2001; SFSCG 2006
Is cost effective; contains the maximum amount of information per unit of cost or effort	Harwell *et al.* 1999; NRC 2000; USEPA 2000
Connects with real-world policy and management issues; reveals trends relevant to restoration goals and practices; is readily interpretable, able to be understood	USEPA 2000; SFSCG 2006
Able to establish clear, measurable targets	SFSCG 2006
Specific enough to determine needed management or corrective action	SFSCG 2006

References for Appendix B

Dale VH, Beyeler SC (2001) Challenges in the development and use of ecological indicators. *Ecological Indicators* 1: 3–10.

Harwell MA, Myers V, Young T, Bartuska A, Gassman N, Gentile JH, Harwell CC, Appelbaum S, Barko J, Causey B, Johonson C, McLean A, Smola R, Templet P, Tosini SA (1999) Framework for an ecosystem integrity report card. *BioScience* 49(7): 543–556.

NRC (2000) *Ecological Indicators for the Nation*. National Research Council, National Academy of Sciences, Washington, DC.

SFSCG (2006) *Indicators for Restoration: South Florida Ecosystem Restoration*. September 5, 2006 draft report to the South Florida Ecosystem Restoration Task Force. South Florida Science Coordination Group, Florida International University, Miami, FL.

USEPA (2000) *Evaluation Guidelines for Ecological Indicators*. EPA-620-R-99-005. US Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC.

Appendix C. Indices of Biological Integrity

Biological integrity was defined by Frey (1977) as "the capability of supporting and maintaining a balanced, integrated, adaptive community of organisms having a composition and diversity comparable to that of the natural habitats of the region." In other words, assemblages have integrity if they resemble those at natural or minimally disturbed reference sites in multiple ways, including their variability (Hughes *et al.* 1986; Hughes 1995; Stoddard *et al.* 2006). Various IBIs and predictive models have been developed and tested to do this for major aquatic assemblages (macroinvertebrates, algae, and fish). Predictive models of taxonomic richness are commonly used in the United Kingdom (Moss *et al.* 1987; Wright 1995; Bailey *et al.* 1998) and increasingly in the USA (Hawkins *et al.* 2000).

The IBI has been widely used and modified since 1981, including applications to fish (Karr *et al.* 1986; Simon & Lyons 1995; Miller *et al.* 1988; Hughes & Oberdorff 1999), macroinvertebrates (Kerans & Karr 1994; Klemm *et al.* 2003), algae (Hill *et al.* 2000, 2003; Fore 2003), and riparian birds (Bryce *et al.* 2002; Bryce 2006). Others have proposed using the IBI format, to create a terrestrial index of ecological integrity (Andreasen *et al.* 2001). IBIs have been shown to respond not only to water quality degradation, but to changes in physical habitat structure, flow regime, migration barriers, and energy source (Karr & Chu 1999). The USEPA and several states (e.g., Ohio, Utah, California, Maryland, Texas, Iowa, Florida, Kentucky) use IBIs to assess status and trends of surface waters at local and regional or statewide scales.

Macroinvertebrate and fish IBIs both vary regionally and by user. Macroinvertebrate IBIs typically combine measurements of total taxa richness, richness of major taxonomic groups, dominance by one to three taxa, and percent of individuals in various tolerance and trophic or feeding guilds (Plafkin *et al.* 1989; Kerans & Karr 1994; Karr & Chu 1999; USEPA 2006). Fish assemblage IBIs typically include metrics for total taxa richness, richness of major taxonomic groups, abundance, anomalies, non-native species, and various tolerance, habitat, trophic, reproductive, and life history guilds (Simon & Lyons 1995; Hughes & Oberdorff 1999). IBIs are widely used for assessing fish assemblage condition in cool and coldwater streams (Leonard & Orth 1986; Lyons *et al.* 1996; Moyle & Randle 1998; Mundahl & Simon 1999; McCormick *et al.* 2001; Hughes *et al.* 2004; Whittier *et al.* 2007), coldwater rivers (Hughes & Gammon 1987; Mebane *et al.* 2003), nationally (Miller *et al.*1988; Simon & Lyons 1995), and internationally (Hughes & Oberdorff 1999; Pont *et al.* 2006).

Salmonids also are key indicators of the condition of coldwater fish assemblages because they are sensitive to a number of stressors. Salmon are important in their own right as sport, commercial, and iconic species, and they tend to be more sensitive to most stressors than all resident salmonids with the exception of bull trout, which has very low tolerance for warm water. ODFW has various indicators of salmonid integrity including abundance, recruitment, distribution, and density of each species or population, and has used these metrics in combination to assess coastal coho status and trends.

References for Appendix C

Andreasen JK, O'Neill RV, Noss R, Slosser NC (2001) Considerations for the development of a terrestrial index of biological integrity. *Ecological Indicators* 1: 21–35.

Bailey RC, Kennedy MG, Dervish MZ, Taylor RM (1998) Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. *Freshwater Biology* 39: 765–774.

Bryce SA (2006) Development of a bird integrity index: Measuring avian response to disturbance in the Blue Mountains of Oregon, USA. *Environmental Management* 38: 470–486.

Bryce SA, Hughes RM, Kaufmann PR (2002) Development of a bird integrity index: using bird assemblages as indicators of riparian condition. *Environmental Management* 30: 294–310.

Fore LS (2003) Response of diatom assemblages to human disturbance: development and testing of a multi-metric index for the Mid-Atlantic Region (USA). In: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (ed Simon TP) CRC Press, Boca Raton, FL.

Frey DG (1977) Biological integrity of water-a historical approach. In: *The Integrity of Water* (eds Ballentine RK, Guarraia LJ) US Environmental Protection Agency, Washington, DC.

Hawkins CP, Norris RH, Hogue JN, Feminella JW (2000) Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications* 10: 1456–1477.

Hill BH, Herlihy AT, Kaufmann PR, DeCelles SJ, Vanter Borgh MA (2003) Assessment of streams of the eastern United States using a periphyton indicator of biological integrity. *Ecological Indicators* 2: 325–338.

Hill BH, Herlihy AT, Kaufmann PR, Stevenson RJ, McCormick FH, Johnson CB (2000) The use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19: 50–67.

Hughes RM (1995) Defining acceptable biological status by comparing with reference conditions. In: *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (eds Davis WS, Simon TP) Lewis Publishing, Boca Raton, FL.

Hughes RM, Gammon JR (1987) Longitudinal changes in fish assemblages and water quality in the Willamette River, Oregon. *Transactions of the American Fisheries Society* 116: 196–209.

Hughes RM, Oberdorff T (1999) Applications of IBI concepts and metrics to waters outside the United States and Canada. In: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (ed Simon TP) CRC Press, Boca Raton, FL.

Hughes RM, Howlin S, Kaufmann PR (2004) A biointegrity index for coldwater streams of western Oregon and Washington. *Transactions of the American Fisheries Society* 133: 1497–1515.

Hughes RM, Larsen DP, Omernik JM (1986) Regional reference sites: a method for assessing stream potentials. *Environmental Management* 10: 629–635.

Karr JR, Chu EW (1999) *Restoring Life in Running Waters: Better Biological Monitoring*. Island Press, Covelo, CA.

Karr JR, Fausch KD, Angermeier PL, Yant PR, Schlosser IJ (1986) *Assessing biological integrity in running waters: a method and its rationale*. Illinois Natural History Survey. Special Publication 5.

Kerans BL, Karr JR (1994) A benthic index biotic integrity (B-IBI) for rivers of the TennesseeValley. *Ecological Applications* 4: 768–785.

Klemm DJ, Blocksom KA, Fulk FA, Herlihy AT, Hughes RM (2003) Development and evaluation of a macroinvertebrate biotic integrity index (MBII) for regionally assessing Mid-Atlantic Highlands streams. *Environmental Management* 31: 656–69.

Leonard PM, Orth DJ (1986) Application and testing of an index of biotic integrity in small, coolwater streams. *Transactions of the American Fisheries Society* 115: 401–415.

Lyons J, Wang L, Simonson TD (1996) Development and validation of an index of biotic integrity for coldwater streams in Wisconsin. *North American Journal of Fisheries Management* 16: 241–256.

McCormick FH, Hughes RM, Kaufmann PR, Peck DV, Stoddard JL, Herlihy AT (2001) Development of an index of biotic integrity for the Mid-Atlantic Highlands region. *Transactions of the American Fisheries Society* 130: 857–877.

Mebane CA, Maret TR, Hughes RM (2003) An index of biological integrity (IBI) for Pacific Northwest rivers. *Transactions of the American Fisheries Society* 132: 239–261.

Miller DL, Leonard PM, Hughes RM, Karr JR, Moyle PB, Schrader LH, Thompson BA, Daniels RA, Fausch KD, Fitzhugh GA, Gammon JR, Halliwell DB, Angermeier PL, Orth DJ (1988) Regional applications of an index of biotic integrity for use in water resource management. *Fisheries* 13(5): 12–20.

Moss D, Furse MT, Wright JF, Armitage PD (1987) The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. *Freshwater Biology* 17: 41–52.

Moyle PB, Randle PJ (1998) Evaluating the biotic integrity of watersheds in the Sierra Nevada, California. *Conservation Biology* 12: 1318–1326.

Mundahl ND, Simon TP (1999). Development and application of an index of biotic integrity for coldwater streams of the upper Midwestern United States. In: *Assessing the Sustainability and Biological Integrity of Water Resources Using Fish Communities* (ed Simon TP) CRC Press, Boca Raton, FL.

Plafkin JL, Barbour MT, Porter KD, Gross SK, Hughes RM (1989) *Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish*. EPA-444-4-89-001. US Environmental Protection Agency, Washington, DC.

Pont D, Hugueny B, Beier U, Goffaux D, Melcher A, Noble R, Rogers C, Roset N, Schmutz S (2006) Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. *Journal of Applied Ecology* 43: 70–80.

Simon TP, Lyons J (1995) Application of the index of biotic integrity to evaluate water resource integrity in freshwater ecosystems. In: *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making* (eds Davis WS, Simon TP) Lewis Publishing, Boca Raton, FL.

Stoddard JL, Larsen DP, Hawkins CP, Johnson RK, Norris RH (2006) Setting expectations for the ecological condition of running waters: the concept of reference condition. *Ecological Applications* 16: 1267–1276.

US Environmental Protection Agency (USEPA) (2006) *Wadeable Streams Assessment: A Collaborative Survey of the Nation's Streams*. EPA-841-B-06-002. US Environmental Protection Agency, Washington, DC.

Whittier TR, Hughes RM, Lomnicky GA, Peck DV (2007) Fish and amphibian tolerance classifications, tolerance values, and an assemblage tolerance index for western USA streams and rivers. *Transactions of the American Fisheries Society* 136: 254–271.

Wright JF (1995) Development and use of a system for predicting macroinvertebrates in flowing waters. *Australian Journal of Ecology* 20: 181–197.

Appendix D. Comparison of multi-metric versus multivariate approaches to ecological indicator development.

Table D1. Review of pros (+) and cons (-) of three categories of methods for combining indicators. Table was reproduced from Rice & Rochet (2005; pages 524–525) with permission from Oxford University Press.

Methods for standardizing indicators

Scoring: Convert indicator values to scores (discrete variation; limited number of classes)
+: easy for qualitative variation
-: usually arbitrary for quantitative variation; no explicit scoring method available; huge scope for subjectivity (Rochet & Rice 2005)

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

Linear interpolation between observed extreme values: Scale all indicators on a common range (e.g., 0, 1), assuming linear variation between minimum and maximum values
+: simple
-: indicator may not show linear variation; sensitive to history of data series

Linear interpolation between reference values: Scale all indicators on a common range using predefined reference values
+: simple
-: linear variation not always relevant; reference values often difficult to define

Multivariate methods: Usually performed on normalized variance, hence indicator standardized by their standard deviations.
+: accounts for uncertainty and variability
-: sample dependent

Weighting methods

Multivariate methods: Projections on maximum inertia axes, so giving lower weight to correlated indicators
+: objective way of reducing redundancy without eliminating potentially useful indicators
-: management objective not taken into account

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 judgments at each level.
+: user-defined weighting
-: number of comparisons increases exponentially with number of indicators and potential values

Methods for combining indicators (graphical)

Kites (Garcia & Staples 2000): One standardized indicator per edge (outer rim = 'good'; center = '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 indicators); potential redundancy

Pie slices (Andreasen *et al.* 2001): One standardized indicator per slice; circumference = ‘degraded’ reference condition; indicator value shaded
+: quick and easy; not too many data manipulations; easy to understand
-: potential redundancy

Amoeba (Collie *et al.* 2001): Circle = reference; arrow lengths = values; arrow directions = correlations between indicators; shape influenced by relative variances

+: takes account of redundancy because based on indicator correlation
-: hard to display multiple indicators

Methods for combining indicators (indices)

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)

Weighted geometric average: Multiplying weighted indicator rather than summing into increase influence of 'bad' scores
+: simple
-: outcome determined by standardization and weights; hard to test weighting validity; prone to eclipsing (good traits may obscure bad ones)

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 indicators; eliminate redundant and inconsistent indicators based on correlations; measure variability in indicators and IBI using multiple sampling at each site and estimate power of IBI
+: 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 indicators

Fuzzy numbers (Tran *et al*. 2002): Normalize indicator with 0 (= ideal) and 1 (= undesirable) by linear interpolation; each normalized indicator with its observed minimum and maximum in a given site make a fuzzy number; compute fuzzy distance of each indicator 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 *a priori* basis, sensitive to assumed distribution

Framework for ecologically sustainable development (Chesson & Clayton 1998): Define hierarchical structure of assessment; standardize indicator (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

Methods for combining indicators (Multivariate ordination methods)

MDS of scored indicators (Pitcher & 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

PCA and canonical correlation analysis (Link *et al.* 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

Multivariate analysis (Charvet *et al*. 2000): Measure indicator in a set of communities; fuzzy scoring and correspondence analysis; hierarchical clustering; for each group, profiles of indicator (frequency distributions of mean scores); reference point possibly given by extreme situations
+: general advantages of multivariate methods
-: interpretation not always obvious (but possibly improved by CCA); not easy to understand

References for Appendix D

Andreasen JK, O'Neill RV, Noss R, Slosser NC (2001) Considerations for the development of a terrestrial index of biological integrity. *Ecological Indicators* 1: 21–35.

Charvet S, Statzner B, Usseglio-Polatera P, 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, Clayton H (1998) *A Framework for Assessing Fisheries with Respect to Ecologically Sustainable Development*. Report of the Bureau of Rural Sciences, Canberra.

Collie J, Gislason H, Vinther M (2001) Using AMOEBAs to integrate multispecies, multifleet fisheries advice. *ICES Document* CM 2001/T: 01.

Garcia SM, Staples DJ (2000) Sustainability reference systems and indicators for responsible marine capture fisheries: a review of concepts and elements for a set of guidelines. *Marine Freshwater Research* 51: 385–426.

Hughes RM, Kaufman PR, Herlihy AT, Kincaid TM, Reynolds L, Larsen DP (1998) A process for developing and evaluating indices of fish assemblages integrity. *Canadian Journal of Fisheries and Aquatic Science* 55: 1618–1631.

Link JS, Brodziak JKT, Edwards SF, Overholtz WJ, Mountain D, Jossi JW, Smith TD, Fogarty MJ (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.

McCormick FH, Hughes RM, Kaufmann PR, Peck DV, Stoddard JL, Herlihy AT (2001) Development of an Index of Biotic Integrity for the Mid-Atlantic Highlands region. *Transactions of the American Fisheries Society* 130: 857–877.

Pitcher T, Preikhost D (2001) RAPFISH: a rapid appraisal technique to evaluate the sustainability status of fisheries. *Fisheries Research* 49: 255–270.

Rice JC, Rochet MJ (2005) A framework for selecting a suite of indicators for fisheries management. *ICES Journal of Marine Science* 62: 516–527.

Tran LT, Knight CG, O'Neill RV, Smith ER, Riitters KH, Wickham J (2002) Fuzzy decision analysis for integrated environmental vulnerability assessment of the mid-Atlantic region. *Environmental Management* 29: 845–859.