

DEFINING AND MEASURING RIVER HEALTH

James R. Karr

University of Washington, Box 357980, Seattle, WA 98195-7980 USA

Running head: River health

SUMMARY

1. Society benefits immeasurably from rivers. Yet over the past century, humans have changed rivers dramatically, threatening river health. As a result, societal well-being is also threatened because goods and services critical to human society are being depleted.
2. “Health”—shorthand for “good condition” (e.g., “healthy” economy, “healthy” communities)—is grounded in science yet speaks to citizens.
3. Applying the concept of health to rivers is a logical outgrowth of scientific principles, legal mandates, and changing societal values.
4. Success in protecting the condition, or health, of rivers, depends on realistic models of the interactions of landscapes, rivers, and human actions.
5. Biological monitoring and biological endpoints provide the most integrative view of river condition, or river health. Multimetric biological indexes are an important and relatively new approach to measuring river condition.
6. Effective multimetric indexes depend on an appropriate classification system, the selection of metrics that give reliable signals of river condition, systematic sampling protocols that measure those biological signals, and analytical procedures that extract relevant biological patterns.
7. Communicating results of biological monitoring to citizens and political leaders is critical if biological monitoring is to influence environmental policies.
8. Biological monitoring is essential to identify biological responses to human actions. By using the results to describe the condition, or health, of rivers and their adjacent landscapes and to diagnose causes of degradation, we can develop "restoration" plans, estimate the ecological risks associated with land-use plans in a watershed, or select among alternative development options to minimize river degradation.

Introduction

Society benefits immeasurably from rivers. Urban centers and the world's most productive agricultural lands are tied to rivers. The energy of rivers generates electric power, and commerce depends on rivers for transportation of raw materials and manufactured goods. Fish and shellfish from rivers supply vital protein. Rivers carry water from the mountains to the sea, fueling the water cycle coupling land, ocean, and atmosphere. Human society depends totally on that cycle. In fact, human society exists only because water, resources associated with water, and the goods and services they provide are present and available in quantities that can support it.

Rivers throughout the world have changed dramatically during the past century because of what humans have done. But do those changes mean that people have degraded "river health"? Depends on whom you ask. To irrigators, rivers are healthy if there is enough water for their fields. For a power utility, rivers are healthy if there is enough water to turn the turbines. For a drinking-water utility, rivers are healthy if there is enough pure, or purifiable, water throughout the year. To sport or commercial fishers, rivers are healthy if there are fin-fish and shellfish to harvest. For recreationists, rivers are healthy if swimming, water skiing, or boating do not make people ill. But every one of these perceptions is only part of the picture. Each trivializes the other uses of the river—not to mention nonhuman aspects of the river itself— while assigning value only to its own desires. To protect all river uses and values, shouldn't we seek broader definitions of river health?

What Is Health?

Webster's dictionaries define *health* as a flourishing condition, well-being, vitality, or prosperity. A healthy person is free from physical disease or pain; a healthy person is sound in mind, body, *and* spirit. An organism is healthy when it performs all its vital functions normally and properly, when it is able to recover from normal stresses, when it requires minimal outside care. A country is healthy when a flourishing economy provides for the well-being of its citizens. An environment is healthy when the supply of goods and services required by both human and nonhuman residents is sustained. Healthy is a short-hand for good condition.

Despite—or perhaps because of—the simplicity and the breadth of this concept, the intellectual literature is rife with arguments on whether it is appropriate to use *health* in an ecological context. Is it appropriate to speak of "ecological health" or "river health"?

The arguments mounted against health as an ecologically useful concept go something like the following. Suter (1993) insists that the health metaphor is inappropriate because health is not an observable ecological property. According to Suter, health is a property of organisms, a position that acknowledges only the first, and narrowest, of the dictionary's definitions. Scrimgeour and Wicklum (1996) believe that no objective ecosystem state can be defined that is preferable to alternative states. Calow (1992) asserts that the idea of health in organisms involves different principles from the concept "as applied to ecosystems." He distinguishes between applying the concept in a weak form to signal normality (an expected condition) and in a strong

form to signal the existence of an active homeostatic process that returns disturbed systems to that normality. The latter, he suggests, requires a system-level control that does not exist in ecosystems. Neither does it exist in any dictionary definitions of health. Why, then, do Calow and Suter hold that characteristic to be central to a health concept in an ecological context?

"Societal values" also enter the discussion, sometimes as an essential, sometimes as an inappropriate consideration. Policansky (1993) and Wicklum and Davies (1995) contend that health is a "value-laden concept" and therefore inappropriate in science. Yet Rapport (1989) suggests that efforts to protect ecological health must consider "the human uses and amenities derived from the system." Regier (1993) and Meyer (1997) agree with Rapport about the importance of societal values in defining and protecting health. Regier speaks of integrity rather than health, suggesting that the concept integrity is "rooted in certain ecological concepts combined with certain sets of human values."

Other authors have searched for more objective or scientific arguments for referring to health in ecological contexts, often equating *health* with phrases such as "self-organizing," "resilient," and "productive." Haskell, Norton, and Costanza (1992) suggest that an ecosystem is healthy "if it is active and maintains its organization and autonomy over time and is resilient to stress." Costanza (1992) goes one step further, proposing an ecosystem health index as the product of system vigor (primary production or metabolism), organization (species diversity or connectivity), and resilience (the ability to resist or recover from damage). Those criteria do not seem scientifically defensible to me. Applying them, we would define oligotrophic lakes as less healthy than highly productive eutrophic lakes. Can we seriously argue that human actions that convert oligotrophic lakes to a eutrophic state would improve health? A tropical forest might be calculated as more healthy (more diverse and connected) than a spruce-fir forest. A community of sewage sludge worms (Tubificidae) at the outflow of a wastewater treatment plant would, by these criteria, be healthy because it is very resilient to additional disturbance.

Clearly, these criteria create their own problems that neither science nor values can resolve. Using maximum production as a measure of health is the analog of using gross national product as a measure of economic vitality; both measure only one aspect of ecological or economic health. Resilience of biological systems is difficult to define and even more difficult to measure (Karr and Thomas, 1996). Resilient to what? The term must be defined in the context of specific disturbances. A biota can sustain itself—it is very resilient—when faced with normal environmental variation, even when that variation is large (e.g., variation in flow in rivers). But the same biota may not be able to withstand even the smallest disturbance outside the range of its evolutionary experience. Does this concept add any objectivity to our concept of health?

In my view health as a word and concept in ecology is useful precisely because it is a concept all people are familiar with. It is not a huge intuitive leap from "my health" to "ecological health." Granted that, we must "operationalize" the term—define it and find ways to measure it—but as a policy goal, protecting the health and integrity of our landscapes and rivers has at least some chance of engaging public interest and support. Further, protecting biological or ecological "integrity" is the core principle of the US Clean Water Act, Canada's National Park

Act, and the Great Lakes Water Quality Agreement between the United States and Canada. Words like *health* and *integrity* are embedded in these laws *because* they are inspiring to citizens and a reminder to those who enforce the law to maintain a focus on the big picture, the importance of living systems to the well-being of human society.

I contend that we *can* define *health* and *integrity* to make the terms useful in understanding humans' relationship with their surrounding ecological systems. Integrity applies to sites at one end of a continuum of human influence, sites that support a biota that is the product of evolutionary and biogeographic processes (Fig. 1). This biota is a balanced, integrated, adaptive system having the full range of elements (genes, species, assemblages) and processes (mutation, demography, biotic interactions, nutrient and energy dynamics, and metapopulation processes) expected in the region's natural environment (Karr, 1991; Angermeier and Karr, 1994; Karr, 1996). This definition takes into account three important principles: (1) a biota spans a variety of spatial and temporal scales, (2) a living system includes items one can count (the elements of biodiversity) plus the processes that generate and maintain them, and (3) living systems are embedded in dynamic evolutionary and biogeographic contexts. This breadth is important because human society depends on and indeed values both elements and processes (structure and function) in these systems (contra Meyer, 1997).

As human activity changes biological systems, they—and we along with them—move along a gradient, ultimately to a state where little or nothing is left alive (see Fig. 1). Whether such a shift is acceptable to society is certainly a "value" decision—do we value the elements and processes that are lost?—but those decisions ought to be grounded in broad understanding of the consequences of loss. For ultimately, the loss of living systems means the loss of our own basis for existence. I would base those thresholds on two criteria (Karr, 1996). First, human activity should not alter the long-term ability of places to sustain the supply of goods and services those places provide. Second, human uses should not degrade off-site areas, a provision that requires a landscape-level perspective in modern decision making. Such criteria in decisions about environmental policy—from land use to setting fish harvest quotas—would avoid the depletion of living systems.

Two examples illustrate what can happen if environmental consequences are ignored in society's decision making process. Flood-control efforts on Florida's Kissimmee River created a canal that compromised local and regional natural resources in ways not accepted by many Florida citizens. Calls for restoration arose soon after the project was completed, and now—27 years later—a project to reverse the original channelization is underway. The goal is to restore the river and its connections with its floodplain to restore the biological integrity of the Kissimmee River landscape (Karr, 1994; Toth, 1993).

In Colorado, expanding irrigated agriculture has been valued for decades. Irrigation adds moisture and energy to the atmosphere, however, increasing humidity, moderating temperature extremes, and increasing convective storm activity (Rapport et al., in press). The resulting changes in regional heat flux transports more industrial and agricultural pollutants from the plains to the mountains, stressing fragile alpine and subalpine ecosystems with excessive nitrogen

deposition. What Coloradans may gain in agricultural production values, they stand to lose in biology of the Continental Divide, including perhaps the already tenuous water supplies for the cities on the front range. In both Florida and Colorado, decisions based on values have unwittingly compromised regional health and integrity. At the very least decisions about what is healthy for society should be more rigorously grounded in understanding of the long-term consequences for societal well-being.

What Is River Health?

The 1972 US Water Pollution Control Act Amendments [now called the Clean Water Act, sec. 101(a)] set a standard for answering this question: "The objective of this Act is to restore and maintain the chemical, physical, and biological integrity of the Nation's waters." By integrity, the Congress intended to "convey a concept that refers to a condition in which the natural structure and function of ecosystems is maintained," a conception that is explicit in Fig. 1. Arguing for passage of this milestone legislation, Senator Edmund S. Muskie (1972) of Maine asked: "Can we afford clean water? Can we afford rivers and lakes and streams and oceans, which continue to make life possible on this planet? Can we afford life itself? . . . These questions answer themselves." Senator Muskie understood that healthy rivers support living systems that are essential to human well-being. He also understood that "chronic biological impact may be a greater problem than the acute results of discharge of raw sewage or large toxic spills" (Muskie 1992).

Water bodies with integrity, especially rivers, have persisted in, even modified, their region's physical and chemical environment over millennia. The very presence of their natural biota means that they are resilient to the normal variation in that environment. Still, the bounds over which the system changes as a result of most natural events are narrow in comparison with the changes that result from human actions such as row-crop agriculture, timber harvest, grazing, or urbanization. Normal, or expected, conditions constituting integrity vary geographically because each river's biota evolves in the context of local and regional geology and climate and within the biological constraints imposed by the organisms with access to that region. Understanding this baseline must be the foundation for assessing change caused by humans. Only then can we take informed decisions in response to the question, Is this level of change acceptable?

When human activities within a watershed (catchment) are minimal, the biota is determined by the interaction of biogeographic and evolutionary processes. As human populations increase and technology advances, landscapes are altered in a variety of ways. Those changes alter the river's biota and thus the entire biological context of the river, causing it to diverge from "integrity." In some cases, the changes are minor. In others, they are substantial; they may even eliminate all or most of the plants and animals in a river. That much divergence from integrity is not healthy for humans or non-humans.

Goals, Models, and Actions

Consideration of river health or integrity rarely entered decision making by societies bent on conquering some frontier. Water was simply there, a potable liquid to be used. It was there to be allocated, to be consumed, and to be discarded, as likely as not carrying with it society's unwanted wastes. When the goal is to conquer, everything else is in the way. This attitude has threatened—and continues to threaten—the tenuous balance between water and human society, between rivers and the people who depend on rivers. Furthermore, certain human communities often exert power over other, often indigenous or otherwise economically powerless, communities with catastrophic consequences for culture, values, and human and ecological health (Donahue and Johnston, 1998).

Society—oblivious to either human-health or ecological risks of radically altering rivers—has chronically undervalued their biological components. We have behaved as if we could repair or replace any lost or broken parts of regional water resource systems, much as we replace toasters, cars, jobs, and even hearts or livers. This disregard has only worsened the lack of coherence in water law and in regulations regarding water use. The result in the US is a body of federal, state, and local law that fails to make the connections between water quality and quantity, surface water and groundwater, headwater streams and large rivers, and the living and nonliving components of aquatic ecosystems. This disconnectedness was one thing when there were few people living on a vast landscape; now it is quite another.

We need a new approach now, one based in new conceptual models of how rivers, landscapes, and human society interact. Mental models guide much that we do. But models--whether conceptual, physical, or mathematical--can be wrong when they make inappropriate assumptions or focus on the wrong endpoint. They can mislead when they contain inappropriate levels of detail, or they can be irrelevant if they do not apply to the real world. The first rule of modeling is to recognize that "all models are wrong, but some models are useful" (Anderson and Woessner, 1992). Models are most useful when they are routinely evaluated to determine if expectations are being met and if policies based on those models are accomplishing the goals of the society using those models.

In the United States, models for what ails rivers, and how to fix it, began with passage of the 1899 Refuse Act; the model then was to stop dumping raw sewage and oil into waterways. Successive generations of laws attempted to ensure that the human-waste-absorbing capacity of rivers was not exceeded. Several decades ago, the model changed to chemical contamination; rivers would be healthy if we just avoided discharging excessive toxic chemicals into them. The latest model seems to be watershed analysis: a more comprehensive approach to the interactions of landscapes, rivers, and humans. Each of these models is only as good as its ability to reflect the primary societal goals regarding water resources, and those goals, too, have been changing: from taking water for granted to "beneficial use" to protecting their integrity. The challenge before us now is to apply the more useful models and to make progress toward that goal.

A New Model

A new model should inform society not only about the condition of rivers and the landscapes they run through, but also about the lives of people living in those landscapes. That model should focus on biological endpoints as the most integrative measures of river health. Only biological monitoring integrates, and thus registers, the influence of all forms of degradation caused by human actions.

Physical, chemical, evolutionary, and ecological processes have interacted to produce rivers and their landscapes, including the local and regional biota (Fig. 2). Humans degrade biological integrity by altering physical habitat, modifying seasonal water flow, changing the system's food base, changing interactions among stream organisms, and contaminating the water with chemicals. These five factors provide a critical conceptual and analytical framework to judge the interactions of human activities and biological change (Karr, 1991).

By measuring biological condition and evaluating the result as a divergence from baseline biological integrity—biological monitoring, we can thus focus on the most integrative, biological endpoint. During the twentieth century, as knowledge and societal values changed, and human-imposed stresses became more complex and pervasive, biological monitoring evolved rapidly. At least two major approaches developed independently over the past 25 years.

One approach, the multimetric index, arose as an offshoot of basic research in aquatic ecology (Karr, 1981; Karr et al., 1986; Karr, 1991); the concept was adopted quickly by a variety of state (Ohio EPA, 1988) and federal (Plafkin et al., 1989) agencies and in geographic regions throughout the world (Oberdorff and Hughes, 1992; Minns et al., 1994; Davis and Simon, 1995; Rossano, 1996; Koizumi and Matsumiya, 1997; Thorne and Williams, 1997; Weisberg et al., 1997; Deegan et al., 1997; Harris this issue). Not all applications have been very effective. The USEPA version developed for use with invertebrates known as rapid bioassessment protocols (RBP), for example, has been less than successful because the metrics proposed for the original RBP were never adequately tested, and questionable statistical and analytical procedures were used. In a test of 10 standard RBP metrics used in Oregon (Fore et al., 1996), 6 failed under scrutiny according to the criteria for validating metrics that go into the indexes of biological integrity (IBI) for fish and invertebrates.

The other approach relies on multivariate statistical methods to discern pattern in taxonomic composition, often but not always at the family level. Examples include RIVPACS (Wright, 1995), AUSRIVAS (Parsons and Norris, 1996), BEAST (Reynoldson et al., 1995), and the aquatic life classification models used in Maine (Davies et al., 1995). Advocates of such approaches often strenuously criticize multimetric indexes, especially IBI, but their arguments indicate misunderstandings of many of the scientific and policy foundations of multimetric assessments (Karr and Chu, 1997b). Further, a multimetric IBI differs in many ways from the invertebrate RBP of Plafkin et al. (1989). The core principle of the multimetric IBI is to detect

divergence from biological integrity—the product of regional evolutionary and biogeographic processes—divergence attributable to human actions. The goal is not to document and understand all the variation that arises in natural systems. (Karr and Chu [1997b] present a full review.)

Effective multimetric biological indexes avoid indicators that are either theoretically or empirically flawed. They incorporate components of biological systems that are sensitive to a broad range of human actions (sedimentation, organic enrichment, toxic chemicals, flow alteration). Promising biological attributes are first identified to span the biological hierarchy from individual health to landscape dynamics. Before any attribute is included as a metric in the index, however, it is rigorously defined, measured, and tested. The result is an index that integrates the behavior of the elements and processes of biological systems. Common metrics include those that illustrate changes in taxa richness (biodiversity), shifts in species composition reflecting human effects (sedimentation or nutrient enrichment), individual health, food web organization, and other biological attributes that respond to human influence. Multimetric indexes thus integrate a number of dimensions of inherently complex systems. In this respect, they are similar to the indexes used to measure the health of regional and national economies (e.g., index of leading economic indicators or consumer price index in the United States).

Integrative multimetric biological indexes are well suited to judging river health against a defined goal or water quality standard (called "criteria" in the United States). These biological measures are more comprehensive and robust than chemical water quality standards; they are more effective at diagnosing degradation, defining its cause(s), and suggesting treatments to halt or reverse the damage. Furthermore, they can be used to evaluate the success of management decisions. Because most restoration efforts aim at explicitly biological goals (e.g., return of fish), biological endpoints can provide both guide and goal for ecological restoration.

Developing Multimetric Indexes

Five tasks are critical to the development and use of an effective multimetric biological index (Karr and Chu, 1997a, b):

1. Classifying environments to define homogeneous sets within or across regions (e.g., streams, lakes, or wetlands; large or small streams; warm-water or cold-water streams; high- or low-gradient streams).
2. Selecting measurable biological attributes that provide reliable and relevant signals about the biological effects of human activities.
3. Developing sampling protocols and designs that ensure that those biological attributes are measured accurately and precisely.
4. Devising analytical procedures to extract and understand relevant patterns in the those data.
5. Communicating the results to citizens and policymakers so that all concerned communities can contribute to environmental policy.

Classify to define homogeneous sets

Like a taxonomy of organisms, classification attempts to distinguish and group distinct environments, communities, or ecosystem types. The proper approach to classification may vary, however, according to specific goals. Hydrologists or geomorphologists may need a different river classification system than a biologist, for example, even though geophysical context is a fundamental determinant of variation in biological systems.

Classification based on the geomorphologists' view of stream channel types, or on other landforms occupied by biological systems, is not necessarily the proper level for assessing the biological condition of those systems. In the Pacific Northwest, geomorphologists identify some 50 to 60 channel types based on the interplay of physical and chemical processes that shape stream channels (MacDonald et al., 1991). But recognizing these channel types does not necessarily mean that an equal number of biological classes is needed. The taxonomic and ecological characteristics of the native biota may not, for example, be unique to each channel type. Further, even if some species replacement occurs, each assemblage may not need a special class. Fewer than 50 biological categories may therefore be effective for biological monitoring. For aquatic systems, biological classification (sometimes referred to as community classification) generally lags far behind classification by physical environment or habitat type for aquatic systems (Angermeier and Schlosser, 1995). Classification at levels appropriate for biological monitoring and assessment, classification that focuses on biological responses to human actions lags even more.

Excessive emphasis on classification, or inappropriate classification, can impede development of cost-effective and sensible monitoring programs. Using too few classes fails to recognize important distinctions among places; using too many unnecessarily complicates development of biological criteria. The challenge is to create a system with only as many classes as are needed to represent the range of relevant biological variation in a region and the level appropriate for detecting and defining the biological effects of human activity.

Another common error is classification based on a matrix of species and abundances, an approach that can obscure important natural history patterns. Many multivariate approaches classify narrowly according to species lists, often excluding rare taxa to avoid zeroes in the data matrix, for example. In this circumstance, mathematical and statistical tractability imposes decisions that diminish our ability to detect and understand biological signal. Using species-level community comparisons such as percentage similarity indexes can also be misleading. Regional classifications based on species overlap limit one's view by focusing on species composition rather than higher-level taxonomic and ecological structure (Karr and Chu, 1997b). Ecological organization and regional natural history are better guides.

Furthermore, no matter how much it enhances our knowledge of natural landscape variation, characterizing ecoregions should not get in the way of testing and using metrics diagnostic of human impact. Slavish adherence to geographically delineated ecological regions

defined by prevailing geophysical and climatic regimes (Omernik, 1995) is shortsighted. It makes little biological sense, for example, to group large, meandering stream reaches with small, fast-flowing streams even if they are in the same lowland ecoregion. The point of classification is to group places where the biology is similar in the absence of human disturbance and where the responses are similar after human disturbance. In some cases, these groupings may coincide with ecoregion boundaries; in others, they may cross those boundaries.

Thus, classification based on ecological dogma, on strictly chemical or physical criteria, or even on the logical biogeographical factors used to define ecoregions is not necessarily sufficient for biological monitoring. The good biologist uses the best natural history, biogeographic, and analytical information available to develop a classification system appropriate to the region.

Select appropriate metrics

Perhaps more than on anything else, successful application of a multimetric index depends on a rigorous process to identify and test metrics, the multiple biological attributes at the heart of multimetric analysis. Failure to properly define metrics can give incorrect signals about resource condition and lead to numerous errors in both science and management. Generally, multimetric indexes incorporate a richer array of signals than analyses based on species composition and abundance matrices (Karr and Chu, 1997b).

Selecting metrics for a multimetric index requires several important steps. First, sampling must cross sites with different intensities and types of human influence; that is, one must sample across a gradient of human disturbance. Without this essential step, how can one detect or understand biological responses to *human* influence? Second, biological monitoring must adhere to rigorous standards about what is measured and how those measurements are used. Knowledge of natural history and familiarity with ecological principles and theory guide the definition of attributes and predictions of how they will behave under varying human influences. But successful biological monitoring depends most on demonstrating that an attribute has a reliable empirical relationship—a consistent quantitative change—across a range, or gradient, of human influence.

Unfortunately, this crucial step is often omitted in many local, regional, and national efforts to develop multimetric indexes. Reputed tests of the effectiveness of multimetric and multivariate approaches (e.g., Reynoldson et al., 1997) are often inadequate because the original selection of metrics did not adhere to this core principle. This principle was compromised in the development and advocacy of rapid bioassessment protocols in the United States (Plafkin et al., 1989; see Karr and Chu, 1997b). "Rapid" is less important than accurate and effective if biologists are to contribute to water resource management.

Only a small number of biological attributes change consistently and quantitatively across a gradient of human influence. Graphs are particularly good for identifying these attributes because they force us to confront the obvious. A graph whose *y*-axis represents a biological response and whose *x*-axis is a range measure of human influence is the ecological analog of a toxicological dose-response curve (Fig. 3). These ecological dose-response curves show a

measured biological response to the cumulative ecological exposure, or "dose," of all events and human activities within a watershed. The number of unique native fish taxa in a midwestern stream sampled today, for example, reflects the cumulative effects of human influence up to the present. Graphs highlight idiosyncracies in patterns of data that, when examined closely, may give insight into the causes of a particular biological response. For example, one can explore whether unique situations exist at sites appearing as "outlying" points on a graph, which cause them to appear as outliers.

Too often, attempts to use, evaluate, or test multimetric indexes do not go through a rigorous selection process. As a result, some attributes have been retained in assessment programs despite widespread evidence that they do not give reliable signals of human effects. The study of populations has dominated much ecological research for decades, for example, so researchers assume that population size (expressed as abundance or density) provides a reliable signal about water resource condition. But because abundances vary so much as a result of natural environmental variation, even in pristine areas, population size is rarely a reliable indicator of human influence. Large numbers of samples (>25) were required, for example, to detect small (<20%) differences in number of fish per 100 m² of stream surface area in small South Carolina streams (Paller, 1995).

Similarly, responses of functional feeding groups of invertebrates are not good indicators of human disturbance (Karr and Chu, 1997b), even though such groups are good indicators for fish. Invertebrates probably do not always feed according to their assumed groups. Further, investigators assign invertebrates to functional feeding groups as often as not by guessing; better quantitative data are available for fish. Responses of functional feeding groups appear to vary with stream size and biogeographic region and, in addition, with kind of human activity (livestock grazing, row-crop agriculture, point-source pollution). As a result, consistency of pattern across studies for invertebrate functional feeding groups is much lower than for other attributes; it is these other attributes that become part of a multimetric benthic IBI (B-IBI; Table 1). The only functional feeding metric that seems moderately reliable is relative abundance of predators.

Failure to apply rigorous standards for defining metrics has derailed numerous efforts to develop, test, and use multimetric indexes. No vigilant medical community would permit the use of tests that had not been demonstrated to accurately diagnose a disease. The same rigor should be applied to choosing metrics for a multimetric index. One accepts or rejects proposed metrics by asking, Does this attribute vary systematically through a range of human influence? By selecting and organizing metrics systematically, an effective multimetric index can emerge from the chaos of biological attributes that can be measured (Karr and Chu, 1997b).

In short, the selection of biological signal used to detect the effects of human actions should use the insights visible in graphs and supplement those insights with thoughtful use of conventional statistics and knowledge of regional natural history. Consistently, reliable metrics include a number of taxa richness attributes (number of unique taxa in a sample, including rare ones) and percentages of individuals belonging to tolerant taxa. In study after study, the same major attributes give reliable signals of resource condition in different circumstances (Karr and

Chu, 1997b; Karr, 1998). As a result, every local, regional, or national project need not test and define its own locally applicable metrics. Scientists and resource managers can implement local biological monitoring and assessment programs based on the results of other studies.

Develop sampling protocols

Few topics provoke more arguments among field biologists than a claim that certain field methods to sample biological systems are "correct." After three decades of sampling a variety of organisms (birds, fish, plants, insects) in various environments (temperate and tropical forests, streams and rivers, wetlands, desert shrub), I find this issue less crucial to the effectiveness of multimetric indexes than the other steps in multimetric biological monitoring. A variety of sampling procedures, even samples based on different taxa, provide data of sufficient quality to make inferences about biological condition. The key is to define and use a protocol rigorously and apply appropriate analytical procedures to set metric-scoring criteria (Karr et al., 1986; Karr, 1991; Karr and Chu, 1997b) based on that method. Scoring criteria should be established for each sampling protocol or taxon.

Successful biological monitoring programs depend on accurate measures of a site's fauna or flora, especially those components influenced most by human disturbance. Thus the spatial and temporal scale of sampling should detect and foster understanding of human influences, not document the magnitude and sources of natural seasonal or successional variation in the same system.

In my view, it is better to do bugs right than to do bugs wrong. It is better to do fish right, than to do bugs wrong. The taxon selected is less important than what is done to detect signal about river condition. Moreover, sampling strategies vary among circumstances. For fish sampled in small streams, for example, it is better to sample across all habitats (pools, riffles, etc.) than to sample only specific habitats because the same gear would be used for all major macrohabitats in the former strategy. In very large rivers, where multiple sampling procedures may be required for fish, it is better to sample discrete habitats than to attempt to combine samples collected by methods of differing efficiency.

Finally, it is best to avoid composite samples from multiple habitats and sampling methods. Mixing insects collected across several macrohabitats, for example, yields samples of unknown heterogeneity. It is especially difficult to avoid different levels of heterogeneity from different sample teams or places when samples are composited, especially when samples from multiple habitats are to be collected based on general rules like collect "in proportion to the abundance of those habitats." Such judgement calls are an open invitation to create data interpretation problems.

For the past decade, my colleagues and I have been developing a stream benthic IBI to fulfill the promise never really attained by RBP. During that period, we have examined about a dozen invertebrate data sets from various sources collected by a variety of methods (Table 2). That work has made me cautious about absolutes. Generally, we have been more successful with

data from single habitats than with composite data from multiple habitats. Although most of the data came from riffles, pool samples yielded good indicators of human influence in watersheds (Kerans, Karr, and Ahlstedt, 1992). Sampling method (Surber, Hess, kicknet, Dendy) was less important than one might expect as long as rigorous procedures were applied throughout sampling and in selecting metrics and developing scoring criteria.

Still several general lessons did emerge over the years. All samples from a study should come from a relatively short time period (preferably one month); data from different seasons should not be mixed in a single analysis. Several factors influence the accuracy of an assessment and should be given adequate attention: definition of subsampling protocols and minimum subsample size; handling of replicate samples; and level of taxonomic identification. In the end, multimetric assessment can be robust to differences if data are handled well (Karr and Chu, 1997b).

Level of identification (family, genus, species) is an especially problematic issue. We have found that generic level is adequate for even the most technical studies and all 10 B-IBI metrics are useful with generic level identification. Family-level identification is a compromise when expertise, time, or funding is limited; only 5 of 10 B-IBI metrics are reliable, however. This is a reasonable compromise in some situations (e.g., citizen monitoring programs). Loss of discrimination of condition is compromised very little, but the ability to diagnose causes of degradation is compromised.

In sum, sampling protocols do affect the success of monitoring efforts and their ability to detect differences in human influence. For benthic invertebrates, subsampling, replicate sampling, and level of taxonomic identification affect the quality of data and the accuracy of assessments. Riffle habitats are typically sampled because they are easy to identify and functionally similar across streams. Type of sampling gear per se matters relatively little because standardized analysis methods can be applied reliably to each sampling technique. In many respects, the analytical protocol is more important than the field protocol to discover interpretable pattern. Thus, an organized and systematic approach to sample invertebrates, which applies reasonable quality control, yields data of sufficient resolution to detect the effects of human actions and diagnose causes of degraded river health.

Analyze data to reveal biological patterns

Multimetric biological monitoring should combine biological insight with statistical power in ways that enable us to understand how a resident biota has been altered by human actions. Regional biology and natural history—not a search for statistical relationships and significance (Stewart-Oaten, 1996)—should drive both sampling design and analytical protocol. We know much about the biology of rivers, and their responses to human activities. We should use that knowledge rather than defer to numerical pattern analysis. Simple graphs reveal, better than strictly statistical tools, relationships between biological attributes and human influence (Karr and Chu, 1997b). Graphs illustrate variation in responses to specific disturbances—among taxa and among

biological attributes chosen as metrics; they also reveal the direction and magnitude of change, for example, along a longitudinal transect down a stream.

Although statistics can and should be used to validate metric choices and predictions while building a multimetric index, excessive dependence on the outcome of statistical tests can obscure meaningful biological patterns. Too often, a narrow focus on p -values rather than on biological consequences limits the value of biological monitoring (Stewart-Oaten, Murdoch, and Parker, 1986; Stewart-Oaten, Bence, and Osenberg, 1992; Stewart-Oaten, 1996). Dependence on narrow statistical approaches overlooks the fact that a statistically significant result (small p -value) may not equate with a large important effect, as researchers often assume; similarly, a statistically insignificant effect (large p -value) may well be biologically important (Yoccoz, 1991; Stewart-Oaten, 1996).

That said, much is known about the statistical properties of multimetric indexes (Fore, Karr, and Conquest 1994; Karr and Chu 1997b). They are statistically versatile and amenable to application of familiar statistical tests (e.g., t-tests or analysis of variance). From statistical power analysis, a properly formulate IBI can detect six distinct categories of resource condition. Finally, at both the individual metric and index level, analyzing the components of variance in a data set can refine sampling protocols and thus improve the inferences to be made from biological monitoring and assessment. Whether the goal is regulatory or to evaluate where restoration funds can most usefully be spent, multimetric IBI provides an analytical tool with considerable potential to guide environmental decisions.

Communicate biological condition

What good is the most rigorous analysis if it cannot be communicated? Communicating the condition of biological systems, and the consequences of human activities to those systems, is the ultimate purpose of biological monitoring. Effective communication can transform biological monitoring from a scientific exercise into an effective tool for environmental decision making. Politics plays an enormous role in environmental policy decisions; how can scientists hope to affect those decisions if they cannot communicate effectively to the decision makers?

Of course biologists must extend what they have learned about monitoring in fresh water to other environments and other taxonomic groups. But they must also avoid gathering and becoming overwhelmed by too much information. Like any scientific method, biological monitoring generates many new and interesting questions, methods, and refinements. But scientists and managers need to realize that they already know enough about how biological systems respond to human influence to make decisions that will halt the decline of rivers. Managers must *use* what they already know.

With multimetric indexes that explain biological condition in numbers and words, biologists can make use of what they know, now. By talking and writing well beyond the confines of academic journals, they can root out the call for more research and call instead for widespread understanding of condition and trends in river health. People need, want, and deserve to understand these issues.

In the end, biologists are themselves partly to blame for the gulf between science and policy making for river health. Biologists, managers, regulators, and decision makers cannot protect river health if they cannot break away from thinking in regulatory dichotomies or continually equate "habitat" with "inhabitants." Too often restoration efforts focus on physical and chemical processes rather than biological context. In the end, a healthy river is a living river. Failing to recognize this essential principle is to fail, as when the focus is on much desired "functions," rather than biological condition. Even when restoration focuses on keystone or indicator taxa, it often fails the biological endpoint test. In the end, a healthy river is a living river. Failing to recognize this essential principle is to fail our rivers and, ultimately, our own health.

Acknowledgments

This paper was prepared with support from the Consortium for Risk Evaluation with Stakeholder Participation (CRESP) by Department of Energy Cooperative Agreement #DE-FC01-95EW55084.S and was aided by grants from the US Environmental Protection Agency. Ellen Chu provided invaluable advice that substantially improved the paper.

References

- Anderson, M. P., and W. W. Woessner. (1992) *Applied Groundwater Modeling: Simulation of Flow and Advective Transport*. Academic Press, San Diego, CA.
- Angermeier, P. L., and J. R. Karr. (1994) Biological integrity versus biological diversity as policy directives. *Bioscience*, **44**, 690–697.
- Angermeier, P. L., and I. J. Schlosser. (1995) Conserving aquatic biodiversity. *American Fisheries Society Symposium*, **17**, 402–414.
- Calow, P. (1992) Can ecosystems be healthy? Critical consideration of concepts. *Journal of Aquatic Ecosystem Health*, **1**, 1–5.
- Costanza, R. (1992) Toward an operational definition of ecosystem health. *Ecosystem Health: New Goals for Environmental Management*. (eds R. Costanza, B. G. Norton, and B. D. Haskell), pp. 239–256. Island Press, Washington.
- Davies, S. P., L. Tsomides, D. L. Courtemanch, and F. Drummond. (1995) Maine biological monitoring and biocriteria development program. Maine Department of Environmental Protection, Bureau of Land and Water Quality, Division of Environmental Assessment, Augusta.
- Davis, W. S., and Simon, T. P., eds. (1995) *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. Lewis Publishers, Boca Raton, FL.

- Deegan, L. A., J. T. Finn, S. G. Ayvazian, C. A. Ryder-Kieffer, and J. Buonaccorsi. (1997) Development and validation of an estuarine biotic integrity index. *Estuaries*, **20**, 601–617.
- DeShon, J. E. (1995) Development and application of the invertebrate community index (ICI). *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. (eds W. S. Davis and T. P. Simon), pp. 217-244. Lewis Publishers, Boca Raton, FL.
- Donahue, J. M. and B. R. Johnston, eds. (1998) *Water, Culture, and Power: Local Struggles in a Global Context*. Island Press, Washington.
- Fore, L. S., J. R. Karr, and L. L. Conquest. (1994) Statistical properties of an index of biotic integrity used to evaluate water resources. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1077–1087.
- Fore, L. S., Karr, J. R., and Wisseman, R. W. (1996) Assessing invertebrate responses to human activities: Evaluating alternative approaches. *Journal North American Benthological Society*, **15**, 212–231.
- Haskell, B. D., B. G. Norton, and R. Costanza. (1992) What is ecosystem health and why should we worry about it? *Ecosystem Health: New Goals for Environmental Management*. (eds R. Costanza, B. G. Norton, and B. D. Haskell), pp. 3–20. Island Press, Washington.
- Karr, J. R. (1981) Assessment of biotic integrity using fish communities. *Fisheries*, **6(6)**, 21–27.
- Karr, J. R. (1991) Biological integrity: A long-neglected aspect of water resource management. *Ecological Applications*, **1**, 66–84.
- Karr, J. R. (1994) Landscapes and management for ecological integrity. *Biodiversity and Landscapes: A Paradox of Humanity*. (eds K. C. Kim and R. D. Weaver), pp. 229–251. Cambridge University Press, Cambridge.
- Karr, J. R. (1996) Ecological integrity and ecological health are not the same. *Engineering within Ecological Constraints* (ed P. Schulze). pp. 100–113. National Academy Press, Washington, DC.
- Karr, J. R. (1998) Rivers as sentinels: Using the biology of rivers to guide landscape management. *The Ecology and Management of Streams and Rivers in the Pacific Northwest Coastal Ecoregion*. (eds R. E. Bilby and R. J. Naiman). pp. xxx–xxx. Springer-Verlag, New York.
- Karr, J. R. and E. W. Chu. (1997a) Biological monitoring: essential foundation for ecological risk assessment. *Human and Ecological Risk Assessment*, **3**, 993-1004.

- Karr, J. R. and E. W. Chu. (1997b) *Biological Monitoring and Assessment: Using Multimetric Indexes Effectively*. EPA 235-R91-001. University of Washington, Seattle.
- Karr, J. R., and T. Thomas. (1996) Economics, ecology, and environmental quality. *Ecological Applications*, **6**, 31-32.
- Karr, J. R., Fausch, K. D., Angermeier, P. L., Yant, P. R., and Schlosser, I. J. (1986) Assessing biological integrity in running waters: A method and its rationale. *Illinois Natural History Survey Special Publication 5*. Champaign, IL.
- Kerans, B. L., and J. R. Karr. (1994) A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications*, **4**, 768-785.
- Kerans, B. L., J. R. Karr, and S. A. Ahlstedt. (1992) Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. *Journal of the North American Benthological Society*, **11**, 377-390.
- Koizumi, N., and Y. Matsumiya. (1997) Assessment of fish habitat based on index of biotic integrity. *Bulletin of the Japanese Society of Fisheries Oceanography*, **61**, 144-156.
- MacDonald, L. H., A. Smart, and R. C. Wissmar. (1991) Monitoring guidelines to evaluate effects of forestry activities on streams in the Pacific Northwest and Alaska. EPA/910/9-91-001. US Environmental Protection Agency, Seattle.
- Meyer, J. L. (1997) Stream health: incorporating the human dimension to advance stream ecology. *Journal of the North American Benthological Society*, **16**, 439-447.
- Minns, C. K., Cairns, V. W., Randall, R. G., and Moore, J. E. (1994) An index of biotic integrity (IBI) for fish assemblages in the littoral zone of Great Lakes' areas of concern. *Canadian Journal of Fisheries and Aquatic Sciences*, **51**, 1804-1822.
- Muskie, E. S. (1972) Senate consideration of the report of the Conference Committee, October 4, 1972. Amendment of the Federal Water Pollution Control Act. US Government Printing Office, Washington, DC.
- Muskie, E. S. (1992) Testimony of Edmund S. Muskie before the Committee on Environment and Public Works, on the Twentieth Anniversary of Passage of the Clean Water Act. September, 22, 1992. Reprinted S. Doc. 104-17; Memorial Tribute Delivered in Congress. Edmund S. Muskie, 1914-1996. US Government Printing Office, Washington, DC.
- Oberdorff, T., and Hughes, R. M. (1992) Modification of an index of biotic integrity based on fish assemblages to characterize rivers of the Seine-Normandie basin, France. *Hydrobiologia*, **228**, 117-130.

- Ohio EPA (Environmental Protection Agency). (1988) *Biological Criteria for the Protection of Aquatic Life*, vol. 1–3. Columbus, Ecological Assessment Section, Division of Water Quality Monitoring and Assessment, Ohio EPA.
- Omernik, J. M. (1995) Ecoregions: A spatial framework for environmental management. *Biological Assessment and Criteria: Tools for Water Resource Planning and Decision Making*. (eds W. S. Davis and T. P. Simon), pp. 49–62. Lewis, Boca Raton, FL.
- Paller, M. H. (1995) Interreplicate variance and statistical power of electrofishing data from low-gradient streams in the southeastern United States. *North American Journal of Fisheries Management*, **15**, 542–550.
- Parsons, M., and R. H. Norris. (1996) The effect of habitat-specific sampling on biological assessment of water quality using a predictive model. *Freshwater Biology*, **36**, 419–434.
- Plafkin, J. L., Barbour, M. T., Porter, K. D., Gross, S. K., and Hughes, R. M. (1989) Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish. EPA/440/4-89-001. Washington, DC, Assessment and Water Protection Division, US Environmental Protection Agency.
- Policansky, D. 1993. Application of ecological knowledge to environmental problems: ecological risk assessment. *Comparative Environmental Risk Assessment*. (ed in C. Cothorn), pp. 37–51. Lewis Publishers, Boca Raton, FL.
- Rapport, D. J. (1989) What constitutes ecosystem health? *Perspectives in Biology and Medicine*, **33**, 120-132.
- Rapport, D. J., C. Gaudet, J. R. Karr, J. S. Baron, C. Bohlen, W. Jackson, B. Jones, R. J. Naiman, B. Norton, and M. M. Pollock. (1998) Evaluating landscape health: integrating societal goals and biophysical processes. *Journal of Environmental Management*, in press.
- Regier, H. A. (1993) The notion of natural and cultural integrity. *Ecological Integrity and the Management of Ecosystems* (eds S. Woodley, J. Kay, and G. Francis), pp. 3–18. St. Lucie Press, Delray Beach, FL.
- Reynoldson, T. B., R. C. Bailey, K. E. Day, and R. H. Norris. (1995) Biological guidelines for freshwater sediment based on Benthic Assessment of SedimenT (the BEAST) using a multivariate approach for predicting biological state. *Australian Journal of Ecology*, **20**, 198-219.
- Reynoldson, T. B., R. H. Norris, V. H. Resh, K. E. Day, and D. M. Rosenberg. (1997) The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. *Journal of the North American Benthological Society*, **16**, 833–852.

- Rossano, E. M. (1996) *Diagnosis of Stream Environments with Index of Biological Integrity*. (In Japanese and English.) Museum of Streams and Lakes, Sankaido Publishers, Tokyo, Japan. ISBN 4-381-00868-5.
- Scrimgeour, G. J. and D. Wicklum. (1996) Aquatic ecosystem health and integrity: problems and potential solutions. *Journal of the North American Benthological Society*, **15**, 254–261.
- Stewart-Oaten, A. (1996) Goals in environmental monitoring. *Detecting Ecological Impacts: Concepts and Applications in Coastal Habitats*, (eds R. J. Schmitt, and C. W. Osenberg), pp. 17–28. Academic Press, San Diego, CA.
- Stewart-Oaten, A., Murdoch, W. W., and Parker, K. R. (1986) Environmental impact assessment: Pseudoreplication in time? *Ecology*, **67**, 929–940.
- Stewart-Oaten, A., Bence, J. R., and Osenberg, C. W. (1992) Assessing effects of unreplicated perturbations: No simple solutions. *Ecology*, **73**, 1396–1404.
- Suter, G. W. (1993) A critique of ecosystem health concepts and indexes. *Environmental Toxicology and Chemistry*, **12**, 1533–1539.
- Toth, L. A. (1993) The ecological basis of the Kissimmee River restoration plan. *Florida Scientist*, **56**, 25–51.
- Weisberg, S. B., J. A. Ranasinghe, L. C. Schaffner, R. J. Diaz, D. M. Dauer, and J. B. Frithsen. (1997) An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries*, **20**, 149–158.
- Wicklum, D. and Davies, R. W. (1995) Ecosystem health and integrity? *Canadian Journal of Botany*, **73**, 997–1000.
- Wright, J. F. (1995) Development and use of a system for predicting the macroinvertebrate fauna in flowing waters. *Australian Journal of Ecology*, **20**, 181–197.
- Yoccoz, N. G. (1991) Use, overuse, and misuse of significance tests in evolutionary biology and ecology. *Bulletin Ecological Society of America*, **71**, 106–111.

Figure Captions

Figure 1. At one end of a continuum of human influence on biological condition, severe disturbance eliminates all life; at the other end of the gradient are "pristine," or minimally disturbed, living systems (top). A parallel gradient (bottom) from integrity toward nothing alive passes through healthy, or sustainable, condition or activities. Below a threshold defined by specific criteria (see text), the conditions or activities are no longer healthy or sustainable in terms of supporting living systems.

Figure 2. Relationships among the elements and processes of natural systems, the kinds of changes that occur as a result of human actions, and a framework of environmental policies that might come from an assessment of biological condition, the endpoint of primary concern to society.

Figure 3. Average taxa richness of mayflies (Ephemeroptera) plotted against percentage of impervious surface area surrounding Puget Sound lowland streams. Note the general dose-response curve relationship: as human influence (impervious area) increases, taxa richness declines. Sites B and C had relatively intact riparian areas (wetlands); site A is downstream of a coal mine (no longer active) that continues to leak contaminants or sediment into the stream.