Freshwater Biology (1999) **41**, 221–234

Defining and measuring river health

JAMES R. KARR
University of Washington, Box 357980, Seattle, WA 98195–7980, U.S.A.

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 indices are an important and relatively new approach to measuring river condition.
6. Effective multimetric indices 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.

Keywords: river health, society, ecology, IBI, multimetric index.

Introduction

Society benefits immeasurably from rivers. Yet over the past century, humans have changed rivers dramatically. Do those changes mean that people have degraded river health? The answer 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 non-human aspects of the river itself — while assigning value only to its own desires. To protect all river uses and values, should we not seek broader definitions of river health?

Correspondence: J. R. Karr, University of Washington, Box 357980, Seattle, WA 98195–7980, U.S.A.
E-mail: jrKarr@u.washington.edu

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 non-human residents is sustained. 'Health' is shorthand for 'good condition'.

Despite — or perhaps because of — the simplicity and the breadth of this concept, the 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?

Several arguments are advanced against 'health' as an ecologically useful concept. 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 definitions. Scrimgeour & Wicklum (1996) believe that no preferred ecosystem state can be objectively defined. 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. Yet no such system-level control exists in any dictionary definitions of health. Why, then, do Calow and Suter hold that system-level control is central to the concept of health in an ecological context?

Societal values also enter the discussion, sometimes as an essential consideration, sometimes as an inappropriate consideration. Policansky (1993) and Wicklum & 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 suggests that health as a concept, although he uses the word 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 terms such as self-organizing, resilient, and productive. Haskell, Norton & 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 vigour (a cardinal measure of system activity, metabolism, or primary productivity), organization (species diversity and connectivity), and resilience (the ability to resist or recover from damage). Further, Costanza makes assumptions about the interactions of these variables. He suggests, for example, that eutrophication represents an increase in metabolism that is outweighed by decrease in organization or resilience. I am not convinced of the inevitability of these connections.

These constructs remind me of the often-heard admonition that unless we can tie something to an ecological theory it cannot or should not be used. I contend that much that we have concluded on the basis of theory — such as the interactions of system vigour, organization, and resilience — has not been empirically verified. In many circumstances, excessive reliance on theory misleads ecology (the science). More important, if those constructs are translated to the environmental policy arena without adequate validation, society will be misled in much the same way that economists have misled themselves and society when they failed to deal with externalities, environmental or social. Too much conventional neoclassical economic theory is simply at odds with reality (Prugh *et al*., 1995). Many ecologists recognize this fact and are working to expand the perspectives of economic theory. Unfortunately, too many ecologists are reluctant to use the same rigour in their exploration of ecological theory. At the very least, theoretical constructs developed by ecologists should be validated in the real world before they are used to guide public policy.

Taken singly, the components of health and the inevitability of their connections, as proposed by Costanza and colleagues (Costanza, 1992; Mageau, Costanza & Ulanowicz, 1995), seem to me neither logical nor scientifically defensible. Until they are extensively validated they should not be used as guides to public policy designed to protect ecological health. If we were to apply them, we would define oligotrophic lakes (low production and diversity) as less healthy than highly productive and diverse eutrophic lakes. A tropical forest might be calculated as more healthy (more diverse and connected with higher primary production) 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.

The use of the product of these three measures as a measure of health also deserves more careful scrutiny. Using maximum production as a measure of health is the analogue 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 & Thomas, 1996). Resilient to what? The term must be defined in the context of specific disturbances. A biota can sustain itself — it is resilient — when faced with normal environmental variation (that is, within the range of its evolutionary experience), 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 (extreme flows out of season). Does this concept of resilience 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, the protection of 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 in ways that will operationalize the terms, using them to help us understand humans' relationship with their surroundings. The term 'integrity' applies to the condition of places at one end of a continuum of human influence: those that support a biota that is the product of evolutionary and biogeographic processes with minimal influence from modern human society (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 areas with minimal influence from modern human society (Karr, 1991, 1996; Angermeier & Karr, 1994). This definition takes into account three important principles: (i) a biota spans a variety of spatial and temporal scales, (ii) a living system includes an array of kinds of things (the elements of biodiversity) plus the processes that generate and maintain them, and (iii) living systems are embedded in dynamic evolutionary and biogeographic contexts. This breadth is important because human society depends both on elements and on processes (structure and function) in these systems. Moreover, society values the elements and the processes of natural systems (*contra* Meyer, 1997).

As human activity changes biological systems, they, and we with them, move along a gradient, ultimately to a state where little or nothing is left alive (see

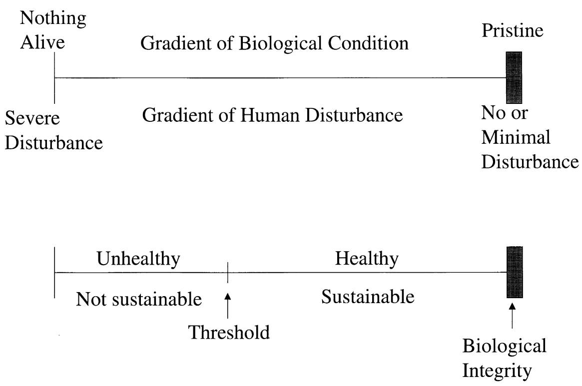

Fig. 1 At one end of a continuum of human influence on biological condition, severe disturbance eliminates all life; at the other end of the gradient are pristine, or minimally disturbed, living systems (top). A parallel gradient (bottom), from integrity towards nothing alive, passes through healthy, 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.

Fig. 1). Whether or not 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 that those places provide. That is, agriculture should not deplete soil or water so that agriculture cannot be sustained; industrial sites should not become so polluted that people cannot continue to work at those sites. Second, human uses should not degrade other areas (e.g. downstream or downwind), 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, 28 years later, a project to reverse the original channelization is underway. The explicit goal is to restore the river and its connections with its floodplain to restore the biological integrity of the Kissimmee River landscape (Toth, 1993; Karr, 1994).

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.*, 1998). The resulting change in regional heat flux transports more industrial and agricultural pollutants from the plains to the mountains, stressing alpine and subalpine ecosystems by excessive nitrogen deposition. Those ecosystems are the foundation of a lucrative tourist industry and the source of water for cities at the base of the mountains. What Coloradans may gain in agricultural production values, they stand to lose in biology of the Continental Divide, including perhaps the already dwindling water supplies for its cities. In both Florida and Colorado, decisions based on values have unwittingly compromised regional natural systems.

What is river health?

The 1972 US Water Pollution Control Act Amendments (now called the Clean Water Act, section 101(a)]) set a standard for answering the question, 'What is river health?' It said: '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 concept that is explicit in Fig. 1. Arguing for passage of this 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), a perspective that is also widely understood according to a national survey of 400 experts, in fields ranging from botany to genetics, conducted by the American Museum of Natural History (Anonymous, 1998). Those scientists identified loss of biodiversity as a more serious problem than ozone depletion, global warming, or pollution and contamination.

Water bodies with integrity, especially rivers, have persisted in, even modified, their regions' physical and chemical environments over millennia. The very presence of their natural biota means that the biota is resilient to the normal variation in that environment. Still, the bounds over which the system changes from its normal or expected condition 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 constraints and opportunities. Understanding this baseline must be the foundation for assessing change caused by humans. Only then can we make informed decisions in response to the questions, 'Is this level of change acceptable? Are the landscape and its rivers healthy?'

When human activities within a watershed (catchment) are minimal, the biota is determined by the interaction of biogeographic and evolutionary processes in the regional climatic and geological context. 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.

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 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 & Johnston, 1998).

Society, oblivious either to human-health risks or to the ecological risks of radically altering rivers, has chronically undervalued rivers' 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 USA 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 non-living 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, one based on new conceptual models of how rivers, landscapes, and human society interact. In the USA, models for what ails rivers, and how to protect or restore rivers, 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 would just avoid 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 biological integrity. The challenge before us now is to apply the more useful models and to make progress towards actual protection.

A new model

A new model should inform society not only about the conditions of rivers and their landscapes, 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. A comprehensive approach to biological monitoring permits a new level of integration because living systems as a monitoring focus are likely to register the influences 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 alter the biological systems in a river by altering physical habitat, modifying seasonal water flow, changing the system's food base, changing interactions among stream organ isms, 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

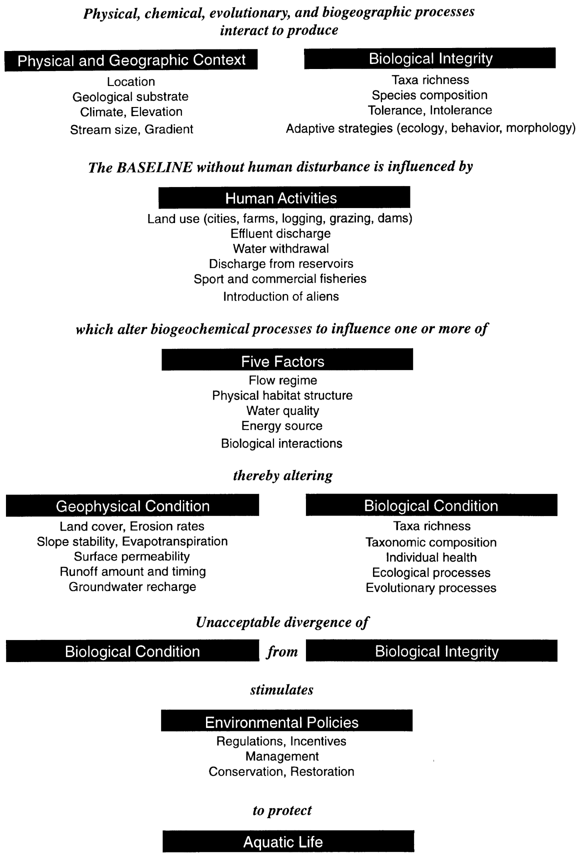

Fig. 2 Relationships among the elements and processes of natural systems, the kinds of changes that occur as a result of human actions, and a framework of environmental policies that might come from an assessment of biological condition, the endpoint of primary concern to society. (From Karr & Chu, 1999).

integrity, 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, 1991; Karr *et al.*, 1986); the concept was adopted quickly by a variety of state (Ohio Environmental Protection Agency, 1988) and federal (Plafkin *et al.*, 1989) agencies and in geographic regions throughout the world (Oberdorff & Hughes, 1992; Minns *et al.*, 1994; Davis & Simon, 1995; Rossano, 1996; Deegan *et al.*, 1997; Koizumi & Matsumiya, 1997; Thorne & Williams, 1997; Weisberg *et al.*, 1997; Harris & Silveira, 1999). These applications have not been equally 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 ten standard RBP metrics used in Oregon (Fore, Karr & Wisseman, 1996), six failed under scrutiny according to the criteria for validating metrics that go into the indices 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 & 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 indices, especially IBI, but their arguments indicate misunderstandings of many of the scientific and policy foundations of multimetric assessments (Karr & Chu, 1999). 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.

Effective multimetric biological indices avoid indicators that are either theoretically or empirically flawed; see Karr & Chu (1999) for a review. They incorporate components of biology 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 behaviour of the elements and processes of biological systems. Common metrics include those that illustrate changes in taxonomic 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 indices thus integrate multiple dimensions of complex systems. In this respect, they are similar to the indices used to measure the health of regional and national economies (e.g. index of leading economic indicators or consumer price index in the USA).

Integrative multimetric biological indices are well suited to judging river health against defined goals or water quality standards (called 'criteria' in the USA). 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 a guide and a goal for ecological restoration.

Developing a multimetric index

Five tasks are critical to the development and use of an effective multimetric biological index (Karr & Chu, 1997, 1999): classify to define homogeneous sets; select appropriate metrics; develop sampling protocols; analyze data to reveal biological patterns; and communicate the results.

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 river classification system that differs from that used by biologists, for example, even though geophysical context is a fundamental determinant of variation in biological systems. In the Pacific Northwest, geomorphologists identify some fifty to sixty channel types based on the interplay of physical and chemical processes that shape stream channels (MacDonald, Smart & Wissmar, 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 fifty biological categories may therefore be needed for effective biological monitoring. Community classification generally lags behind classification by physical environment or habitat type (Angermeier & Schlosser, 1995). Classification at levels appropriate for biological monitoring and assessment, especially 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. 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 zeros in the data matrix, for example. In this circumstance, mathematical and statistical tractability imposes decisions that diminish our ability to detect and understand biological signals. Use of species-level community comparisons, such as percentage similarity indices, 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 & Chu, 1999).

The point of classification is to group places where living systems are similar at higher taxonomic and ecological levels in the absence of human disturbance, and where the biological responses are similar after human disturbance. 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 and environmental policy goals.

Select appropriate metrics

Successful application of a multimetric index depends on a rigorous process to identify and test metrics, i.e.

the measurable biological attributes that provide reliable and relevant signals about the biological effects of human activities. Failure to properly define metrics can give incorrect signals about resource condition and lead to numerous errors in both science and management. Generally, multimetric indices incorporate a richer array of signals than do analyses based on species composition and abundance matrices (Karr & Chu, 1999).

Selection of metrics for a multimetric index involves several important steps. First, sampling must include a series of 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 a third step, the demonstration 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 indices. Studies intending to test the effectiveness of multimetric approaches fail when their metrics are not selected according to this core principle (e.g. Reynoldson *et al*., 1997). The development and advocacy of rapid bioassessment protocols in the United States (Plafkin *et al*., 1989) also neglected to select metrics that changed consistently across a gradient of human influence (see Karr & Chu, 1999).

Graphs are particularly helpful for identifying biological attributes that change consistently and quantitatively across a gradient of human influence because they force us to confront the obvious. A graph whose y-axis represents a biological response and whose x-axis is a measure of human influence is the ecological analogue of a toxicological dose–response curve (e.g. Figure 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. Graphs highlight idiosyncrasies in patterns of data that, when

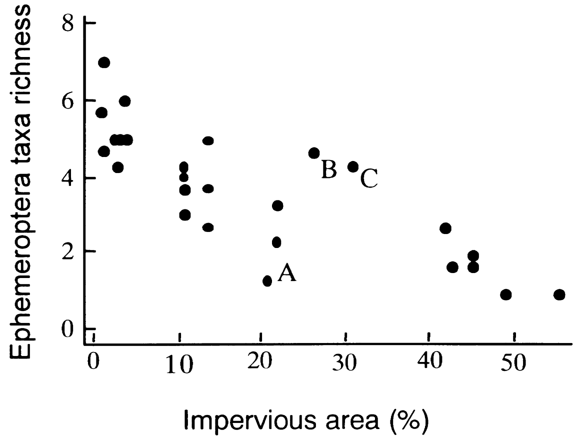

Fig. 3 Average taxa richness of mayflies (Ephemeroptera) plotted against percentage of impervious surface area surrounding Puget Sound lowland streams. Note the general dose–response curve relationship: as human influence (impervious area) increases, taxa richness declines. Sites B and C had relatively 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. (From Karr & Chu, 1999).

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 indices do not follow this rigorous selection process. Many 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 m^2 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 & Chu, 1999), 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 indices. Conversely, careful screening of attributes to define reliable metrics can provide an effective multimetric index (Karr & Chu, 1999). No vigilant medical community would permit the use of tests that had not been demonstrated to accurately diagnose a disease. The same rigour should be applied to choosing metrics for a multimetric index. In short, the selection of biological signal used to detect the effects of human actions should use the insights provided by graphs and supplement those insights with thoughtful use of conventional statistics and knowledge of regional natural history. In study after study, the same major attributes give reliable signals of resource condition in different circumstances (Miller *et al.*, 1988; Karr & Chu, 1999; 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 for sampling biological systems are more correct than others. 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 indices than the other steps in multimetric biological monitoring. A variety of sampling procedures, even samples based on different taxa, provide data of

Table 1 Biological attributes in two groups, those selected for the benthic index of biological integrity (B-IBI) and attributes corresponding to functional feeding groups. The latter were not included in B-IBI because they did not give consistent dose–response curves across gradients of human influence for multiple data sets. Attributes that responded to human-induced disturbance for a data set are indicated by †; those marked with a dash (–) were not tested. Blanks indicate no consistent response. See Karr & Chu (1999) for additional details on all these results.

Biological attributes	Predicted response	Tenn. Valley[a]	SW Ore.[b]	Eastern Ore.[c]	Puget Sound[d]	Japan[e]	NW Wyo.[f]
Metrics used in benthic index of biological integrity							
Total number of taxa	Decrease	†	†		†	†	
Ephemeroptera taxa	Decrease	†	†		†	†	†
Plecoptera taxa	Decrease	†	†	†	†		†
Trichoptera taxa	Decrease	†	†	†	†	†	
Long-lived taxa	Decrease	–	†		†	–	
Intolerant taxa	Decrease	†	†	†	†	†	†
% Tolerant	Increase	†	†		†	†	†
'Clinger' taxa richness	Decrease	–	–	–	†	†	–
Dominance	Increase	†	†			†	†
Attributes based on functional feeding groups							
% Predators[g]	Decrease	†		†	†		†
% Scrapers	Variable	†		†			†
% Gatherers	Variable			†			
% Filterers	Variable	†					
% Omnivores	Increase	†					
% Shredders	Decrease			†			†

[a]Kerans & Karr (1994); [b]Fore *et al.* (1996); [c]Fore, Karr & Tait, unpublished; [d]Kleindl & Karr, unpublished; [e]Rossano (1996); [f]Patterson, Karr & Luchtel, unpublished.
[g]The only functional feeding group metric included in the 10-metric B-IBI.

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 & Chu, 1999) on the basis of 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.

The choice of taxon is less important than the methods and analysis used to detect signals. Sampling strategies will vary among circumstances. For fish sam-pled 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 collected on the basis of general rules such as 'collect in proportion to the abundance of those habitats'. Such judgement calls are an open invitation to create problems in data interpretation .

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 & 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 emerged 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

Table 2 Data sets examined from diverse geographic areas, noting the kind of human influence assessed, aspects of the field methods (replication, habitat sampled, sampling method) and laboratory methods (level of identification, subsampled or not) used to collect and handle data, and the reliability of inferences about doses–responses across gradients of human influence (good = relatively tight, fair = very loose, poor = none) drawn from those data.

Location	Human influence	Number of replicates	Habitat sampled	Sampling method	Taxonomic level	Subsample?	Result
Tennessee[a]	Mixed	4–9	Pools	Hess	Genus	No	Good
Tennessee[a]	Mixed	4–9	Riffles	Surber	Genus	No	Good
South-west OR[b]	Logging and mixed	5 (composite)	Riffle	Kicknet	Species	Only if large	Fair
North-east OR[c]	Riparian damage	5	Riffle	Hess	Genus	No	Good
Eastern WA and OR[d]	Mixed	–	Multiple	Travel Kicknet	Species	Yes	Poor
Puget Sound, WA[e]	Urban mixed	3	Riffle	Surber	Genus	No	Good
Grand Tetons, WY[f]	Recreation	3	Riffle	Surber	Genus	No	Good
Japan[g]	Mixed	3 (composite)	Riffle	Kicknet	Species	No	Good
Ohio[h]	Mixed	1	Riffle	Dendy	Species	No	Good

[a]Kerans *et al*. (1992), Kerans & Karr (1994); [b]Fore *et al*. (1996); [c]Fore, Karr & Tait, unpublished; [d]Fore & Karr, unpublished; [e]Kleindl & Karr, unpublished; [f]Patterson, Karr & Luchtel, unpublished; [g]Rossano (1996); [h]DeShon (1995).

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 & Chu, 1999).

Level of identification (family, genus, species) and subsampling are especially controversial. 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 (Karr & Chu, 1999). Family-level identification is a reasonable compromise when expertise, time, or funding is limited; only 5 of 10 B-IBI metrics are reliable, however, when invertebrates are identified to family level. This compromise is reasonable in some situations. A recent study, for example, documents the effectiveness of volunteer and professional programs to measure stream condition (Fore, Paulsen & O'Laughlin, unpublished). When citizen volunteers collected and sorted benthic invertebrate samples, they were able to discriminate four distinct levels of biological condition at 80% statistical power. Professional identification of most groups to generic level improved discrimination to only about 4.4 levels of biological condition (80% statistical power); full professional laboratory protocols improved discrimination to 5.8 levels of biological condition.

Subsampling, a standard approach used to save money in many bioassessment protocols (Barbour & Gerritsen, 1996), is less reliable, according to a recent study (Doberstein, Karr & Conquest, unpublished). Doberstein and colleagues analyzed 500 random 100-individual subsamples of invertebrates from a minimally disturbed Puget Sound lowland stream (Washington State, USA). They could statistically discern 8.2 classes of biological condition on the basis of a 10-metric B-IBI when complete samples were counted. Only 2.8 classes of stream condition could be discerned with counts of 100 random individuals from the same samples. They concluded that the potential for ill-informed water resource decisions was unacceptably high when based only on 100-individual subsamples.

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. Typically, riffle habitats are 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 the sampling of 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, more clearly than strictly statistical tools, relationships between biological attributes and human influence (Karr & Chu, 1999). 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 & Parker, 1986; Stewart-Oaten, Bence & 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 indices (Fore, Karr &

Conquest, 1994; Karr & Chu, 1999). They are statistically versatile and amenable to application of familiar statistical tests [e.g. *t*-test or analysis of variance (ANOVA)]. From statistical power analysis, a properly formulated 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 managerial (e.g., 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 for 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 indices 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.

Biologists themselves are 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 if they continually equate habitat and inhabitants. Too often, restoration efforts focus on physical and chemical processes, or functions, rather than biological context. Even when restoration focuses on key 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.

Acknowledgements

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 numerous grants from the US Environmental Protection Agency. Ellen Chu provided invaluable advice that substantially improved the paper.

References

Angermeier P.L. & Karr J.R. (1994) Biological integrity versus biological diversity as policy directives. *Bioscience*, **44**, 690–697.

Angermeier P.L. & Schlosser I.J. (1995) Conserving aquatic biodiversity. *American Fisheries Society Symposium*, **17**, 402–414.

Anonymous (1998) Museum pieces. *Environment*, **40**, 23.

Barbour M.T. & Gerritsen J. (1996) Subsampling of benthic samples: a defense of the fixed-count method. *Journal of the North American Benthological Society*, **15**, 386–391.

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, DC.

Davies S.P., Tsomides L., Courtemanch D.L. & Drummond F. (1995) *Maine Biological Monitoring and Biocri-*

teria Development Program. Maine Department of Environmental Protection, Bureau of Land and Water Quality, Division of Environmental Assessment, Augusta.

Davis W.S. & 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., Finn J.T., Ayvazian S.G., Ryder-Kieffer C.A. & Buonaccorsi J. (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 & T.P. Simon), pp. 217–244. Lewis, Boca Raton, FL.

Donahue J.M. & Johnston B.R. (Eds) (1998) *Water, Culture, and Power: Local Struggles in a Global Context.* Island Press, Washington, DC.

Fore L.S., Karr J.R. & Conquest J.L. (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. & Wisseman R.W. (1996) Assessing invertebrate responses to human activities: evaluating alternative approaches. *Journal of the North American Benthological Society*, **15**, 212–231.

Harris J.H. & Silveira R. (1999) Large-scale assessments of river health using an index of biotic integrity with low diversity fish communities. *Freshwater Biology*, **41**, 235–252.

Haskell B.D., Norton B.G. & Costanza R. (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, DC.

Karr J.R. (1981) Assessment of biotic integrity using fish communities. *Fisheries*, **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 & 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. 97–109. National Academy Press, Washington, DC.

Karr J.R. (1998) Rivers as sentinels: using the biology of rivers to guide landscape management. *River Ecology and Management: Lessons from the Pacific Coastal Ecoregion* (Eds R.J. Naiman and R.E. Bilby), pp. 502–528. Springer-Verlag, New York.

Karr J.R. & Chu E.W. (1997) Biological monitoring: essential foundation for ecological risk assessment. *Human and Ecological Risk Assessment*, **3**, 993–1004.

Karr J.R. & Chu E.W. (1999) *Restoring Life in Running Waters: Better Biological Monitoring.* Island Press, Washington, DC.

Karr J.R., Fausch K.D., Angermeier P.L., Yant P.R. & Schlosser I.J. (1986) Assessing biological integrity in running waters: a method and its rationale. *Illinois Natural History Survey Special Publication* 5. Champaign, IL.

Karr J.R. & Thomas T. (1996) Economics, ecology, and environmental quality. *Ecological Applications*, **6**, 31–32.

Kerans B.L. & Karr J.R. (1994) A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. *Ecological Applications*, **4**, 768–785.

Kerans B.L., Karr J.R. & Ahlstedt S.A. (1992) Aquatic invertebrate assemblages: spatial and temporal differences among sampling protocols. *Journal of the North American Benthological Society*, **11**, 377–390.

Koizumi N. & Matsumiya Y. (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., Smart A. & Wissmar R.C. (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, WA.

Mageau M.T., Costanza R. & Ulanowicz R.E. (1995) The development and initial testing of a quantitative assessment of ecosystem health. *Ecosystem Health*, **1**, 201–213.

Meyer J.L. (1997) Stream health: incorporating the human dimension to advance stream ecology. *Journal of the North American Benthological Society*, **16**, 439–447.

Miller D.L., Leonard P.M., Hughes R.M., Karr J.R., Moyle P.B., Schrader L.H., Thompson B.A., Daniels R.A., Fausch K.D., Fitzhugh G.A., Gammon J.R., Halliwell D.B., Angermeier P.L. & Orth D.J. (1988) Regional applications of an index of biotic integrity for water resource management. *Fisheries*, **13**, 12–20.

Minns C.K., Cairns V.W., Randall R.G. & 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, 1914–96. US Government Printing Office, Washington, DC.

Oberdorff T. & 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 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.

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. & Norris R.H. (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. & 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. C. Cothern), pp. 37–51. Lewis Publishers, Boca Raton, FL.

Prugh T., Costanza R., Cumberland J.H., Daly H., Goodland R. & Norgaard R. (1995) *Natural Capital and Human Economic Survival.* International Society for Ecological Economics Press, Solomons, MD.

Rapport D.J. (1989) What constitutes ecosystem health? *Perspectives in Biology and Medicine*, **33**, 120–132.

Rapport D.J., Gaudet C., Karr J.R., Baron J.S., Bohlen C., Jackson W., Jones B., Naiman R.J., Norton B. & Pollock M.M. (1998) Evaluating landscape health: integrating societal goals and biophysical processes. *Journal of Environmental Management*, **53**, 1–15.

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., Bailey R.C., Day K.E. & Norris R.H. (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., Norris R.H., Resh V.H., Day K.E. & Rosenberg D.M. (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, Tokyo, Japan, Museum of Streams and Lakes (in Japanese and English).* Sankaido Publishers, Tokyo, Japan.

Scrimgeour G.J. & Wicklum D. (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., Bence J.R. & Osenberg C.W. (1992) Assessing effects of unreplicated perturbations: no simple solutions. *Ecology*, **73**, 1396–1404.

Stewart-Oaten A., Murdoch W.W. & Parker K.R. (1986) Environmental impact assessment: pseudoreplication in time? *Ecology*, **67**, 929–940.

Suter G.W. (1993) A critique of ecosystem health concepts and indexes. *Environmental Toxicology and Chemistry*, **12**, 1533–1539.

Thorne R.St.J. & Williams W.P. (1997) The response of benthic invertebrates to pollution in developing countries: a multimetric system of bioassessment. *Freshwater Biology*, **37**, 671–686.

Toth L.A. (1993) The ecological basis of the Kissimmee river restoration plan. *Florida Scientist*, **56**, 25–51.

Weisberg S.B., Ranasinghe J.A., Schaffner L.C., Diaz R.J., Dauer D.M. & Frithsen J.B. (1997) An estuarine benthic index of biotic integrity (B-IBI) for Chesapeake Bay. *Estuaries*, **20**, 149–158.

Wicklum D. & 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 of the Ecological Society of America*, **71**, 106–111.

(Manuscript accepted September 1998)