Section III

Multimetric Indexes Convey Biological Information

FIVE ACTIVITIES are central to making multimetric biological indexes effective (Karr and Chu 1997):

1. Classifying environments to define homogeneous sets within or across ecoregions (e.g., streams, lakes, or wetlands; large or small streams; warm-water or cold-water lakes; high- or low-gradient streams; depressional or flow-through wetlands).
2. Selecting measurable 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 those data.
5. Communicating the results to citizens and policymakers so that all concerned communities can contribute to environmental policy.

Understanding biological responses requires measuring across degrees of human influence

Our ability to protect biological resources depends on our ability to identify and predict the effects of human actions on biological systems, especially our ability to distinguish between natural and human-induced variation in biological condition. Thus, even though measures taken at places with little or no human influence (e.g., only from "reference" sites) may tell us something about natural variability from place to place and through time at undisturbed sites, they cannot tell us anything about which biological attributes merit watching for signs of human-caused degradation. To find these signs, sampling and analysis should focus on multiple sites within similar environments, across the range from minimal to severe human disturbance.

One could choose sampling sites that represent different intensities of only one human activity, such as logging, grazing, or chemical pollution. It would then be possible to evaluate biological responses to a changing "dose" of a single human influence. Though rare, such a study opportunity could help identify the biological response signature characteristic of that activity (Karr et al. 1986; Yoder and Rankin 1995b). Knowledge of such biological response signatures would give researchers a diagnostic tool for watersheds influenced by unknown or multiple human activities. In reality, however, it is virtually impossible to find regions influenced by only a single human activity.

In most circumstances, diverse human activities interact (e.g., during urbanization) to affect conditions in watersheds, water bodies, or stream reaches. In such cases, sites can be grouped and placed on a gradient according to activities and their effects: industrial effluent is more toxic than domestic effluent, for example, and both pose more serious threats than low dams, weirs, or levees (Figure 5). Removal of natural riparian corridors damages streams, but conversion to a partially herbaceous riparian area is less damaging than conversion to riprap. Streams grouped this way show striking and systematic differences in biological condition across the gradient of human disturbance (Figure 6).

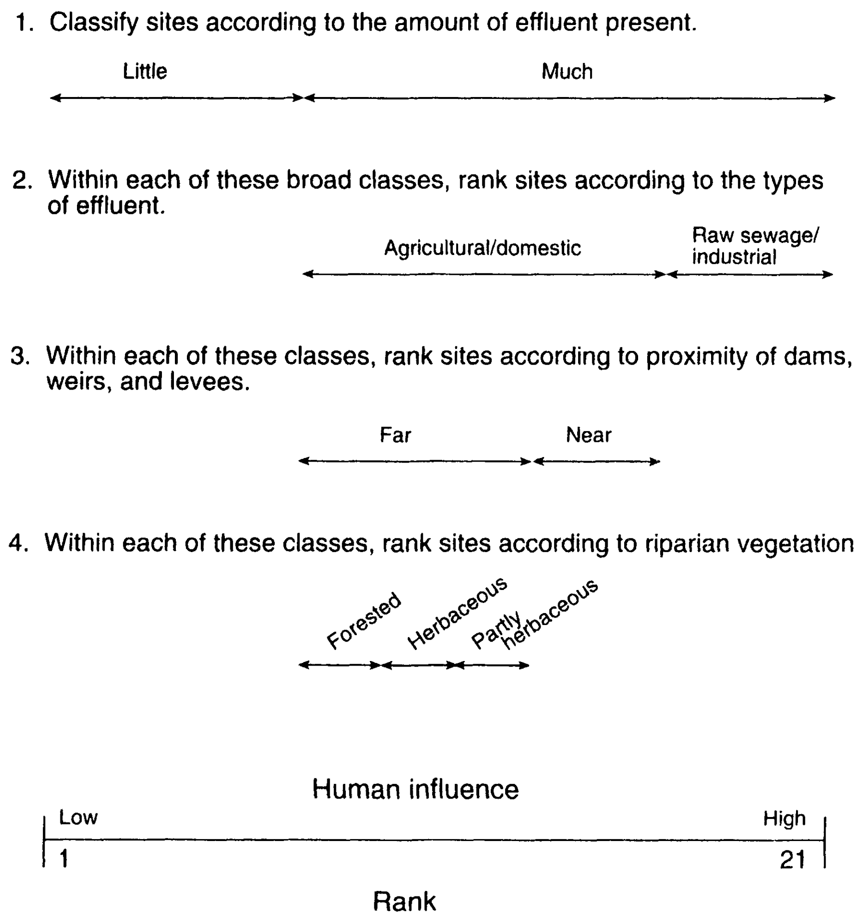

Figure 5. A priori classification system for ranking Japanese streams according to intensity of human influence (Rossano 1995). Sites were assigned to one of 21 possible categories based on amount and type of effluent, proximity of dams and other structural alterations, and type of riparian vegetation. Even without quantitative measures from each site, this approach allowed sites to be ranked across a range of human influence.

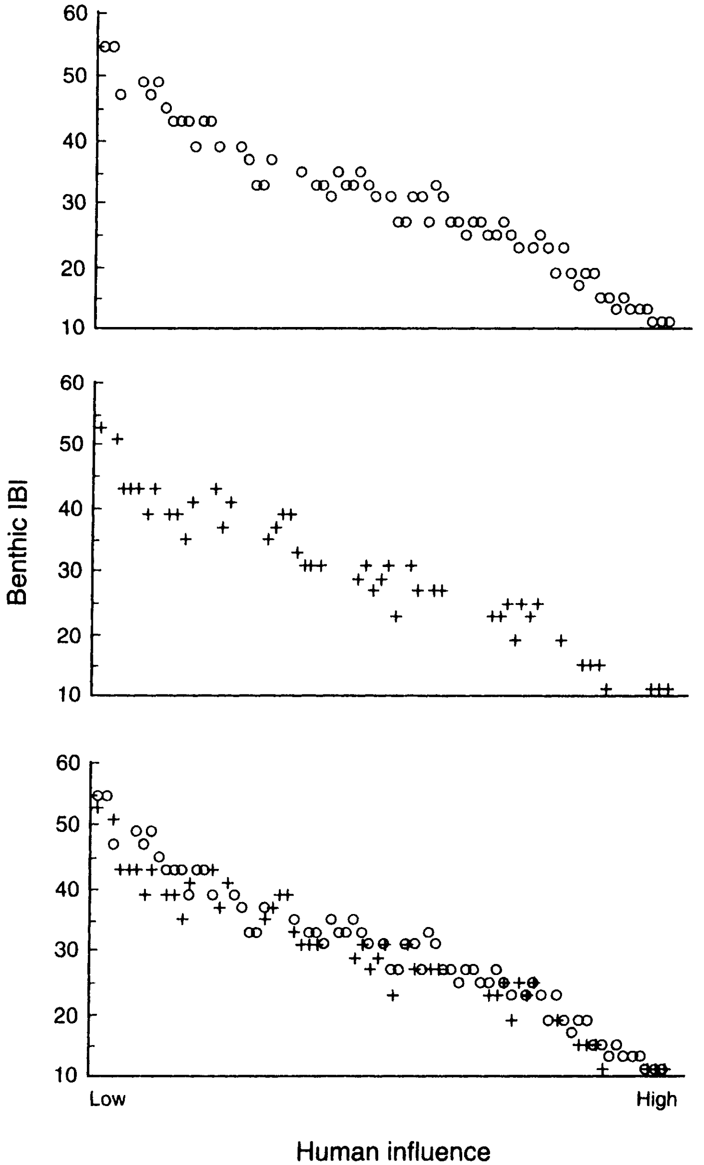

Figure 6. Benthic indexes of biological integrity (B-IBIs) for 115 Japanese streams (from Rossano 1995). The top panel shows B-IBIs calculated from half of the 115-stream data set (circles), which was used to initially select and test metrics for use in the B-IBI. The middle panel shows B-IBI values calculated from the second half of the data set (pluses); the metrics and scoring criteria used for these data were the metrics and criteria developed from the first half. In the bottom panel, all 115 B-IBIs are plotted together; the indexes from both sets correspond closely, ranking the streams comparably according to intensity of land use from low to high. The range of human influence against which the B-IBIs are plotted comes from the classification scheme shown in Figure 5.

Sometimes a single variable can capture and integrate multiple sources of influence. Relatively simple descriptors—human population in the watershed, percentage of impervious area, percentage of land area devoted to agriculture or urban uses, or percentage of developed area—are adequate for regional watershed analyses (Meeuwig and Peters 1996). The percentage of impervious area, for example, summarizes the multiple effects of paving, building, and other consequences of urbanization, as in a recent study of Puget Sound lowland streams (Figure 7; see also Maxted 1997). This measure provides a simple surrogate of human influence that works well at percentages of impervious area from near 0% to 60%. Unfortunately, it is less useful in understanding the often large variation in biological condition at some percentages of imperviousness (e.g., 3% to 8%; see Figure 7). Finding the differences in human activity that can explain these biological differences requires information from the watersheds that is more detailed.

Alternatively, sites may be grouped into qualitative disturbance categories. In a study of recreational influence on stream biology in the northern Rocky Mountains (Figure 8), Patterson (1996) classed sites into four categories associated with different levels of human activity: (1) little or no human influence in the watershed; (2) light recreational use (hiking, backpacking); (3) heavy recreational use (major trailheads, camping areas); and (4) urbanization, grazing, agriculture, or wastewater discharge. Patterson demonstrated that light recreational activity did not substantially reduce

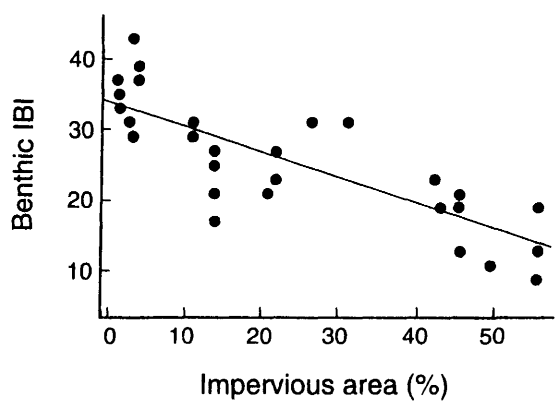

Figure 7. Benthic index of biological integrity (B-IBI) plotted against the percentage of impervious area for urban, suburban, and rural stream sites in the Puget Sound lowlands, Washington (from Kleindl 1995). Though B-IBI clearly decreases with increasing impervious area, this plot offers no insight into B-IBI differences among sites with similar percentages of impervious area, especially at low percentages (3% to 17%).

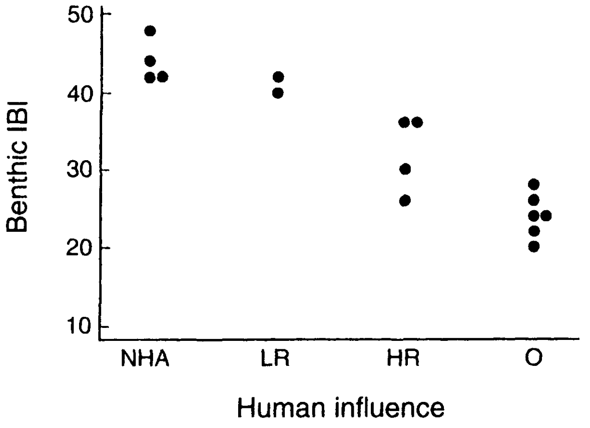

Figure 8. Benthic indexes of biological integrity (B-IBIs) for stream sites in or near Grand Teton National Park, Wyoming (from Patterson 1996). Before B-IBIs were determined, these sites had been placed into four categories of human influence: little or no human activity (NHA), light recreational use (LR), heavy recreational use (HR), and other (O). B-IBIs revealed no significant difference between sites with little or no human activity and those having low recreational use. But B-IBIs were significantly lower for sites used heavily for recreation and lower still for sites subjected to other uses—specifically, urbanization, grazing, agriculture, and wastewater discharge.

B-IBIs in comparison with undisturbed watersheds. Heavy recreational use, in contrast, did significantly alter the benthic invertebrates, although not as much as more-intensive uses including urbanization and agriculture.

A similar approach was taken in a study of biological response to chemical pollution on three continents: South America, Africa, and southeastern Asia (Thorne and Williams 1997). The authors classified sites according to a pollution gradient based on the integration of six measures of chemical pollution. Biological condition, as indicated by metrics such as total taxa richness (families) and mayfly, stonefly, and caddisfly richness, clearly went down as pollution went up. The biological responses in the three tropical regions were similar; the patterns parallel those seen in temperate regions even though the faunas are all very different.

Data collected over a number of years at the same site(s) can also reveal biological responses as human activities change during that period. Regardless of how one represents a range of human influence among study sites, sampling from sites with different intensities and types of human activity is essential to detect and understand biological responses to human influence.

The goal is to compare like environments with like environments—to isolate and understand patterns caused by human activities at sites within those like environments.

Too many existing studies confound patterns of human influence with natural variation over time at undisturbed sites or across different environment types. In other situations, researchers combine measures of human activity, the physical and chemical manifestations of those activities, and their biological consequences in a heterogeneous analysis with ambiguous results. Those analyses may even include measures of physical environment such as stream gradient. When this range of factors (different human influences on different environment types) is lumped in a single analysis, it becomes almost impossible to understand the causes or consequences of human versus natural events.

Consider the following analogy. Three experiments are designed: one to understand variation in natural biological systems as a function of stream size; another to distinguish the effects of pesticide runoff on streams of first, fourth, and sixth order; and a third to define the effects of pesticides on plants and insects. Analyzing samples from the first series of stream sites would tell you about biological responses to changing stream size. Samples from the second series would illustrate changing human influence as a function of stream size. Samples from the third would distinguish responses of different taxa. It would be silly to mix the data from the three studies in a single statistical analysis, without regard to which study the individual samples came from. Yet by using analytical procedures that mix the effects of natural and human-induced variation (in a single correlation matrix, for example), researchers are essentially doing just that: they are ignoring the context of the different components of their data, making it difficult to distinguish the biological signs relevant to resource management or protection. They then confound the sources of the variance they see, even if their initial sampling setup would have permitted discrimination among those sources. Univariate and multivariate analyses all too often suffer from this flaw.

Sampling only from "reference" sites creates a similar problem because it does not provide a way to document which biological attributes vary with human influence (see Premise 31). Careful thought about which variables best summarize human influence and the relationships among those variables should be the foundation of monitoring protocols. Creating opportunities to discover biological patterns in relation to human activity must be foremost.

Only a few biological attributes provide reliable signals about biological condition

The success of biological monitoring programs and their use to define and enforce biological criteria is tied to identifying biological attributes that provide reliable signals about resource condition (Table 3). Choosing from the profusion of biological attributes (Figure 9) that could be measured is a winnowing process, in which each attribute is essentially a hypothesis to be tested for its merit as a metric. One accepts or rejects the hypothesis by asking, Does this attribute vary systematically through a range of human influence? When metrics are selected and organized systematically, an effective multimetric index can emerge.

Knowledge of natural history and familiarity with ecological principles and theory guide the definition of attributes and the prediction of their behavior 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 (e.g., RBP I, II, III; Plafkin et al. 1989).

The study of populations has dominated much ecological research for decades (see Section II), so researchers still assume that population size (expressed as abundance or density) provides a reliable signal about water resource condition. But because species 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 (see Premise 14 and Premise 25). Large numbers of samples (>25) were required, for example, to detect small (<20%) differences in number of fish per 100 square meters of stream surface area in small South Carolina streams (Paller 1995b). Other attributes—such as taxa richness (number of unique taxa in a sample, including

Table 3. Key terms used in defining biological condition.

Term	Definition
Endpoint	A measured characteristic that indicates the condition of a biological, chemical, or physical system
Attribute	Measurable part or process of a biological system
Metric	Attribute empirically shown to change in value along a gradient of human influence
Multimetric index	A number that integrates several biological metrics to indicate a site's condition
Biological monitoring	Sampling the biota of a place (e.g., a stream, a woodlot, or a wetland)
Biological assessment	Using samples of living organisms to evaluate the condition or health of places
Biological criteria	Under the Clean Water Act, numerical values or verbal (narrative) standards that define a desired biological condition for a water body; legally enforceable

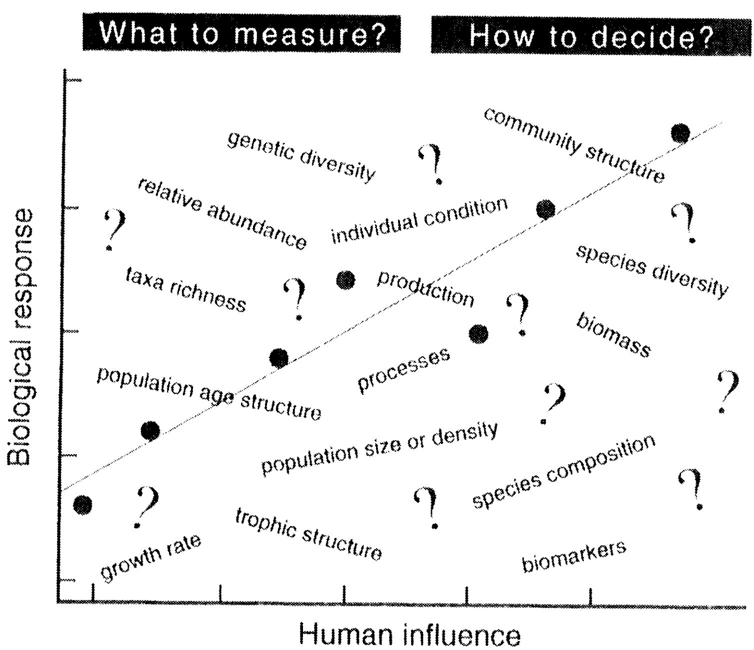

Figure 9. Almost any biological attribute can be measured, but only certain attributes provide reliable signals of biological condition and therefore merit integration into a multimetric index.

rare ones) and percentages of individuals belonging to tolerant taxa—have, in contrast, been found to vary consistently and systematically with human influence. Such attributes, when graphed, give rise to analogues of the toxicological dose-response curve—which we call ecological dose-response curves—where the *y*-axis represents the measured attribute and the *x*-axis, measures of human influence (Figure 10).

Ecological dose-response curves differ in one critical respect from toxicological dose-response curves. Toxicological dose-response curves usually measure biological response in relation to dose of a single compound. Ecological dose-response curves measure a biological response to the cumulative ecological exposure, or "dose," of all events and human activities within a watershed, expressed in terms such as percentage of area logged, riparian condition, or percentage of impervious area. The number of native fish species in a midwestern stream sampled today, for example, reflects the cumulative effects of natural events and human influence up to the present. The very existence of those species is the product of what has occurred before.

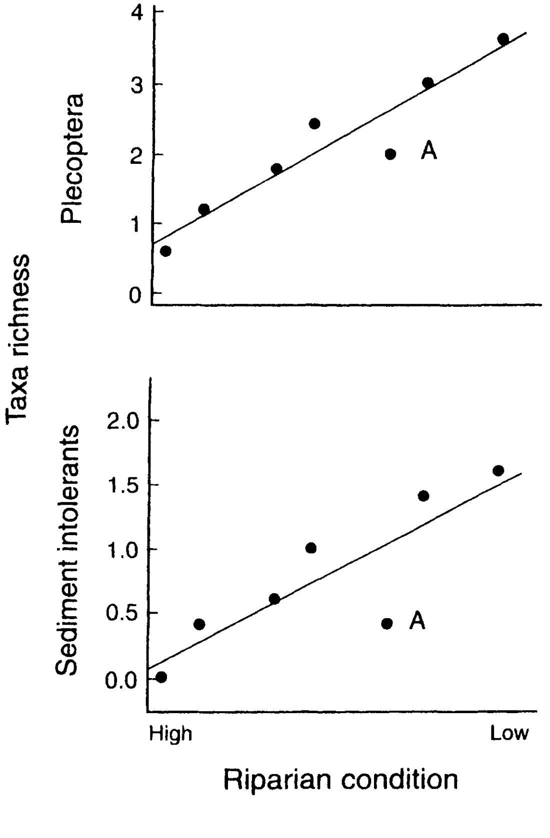

Figure 10. Average taxa richnesses of Plecoptera and sediment-intolerant taxa plotted against riparian condition for seven stream sites in the John Day Basin, Oregon, in 1988. Site A had fewer taxa than expected because although cattle were excluded, intense grazing upstream had affected the site's biota.

Graphs reveal biological responses to human influence

"Often the most effective way to describe, explore, and summarize a set of numbers (even a very large set) is to look at pictures of those numbers. . . . [O]f all methods for analyzing and communicating statistical information, well-designed data graphics are usually the simplest and at the same time the most powerful" (Tufte 1983: 9; see also Tufte 1990, 1997). Tufte's message is nowhere more important than in the display, interpretation, and communication of biological monitoring data.

Graphs reveal the biological responses important for evaluating metrics more clearly than do strictly statistical tools. They exploit "the value of graphs in forcing the unexpected" (Mosteller and Tukey 1977) on whoever looks at them, including researchers, who must then confront and explain the pattern in those graphs. For samples where the relationship between human influence and biological response is strong, statistics and graphs agree (Figure 11). In other cases, meaningful biological patterns can be lost by excessive dependence on the outcome of menu-driven statistical tests. Statistical correlation can miss an important relationship if the x-variable (e.g., percentage of area logged) is measured with low precision or if additional factors beyond those plotted on the x-axis influence metric values but are not included in the statistical analysis.

In Figure 12, for example, we plot two different aspects of biological condition against one measure of human influence, such as the percentage of upstream watershed that has been logged. Sites are assigned a plus or minus on the basis of that measure and other aspects of human influence that are visible and documented but not plotted on the same graph. In forested watersheds, these other aspects might include whether roads are near or far from the stream channel, time since logging, or traits unique to particular watersheds. In some cases, such interacting factors may degrade biological condition (roads near the stream channel would worsen logging's effects), or they may allow good conditions to persist (roads on distant ridges

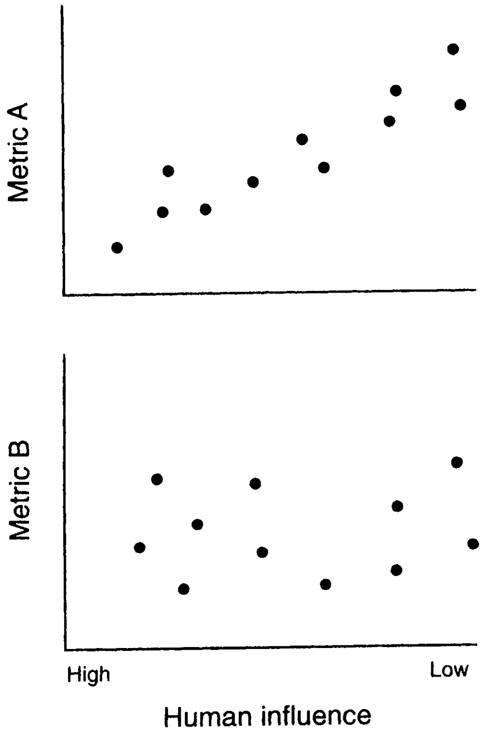

Figure 11. Example of two hypothetical metrics plotted against a gradient of human influence. Here statistical correlation and graphical analysis agree: metric A is a good indicator, and metric B is not. (Compare Figure 12.)

have less effect on streams). The distribution of pluses and boxes in Figure 12 illustrates the fallacy of assuming that a biological metric says nothing about condition because it does not correlate strongly with a single surrogate of that condition, as researchers perennially assume when a biological measure does not correlate with some measure of chemical pollution. Rather, we should conclude that the surrogate is not capturing significant components of human influence and look more closely for the biological explanations behind the data.

Not all aspects of human influence can be easily captured in a single graph or statistical test. When a number of variables influence condition, a single plot against one dimension of human influence will not tell the whole story (Figure 13); neither will a single statistical test. Graphs force us to search for insights that rote application of statistical tests cannot discover.

Weak statistical correlation can also miss important biological patterns

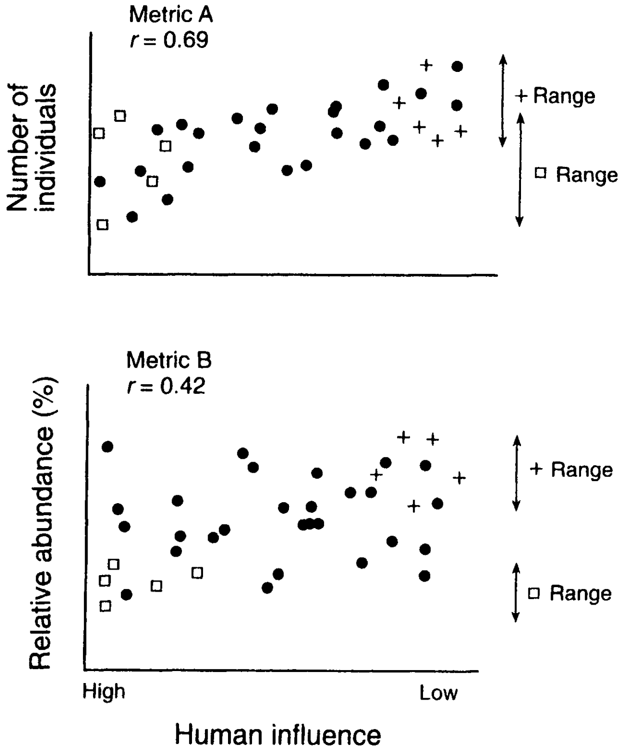

Figure 12. Hypothetical relationships between human influence and candidate biological metrics (from Fore et al. 1996). Metric A is more strongly correlated with resource condition (or r^2 is higher if using regression) than Metric B, initially suggesting that it is a better metric. But comparing the metrics' ability to distinguish between minimally disturbed sites (denoted by plus signs) and severely degraded sites (open boxes; ranges noted by arrows) shows that Metric B is actually a more effective measure of biological condition despite its smaller statistical correlation. (Compare Figure 11.)

when the distribution of the data (e.g., Figure 14) does not lend itself to tests based on standard correlation techniques that detect only linear relationships. Yet nonlinear patterns are common in field data (Figure 15). Consider the plots in Figure 16, for example. The points fall into a wedge-shaped distribution, whose scatter shows little or no statistical significance but can be interpreted biologically. The upper bound of each plot is the hypotenuse of a right triangle (the maximum species richness line) that defines the number of species expected in minimally disturbed streams as a function of stream size (Fausch et al. 1984). The plots illustrate what Thomson et al.

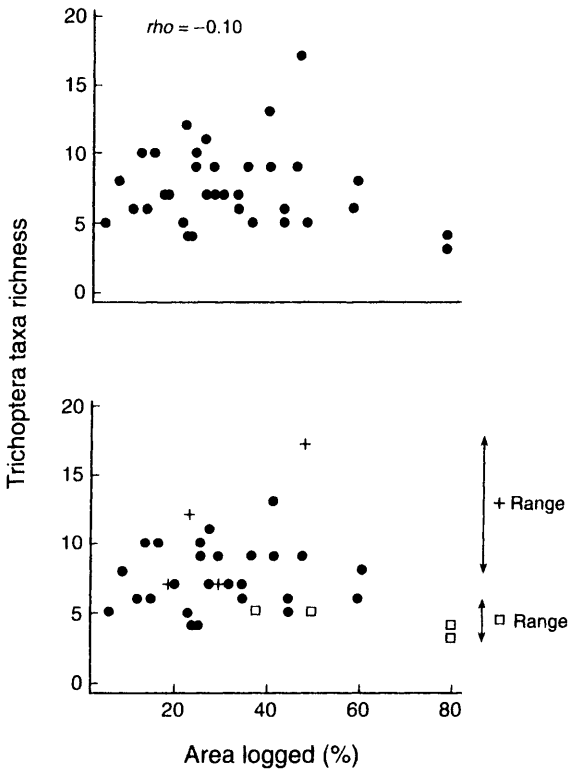

Figure 13. Taxa richness of Trichoptera plotted against the percentage of watershed area that was logged for 32 stream sites in southwestern Oregon. Metric correlation (Spearman's *rho*) was not significant because, alone, the percentage of area logged was an inaccurate measure of human influence; other factors, such as type of logging, presence of roads, and other human influences, were not included. When these other human influences were considered, to identify minimally disturbed sites (denoted by plus signs) and severely degraded sites (open boxes), the response of Trichoptera taxa richness visibly distinguished between different degrees of human disturbance.

(1996) term a "factor ceiling distribution" (see also Blackburn et al. 1992 and Scharf et al. 1998 on ecological inferences from the edges of scatter diagrams). In this case, the ceiling—maximum species richness—is defined by the evolution of the regional biota. Generally at sites where the number of fish species falls below the ceiling, some human activity in the adjacent or upstream watershed has reduced the number of species present; alternatively, sampling might have been inadequate, "dragging" species richness below the line.

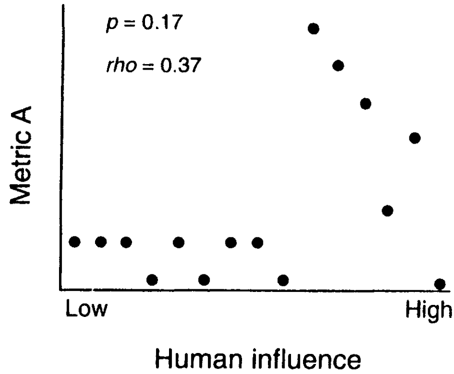

Figure 14. Hypothetical relationship between human influence and a Metric A. Statistical correlation (Spearman's *rho*) is not significant, yet the graphic pattern strongly suggests a biological response. At low levels of human influence, Metric A is not a reliable indicator of biological condition, but where human disturbance is high, the metric does respond.

Graphs highlight idiosyncrasies in data distributions that, when examined closely, may provide insight into the causes of a particular biological pattern. At one extreme, outlying points on a graph may offer key insights about the complex influence of human activities in watersheds. The researcher can then explore what unique situations at those sites cause them to appear as outliers.

Even the spread of data can offer insights, as illustrated by the large range in B-IBIs at sites with 20% to 30% impervious area shown in Figure 17. Sites with high mayfly taxa richness (B and C) lie in reaches of two streams with relatively intact riparian corridors and wetlands. The site with

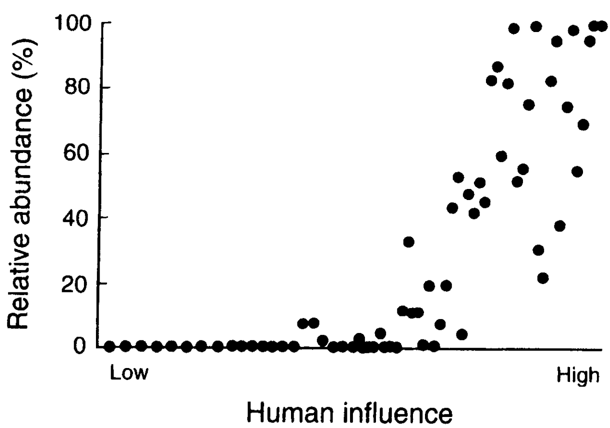

Figure 15. Relative abundance (percentage of total) of individuals belonging to tolerant taxa in samples of benthic invertebrates from 65 Japanese streams ranked according to intensity of human influence (see Figures 5 and 6). (Data provided by E. M. Rossano.)

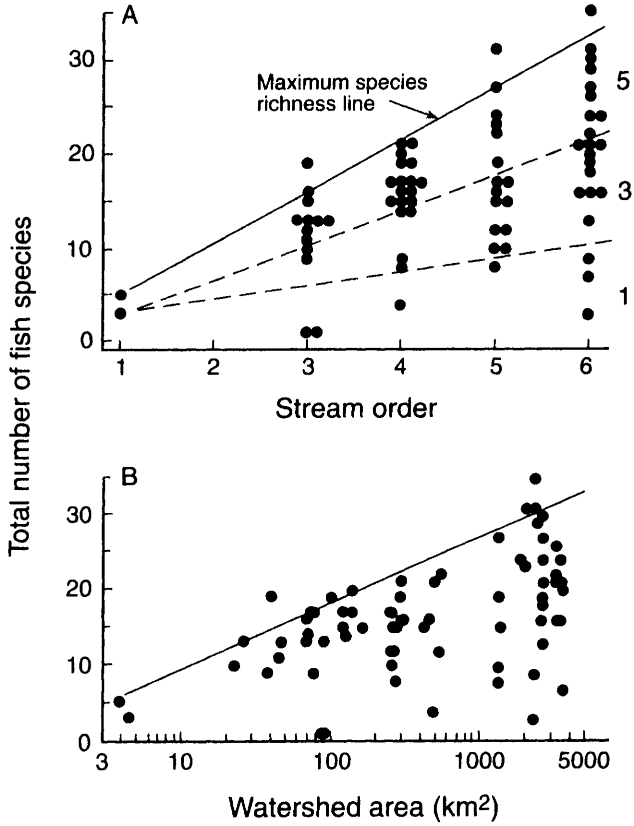

Figure 16. Number of fish species in relation to stream size (*top*) and watershed area (*bottom*); each point represents a site. The maximum species richness line through the highest points on each graph defines the number of species expected in minimally disturbed streams or watersheds. Points below that line represent sites where human activity has reduced the number of species present (from Fausch et al. 1984).

low mayfly taxa richness (A) is located in a stream that receives fine material from an old coal mine. Sites A, B, and C had unique characteristics that were best understood by examining their specific contexts, not by applying a regression or correlation analysis. Finding these patterns then led to subsequent studies in the same and in other places to determine if those patterns were more general.

Graphs also illustrate variation in behavior among taxa in response to a specific disturbance (Figure 18). For example, numbers of taxa for three

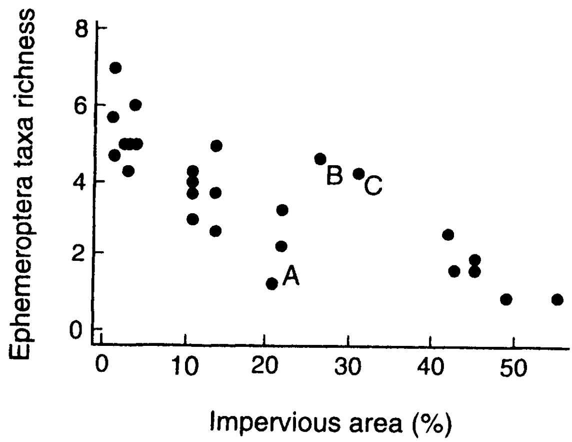

Figure 17. Average taxa richness of Ephemeroptera plotted against percentage of impervious area surrounding Puget Sound lowland streams (from Kleindl 1995). Site A, Coal Creek, had fewer Ephemeroptera than expected. This site has an active mine in its headwaters, and Ephemeroptera are known to be sensitive to mine waste. Sites B and C had relatively intact riparian areas (wetlands).

orders of insects (stoneflies, mayflies, and caddisflies) declined downstream of the outflow from a streamside sludge pond in the Tennessee Valley, but the magnitude of change varied among the taxa (see also Premise 14). The same graph also reveals the direction and magnitude of change along a longitudinal transect down the stream.

Graphs may sometimes allow researchers to avoid naive application of elaborate multivariate techniques (Beals 1973). Principal components analysis, the most often used ordination technique (James and McCullough 1990), defines statistically orthogonal factors, which may or may not be independent biologically; interpreting the results can therefore be complicated (Goodall 1954). Graphs can be a superior approach to methods that focus on maximum variance extracted because they reveal ecological rather than mathematical associations, a more appropriate criterion for organizing and understanding complex information (Beals 1973).

Complex ecological situations require unusual analytical means. Graphs can often be ecologists' most useful tools, permitting the exploration of ecological data "before, after, and beyond the application of 'standard analyses'" (Augspurger 1996). Rather than choose an inappropriately linear statistical model before plotting their data, ecologists should exploit the power of graphs for "reasoning about quantitative information" (Tufte 1983) and

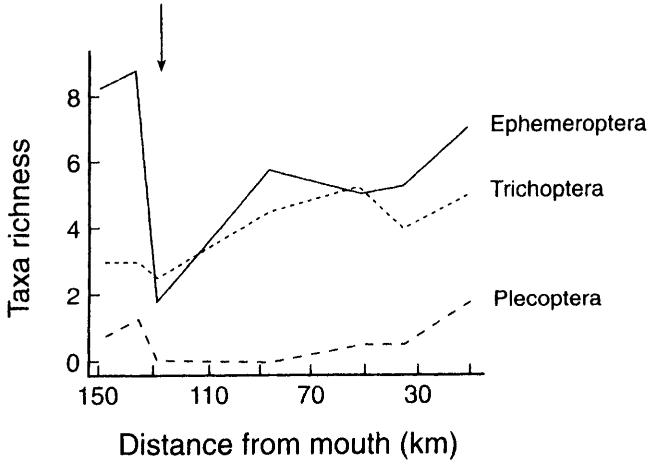

Figure 18. Taxa richness of mayflies, stoneflies, and caddisflies for sites along the North Fork Holston River in the Tennessee Valley in 1976 (from Kerans and Karr 1994). Arrow indicates the position of the streamside sludge pond. Taxa richnesses for all three orders decline at the sludge pond and slowly recover for sites downstream.

then choose and apply appropriate statistics. It is myopic to be a slave of standard statistical rules and procedures—just as it would be myopic to avoid statistics altogether.

Similar biological attributes are reliable indicators in diverse circumstances

A striking conclusion from 15 years' research in selecting metrics is that the same major biological attributes serve as reliable indicators in diverse circumstances. This result has its advantages and disadvantages. On the advantage side, every small project (e.g., at the county or community level) need not test and define its own locally applicable metrics. Scientists and resource managers can implement local biological monitoring and assessment programs on the basis of results from other studies. When local studies cite earlier work, readers can know that the methods have been tested elsewhere; the accumulating body of tests refines, or refutes, the generality of patterns that others have defined.

On the disadvantage side, some applications of multimetric indexes uncritically borrow theoretical or empirical metrics from other studies. This borrowing becomes problematic when the theory is wrong or does not apply in the study circumstance, or when metrics are applied to systems or regions other than those for which they were tested. For example, human impacts may increase taxa richness in cold-water streams (Hughes and Gammon 1987; Lyons et al. 1996), as cool- and warm-water species enter areas where water temperatures have been raised by activities that alter riparian vegetation. In contrast, in eastern warm-water streams, human influence commonly decreases species richness except for aliens (Karr et al. 1986). Thus, you cannot make identical assumptions about metrics of fish taxa richness in the two contexts.

Similarly, a benthic invertebrate metric for soft-bodied organisms (e.g., oligochaetes, tipulid flies, and other grublike forms) often indicates degraded conditions in North America, but in Japan, the better metric is legless organisms, a grouping that includes the soft-bodied organisms but also shelled snails and mussels. In North America, mussels and snails are more often indicators of high-quality environments, but in Japan, most of these taxa are alien or otherwise indicative of degraded conditions.

The bottom line is that metrics should be based on sound ecology and adapted only with great care beyond the regions and habitats for which they were developed. Exploring biological patterns to discover the best biological signals (that is, metrics) should mix graphs, conventional statistics, and thoughtful consideration of regional natural history.

Tracking complex systems requires a measure that integrates multiple factors

Policymakers, citizens, and scientists faced with making decisions about complex systems—economies, personal health, societal well-being, an ecological system—need multiple levels of information. Consider some of the indexes used to track the health of the national economy: the index of leading economic indicators, the producer price index, the consumer price index, the cost-of-living index, and the Dow Jones industrial average. All these indexes integrate multiple economic factors.

The index of leading economic indicators (Mitchell and Burns 1938) tracks the U.S. economy in terms of 12 measures: length of work week; unemployment claims; new manufacturing orders; vendor performance; net business formation; equipment orders; building permits; change in inventories, sensitive materials, and borrowing; stock prices; and money supply. These measures are combined to form the overall index, which takes as its reference point a standardized year (e.g., 1950); the value of the current year's index is expressed in terms of its value in the reference year. Composite economic indexes like these have survived six decades of discussion and criticism and remain widely used by economists, policymakers, and the media to interpret economic trends (Auerbach 1982).

Similarly, physicians and veterinarians rely on multiple measures and multiple tests to assess the health of individual patients. On a single visit to the doctor, you might be "sampled" for urine chemistry, blood-cell counts, blood chemistry, body temperature, throat culture, weight, or chest X-rays. Clearly, these measurements are not independent of one another, for they come from a single individual whose health is affected by many interacting factors. Further, you would not expect your doctor to rely on only one specialized blood test to diagnose your overall health; rather, you assume that multiple measures will give a more accurate diagnosis. Patterns emerging from these multiple measurements enable the doctor to recognize the signature of a particular ailment and to suggest more targeted measurements if she suspects a certain disease. Only then could she prescribe treatment.

Societal well-being obviously comprises many factors, not just the economic. To foster well-being, policy decisions need to consider as many factors likely to be affected by the outcome as possible. Multiattribute models have been developed to aid this kind of decision making by quantifying the effects of alternative decisions on multiple societal attributes (Gregory 1987).

Multimetric biological indexes calculated from ambient biological monitoring data provide a similar integrative approach for measuring condition and "diagnosing" causes in complex ecological systems. The same logical sequence applies in compiling multimetric economic, health, societal, or biological indexes. First, identify reliable and meaningful response variables through testing; then measure and evaluate the system against expectations; finally, interpret the measured values in terms of an overall assessment of system condition. The resulting index (for economic or biological resources) or diagnosis (for patients) allows people without specialized expertise to understand overall condition and to make informed decisions that will then affect the health of those economies, resources, or patients.

Most multimetric biological indexes for aquatic systems comprise 8 to 12 metrics,[1] each selected because it reflects an aspect of the system's biological condition. These metrics are not independent because they are calculated from a single collection of organisms, just as multiple personal health tests are done on a single individual. But even if metrics are statistically correlated, they are not necessarily biologically redundant. Rather, just as a fever plus a high white-blood-cell count reinforces a diagnosis of bacterial infection, multiple metrics all contribute to a diagnosis of ecological degradation, or ecological "disease." Moreover, when more than one metric points to similar reasons for degradation, there is less uncertainty (Smith 1994). Even when some redundancy exists among metrics, multiple lines of evidence are valuable.

The two most common IBIs for streams have been developed, tested, and applied using fish (Karr 1981; Miller et al. 1988; Lyons 1992a; Fore et al. 1994; Lyons et al. 1995, 1996; Simon 1998) and benthic invertebrates (Kerans and Karr 1994; Kleindl 1995; Rossano 1995, 1996; Fore et al. 1996; Patterson 1996). Both incorporate known attributes from multiple levels of biological organization and different temporal and spatial scales. Typically, patterns

[1] For species-poor environments such as cold-water streams, the total number of metrics is likely to be smaller (e.g., Lyons et al. 1996).

emerge that are the signatures of biological responses to particular human activities (Karr et al. 1986; Yoder 1991b; Yoder and Rankin 1995b).

Following the success and widespread use of these two indexes, similar indexes are now being developed by a number of state agencies to use with invertebrates and vascular plants in wetlands (Karr 1998c); with algae and diatoms in streams (Bahls 1993; Kentucky DEP 1993; Florida DEP 1996; Barbour et al., in press); with diverse taxa in lakes (Harig and Bain 1998; Whittier 1998); and with plants, invertebrates, and vertebrates in terrestrial environments (CRESP 1996; Chu 1997; Bradford et al. 1998; Blair, in press; see also Premise 22). Extending IBI to new taxa, environment types, and geographic areas is like learning to practice medicine in humans, pets, livestock, and others: the expectation of what constitutes "health" depends on the animal, but the same fundamental diagnostic strategy applies in all cases.

Multimetric biological indexes incorporate levels from individuals to landscapes

The success of multimetric approaches such as IBI in assessing biological condition depends on choosing and integrating metrics that reflect diverse responses of biological systems to human actions. Ideally, a multimetric index would cover all such responses, but the costs of developing such an index would be much too high. A set of chosen metrics is necessarily a compromise between "too narrow" and "too broad"; it is also a compromise of choices among conveniently measured biological surrogates of important biological phenomena. Metrics will evolve and expand over the next decade as researchers apply this approach in diverse regions and habitats and for different organisms. Still, a fundamental tenet of IBI is to deliberately choose metrics covering the range of biological signals available from disturbed systems.

IBI is not a community analysis in either of the common uses of the word *community*. IBI does not examine all taxa but is generally based on one or two assemblages (which Fauth et al. [1996] defines as phylogenetically related groups of organisms), such as fish or benthic invertebrates. Neither does a multimetric IBI focus on the community level in the standard textbook hierarchy of biology (individual, population, assemblage, community, ecosystem, and landscape). Rather, the choice of measures in a multimetric index reflects an attempt to represent as many of those levels as possible, preferably directly but at least indirectly. The resulting indexes are likely to produce the strongest multimetric view of biological condition (Table 4). The best multimetric indexes are more than a community-level assessment because they combine measures of condition in individuals, populations, communities, ecosystems, and landscapes.

Table 4. Types of metrics, suggested number of metrics of each type, and corresponding levels in the biological hierarchy. Well-constructed multimetric indexes contain the suggested number of metrics from each type and therefore reflect multiple dimensions of biological systems.

Metric type	Number	Individual	Population	Community	Ecosystem	Landscape
Taxa richness	3–5	✓	✓	✓	✓	
Tolerance, intolerance	2–3		✓	✓		
Trophic structure	2–4			✓	✓	✓
Individual health	1–2	✓				
Other ecological attributes	2–3	✓	✓	✓	✓	✓

Individual level. Individual health manifests itself in many ways both internally and externally, through physiological or morphological signs and metabolic or genetic biomarkers reflecting organismal stress. We have not yet seen reliable metabolic or genetic biomarkers that can be applied broadly in field studies, although in certain situations (see Summers et al. 1997 for a promising example), biomarkers may work as secondary tools for diagnosing biological condition. To date, however, IBI metrics of individual health consist of easily detected external abnormalities; their frequency in an assemblage indicates stress on individuals.

In fish, for example, visible signs of stress include skeletal deformities; skin lesions; tumors; fin erosion; and certain diseases that are associated with impaired environments, especially large amounts of toxic substances. Early studies of fish in the seven-county area around Chicago indicated high incidence of external abnormalities (Karr 1981), a pattern also apparent in Ohio (Yoder and Rankin 1995a). Among benthic invertebrates, head-capsule deformities in chironomids (midges) are strong indicators of toxics (Hamilton and Saether 1971; Cushman 1984; Warwick et al. 1987; Warwick and Tisdale 1988). Anomalies in fish are often used as IBI metrics, but chironomid head-capsule deformities are rarely incorporated into the benthic IBI because so much laboratory work is required to stain individual insects, mount them on slides, and count them.

In other studies, tadpoles collected in a coal ash deposition basin had fewer labial teeth than tadpoles from reference areas (Rowe et al. 1996). They also had deformed labial papillae, which would limit the foods they could eat and thus their growth. Fish in Gulf of Mexico estuaries showed higher numbers and frequencies of several pathologies at heavily disturbed sites than at minimally disturbed sites (Summers et al. 1997). Finally, in a metal-contaminated Rocky Mountain river in Colorado, periphytic diatoms of the genus *Fragilaria* had deformed cells (McFarland et al. 1997). The percentage of deformed cells ranged from 0.2% ± 0.2 to 12% ± 2.0 from low to high levels of contamination by heavy metals (cadmium, copper, iron, and zinc).

Population level. Several metrics in both the fish and benthic IBIs indicate, if not the details of population demography, the relative condition of component groups. Usually, a population must be viable at a site before one can consistently detect a species' presence. For example, the lack of intolerant taxa among fish or invertebrates or of clingers (taxa that cling to rocks) among the invertebrates is a strong signal that populations of these organisms are doing poorly. The absence of darters, sunfish, and suckers among the fishes and of mayflies (Ephemeroptera), stoneflies (Plecoptera), and caddisflies (Trichoptera) among the invertebrates suggests that viable populations of many species within these taxa cannot maintain themselves. The presence or absence of certain age classes, such as large old fish, among coldwater salmonids may also be a useful metric.

Assemblage level. Changes in the chemical, physical, and biological environment resulting from human activities alter assemblages. These changes may appear as changes in species composition or species richness (conventional measures of community structure). They may also appear as altered trophic structure, such as decreases in top carnivores or increases in omnivores, or as shifts from specialists to generalists in food or reproductive habits (reflecting shifts in food-web organization or changing availability of microhabitats for spawning). Multimetric indexes incorporate this information by including metrics such as the percentage of predators, omnivores, or other feeding groups and also species richness and the relative abundance of alien fishes (in streams) or of vascular plants (in wetlands and terrestrial environments).

Considerable theoretical discussion has centered on "functional feeding groups" of North American benthic invertebrates (Cummins 1974; Cummins et al. 1989; Cummins et al. 1995). In particular, according to the river continuum hypothesis (Vannote et al. 1980), the relative abundance of these groups is predicted to change along the length of a river or stream. For example, in comparison with headwaters, which are presumed to receive mostly allochthonous organic matter (from outside the stream), downstream reaches might have more filter-feeders or net-spinning caddisflies taking advantage of high in-stream production. But the river continuum hypothesis does not seem to apply consistently across North American streams (Vannote et al. 1980; Winterbourn et al. 1981; Minshall et al. 1983). Metrics based on functional feeding groups among benthic invertebrates (with the possible exception of relative predator abundance) likewise respond differently in different streams (Karr 1998b).

This inconsistent response differs from what might be a more general pattern of trophic metric behavior in fishes; perhaps the trophic structure of fish assemblages in North America is more consistent than for benthic invertebrates. Alternatively, perhaps more is known about the natural history of fishes, permitting better delineation of feeding groups. Or our knowledge of invertebrates may be less precise, or invertebrates may be more opportunistic. The generality of trophic group response to disturbance deserves more careful analysis, but, meanwhile, be careful. Despite a widely accepted theory, metrics pertaining to functional feeding groups among benthic invertebrates may or may not be good indicators; their dose-response relationships to human influence must be carefully tested and established for multiple data sets and circumstances before they should be used in a multimetric index.

Landscape level. Regardless of level in the biological hierarchy (individuals, species, ecosystem), the persistence of living things depends on heterogeneities in space and time. Spatial heterogeneities are visible in littoral zonation, in vegetation bands associated with water depth in marshes, or in association with soil moisture and slope gradients on drier land. Stream fish spend their lifetime in many microhabitats; they are exposed to different flows and other temporal shifts as days and seasons change. Eggs laid in main-channel gravels become fry hiding in side channels and along the banks. Fry grow into juveniles large enough to avoid the predators that would otherwise eat them. Juveniles may then move into the deep pools those predators inhabit and where food supplies also differ.

Finding food, avoiding predators, seeking spawning habitat—any activity in an organism's life cycle—are subject to and dependent on such heterogeneities in space and time. For some species, the scale of movements may extend only a few centimeters; for others, the scale can be hundreds or thousands of miles. The loss of spatial or temporal components of these heterogeneities can change a species' distribution or abundance, or cause it to disappear altogether. The presence or absence of anadromous or other migratory fishes (e.g., salmon, bull trout) is thus a landscape-level indicator. Dams, alien predators, and altered water flows and temperatures interfere with their movements through a landscape, decimating these species.

Incorporating several multimetric indexes (fish IBI, benthic IBI, algal IBI) into a biological monitoring program is a good way to capture the condition of assemblages that respond to human disturbances at different scales. Different taxa in the same or different assemblage reflect the presence of a broad range of heterogeneities. If top predator taxa needing large home ranges or long-lived taxa requiring years to mature are present, for example, you can infer that the spatial and temporal components they require are also present. Excessive in-stream production or numerous herbivorous fishes or invertebrates would characterize heavily grazed landscapes, where riparian corridors may be damaged and excessive nutrients from livestock wastes are entering the stream.

Development of IBI to date has involved a conscious effort to span the variety of biological contexts. But much remains to be done. Better measures of individual health are needed, as are measures better defining demographics. Strengthening the connections between measures of food web and trophic structure and more-direct measures of nutrient cycling and energy flow would also improve multimetric assessment. Finally, landscape metrics that emphasize overall biological condition (number of native community types or cumulative taxa richness across a watershed) are also needed. Ideally, metrics of landscape condition should be more than a sum of site-specific assessments.

Throughout development and use of multimetric indexes, great care must be taken to measure biological condition, not stressor intensity. Biological surrogates of biological condition are essential; chemical and physical surrogates of biological condition are not adequate by themselves.

Developed and applied properly, the multimetric IBI incorporates and depends on known components of biology—components specific to localities and taxa—across the organizational hierarchy and from disparate spatial and temporal scales. The result is a synthesis of biological signals that reveal the effects of human activities at different levels, in different places, on different scales, and in response to a range of human activities.

Metrics are selected to yield relevant biological information at reasonable cost

The index of biological integrity first developed for fish (Karr 1981; Karr et al. 1986) incorporated 12 metrics from three biological categories: species richness and composition, trophic composition, and individual condition. Later work with both fish and invertebrates led to somewhat different groupings: specifically, species richness, taxonomic composition, individual condition, and biological processes (Karr 1993; Barbour et al. 1996b) and community structure, taxonomic composition, individual condition, and biological processes (Fore et al. 1996). Within each broad category, some metrics are proven for many regions and faunas. Others work in some regions or studies but not in others. Still other potential metrics based on theoretical ecology or toxicology may work but have not been adequately tested, because they are either too difficult to measure or too theoretical to define (Table 5).[2] The categories in Table 5 guide metric selection for new regions, faunas, or habitats, but no metric should become part of a multimetric index before it is thoroughly and systematically tested and its response has been validated across a gradient of human influence.

The choice of how to actually express each metric is as important as selecting the metric itself. You could simply count the number of individuals in a target group and express it as population size, abundance, or density (Figure 19, top); you could determine the proportion, or relative abundance, of the total number of individuals belonging to a target group

[2] Unfortunately, untested or too theoretical attributes have been central to EPA's rapid bioassessment protocols (RBP I, II, III), used since 1989. Many measures incorporated into RBP III were never adequately tested, and tests (Barbour et al. 1992; Kerans et al. 1992; Kerans and Karr 1994; Barbour et al. 1996a; Fore et al. 1996) now indicate that they do not meet rigorous standards for accepting metrics.

Table 5. Sample biological attributes, in four broad categories, that might have potential as metrics. Actual monitoring protocols have proven some of these attributes effective; other attributes may work but need more testing; still others are difficult to measure or too theoretical. Ideally, an IBI should include metrics in each of these categories, but untested or inadequately tested attributes should not be incorporated into the final index.

Category	Demonstrated effective	Need more testing	Difficult to measure or too theoretical
Taxa richness	Total taxa richness Richness of major taxa, e.g., mayflies or sunfish	Dominance (relative abundance of most-numerous taxa)	Relative abundance distribution, after Preston (1962)
Tolerance, intolerance	Taxa richness of intolerant organisms Relative abundance of green sunfish Relative abundance of tolerant taxa	Number of rare or endangered taxa	Chironomid species (difficult to identify)
Trophic structure	Trophic organization, e.g., relative abundance of predators or omnivores		Productivity
Individual health	Relative abundance of individual fish with deformities, lesions, or tumors Relative abundance of individual chironomids with head-capsule deformities Growth rates by size or age class	Contaminant levels in tissue (biomarkers)	Metabolic rate
Other ecological attributes		Age structure of target species population	

(number of individuals in the target group divided by the total number of individuals in the sample; Figure 19, middle); or you could count the number of taxa in the entire sample or in particular subgroups (taxa richness; Figure 19, bottom). You could also determine the proportion of the biota from specific taxa (e.g., number of mayfly taxa/total number of taxa). Approaches vary in their ability to reveal consistent dose-response

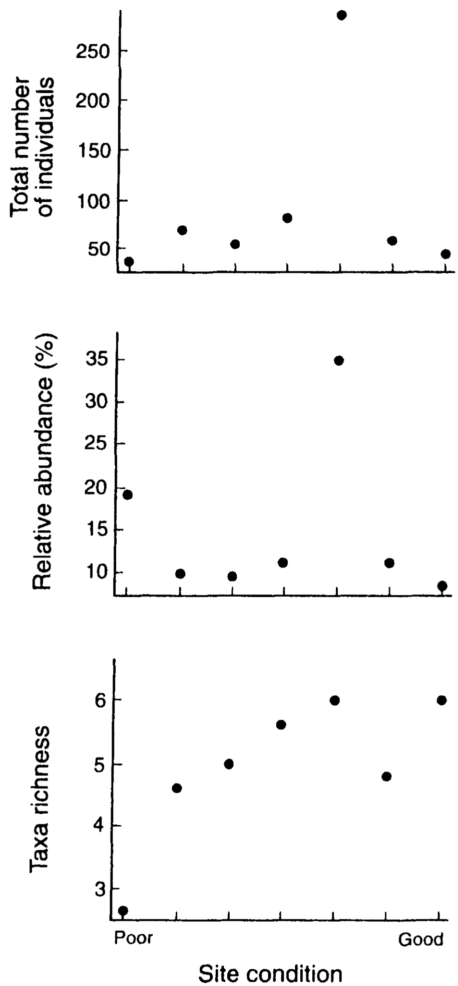

Figure 19. Presence of Trichoptera (caddisflies) in a standard sample, expressed as total number of trichopteran individuals (*top*), relative abundance of trichopteran individuals (*middle*), and richness of trichopteran taxa (*bottom*). These three biological attributes are plotted against riparian condition as an indicator of site condition at seven stream sites in the John Day River basin, north-central Oregon.

relationships, as Figure 19 shows; knowledge of natural history and of which sampling protocols are most efficient should guide your choice.

Population size—besides being difficult and often costly to determine with sufficient precision (Paller et al. 1995b), especially for rare species—is not a good measure because it is naturally too variable, irrespective of human impacts (Karr 1991). Work in Puget Sound lowland streams, for example, found no systematic relationship in two successive years between benthic invertebrate abundance and the percentage of impervious area in the upstream watershed, one measure of human influence (Figure 20).

Similarly, ratios of two groups in an assemblage do not respond systematically to human influence, largely because ratios are composed of two factors that can respond, and thus vary, independently of each other, making it

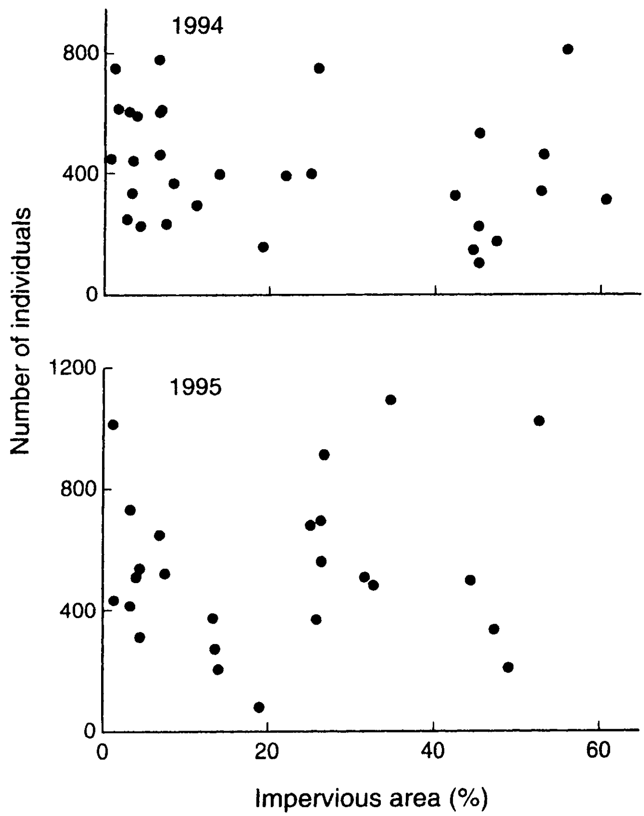

Figure 20. Number of invertebrates plotted against impervious area for lowland Puget Sound streams in two successive years.

impossible to draw firm conclusions about the relationship of the ratio to human influence (see Premise 25). Further, two large numbers and two small numbers may yield the same ratio, although the biological meaning of small and large numbers may be very different (Kerans and Karr 1994). If both components of the ratio are important, it is more appropriate to consider them separately. (This reasoning also applies in the case of diversity indexes, which combine richness and relative abundances. The attributes should be kept distinct with separate metrics.)

Metrics related to feeding ecology or trophic structure are best expressed as relative abundance—for example, the number of individual predators, omnivores, or scrapers divided by the total number of sampled individuals.[3] The relative abundance of organisms at various levels in a stream's trophic organization reflects the condition of the food web, including energy flow and nutrient dynamics, but relative abundances are much easier to measure than true production or energy flow. If we know what to expect from minimally disturbed sites in a region, we can then find the deviations caused by human activities from that expectation. The relative abundance of fish-eating fish in minimally disturbed streams, for example, is likely to be 20% or more; omnivores, 20% or less. In degraded streams, the relative abundance of omnivores is likely to be much higher (>40%).

Major taxonomic groups are best evaluated in terms of taxa richness, because as human activities damage a stream and its watershed, native taxa tend to disappear.[4] A decline in taxa richness is generally one of the most reliable indicators of degradation for many aquatic groups (Ford 1989; Barbour et al. 1995). These include periphyton (Bahls 1993; Pan et al. 1996); phytoplankton (Schelske 1984); zooplankton (Stemberger and Lazorchak 1994); river fish (Karr 1981; Miller et al. 1988; Ohio EPA 1988; Rivera and Marrero 1994; Rodriguez-Olarte and Taphorn 1994; Lyons et al. 1995, 1996; Koizumi and Matsumiya 1997); lake fish (Minns et al. 1994); estuary fish (Thompson and Fitzhugh 1986; Deegan et al. 1993; Weaver and Deegan

[3] Although this metric looks like a ratio, it is actually a proportion—the value of a variable divided by a constant for the sample. Proportions are more reliable as indicators because they are based on a binomial distribution (Fore et al. 1996). In contrast, the ratios of two taxa or two functional feeding groups comprise two variables from the sample; combining them in a ratio has serious statistical consequences, producing a Cauchy distribution (Hannaford and Resh 1995).

[4] Taxa richness can be standardized per unit of area (e.g., taxa per 0.1 square meter) or per count of individuals (e.g., taxa per 500 individuals). The proper choice is hotly debated, a topic we cover in more detail in Premise 29.

1996; Deegan et al. 1997; Hartwell et al. 1997); freshwater invertebrates (Ohio EPA 1988; Reynoldson and Metcalfe-Smith 1992; Kerans and Karr 1994; DeShon 1995; Fore et al. 1996; Thorne and Williams 1997); and marine invertebrates (Summers and Engle 1993; Engle et al. 1994; Weisberg et al. 1997).

An exception to this pattern arises when alien taxa are involved. The presence of alien taxa is a clear indication of human influence. Furthermore, in diverse regions, from 10% to 15% of alien taxa are considered harmful because they have "a significant impact on ecosystem health" (Mills et al. 1998). For example, of more than 4500 alien taxa in the United States, nearly 700 are considered harmful.

Taxa richness may be calculated for an entire sample or for subgroups, such as fish families or insect orders, that use the stream environment in a particular way. Sunfish, for example, feed in the water column or at the surface of pools, whereas suckers feed in benthic pool environments, and darters or sculpins feed in benthic riffle environments. Each requires the unique structural complexity and cover associated with those particular feeding environments; the interactions of cover, structural complexity, and changing food abundances resulting from human actions may cause declines in all these groups. Because their natural histories differ, these three taxa provide information about the condition of three different habitat types within a stream. Loss of sucker taxa points to a problem, such as sedimentation, within the benthic pool environment. Loss of sunfish suggests loss of physical cover and their invertebrate foods from the pelagic and surface zones of pools; indeed, insects decline at the surface when riparian vegetation is lost. Similar information may be gained from the taxa richness of lithophilous ("rock-loving") spawners or nursery species.

Among benthic invertebrates, the taxa richnesses of Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddisflies) reflect different types of degradation. Ephemeroptera taxa are lost when toxic chemicals like those from mine wastes foul a stream (see Figure 18; Hughes 1985; Kiffney and Clements 1994). Plecoptera taxa disappear as riparian vegetation is lost and sediment clogs the interstitial spaces among cobbles. Plecoptera tend to decline at less intense levels of human influence than Trichoptera or Ephemeroptera. Therefore, combining these three taxa into a single "EPT"[5] metric (as in RBP III and others; Plafkin et al. 1989; Lenat

[5] EPT is the sum of the mayflies (Ephemeroptera), stoneflies (Plecoptera), and caddisflies (Trichoptera) found in a benthic invertebrate sample.

and Penrose 1996) may obscure real differences that could help diagnose both the types and sources of degradation at a site.

The signals provided by intolerant and tolerant taxa mean that the best expression of metrics based on these taxa is not the same for intolerants and tolerants. The mere presence of very sensitive, or intolerant, taxa (as apparent from taxa richness) is a strong indicator of good biological condition; the relative abundance of intolerant taxa, in contrast, is difficult to estimate accurately without extensive and costly sampling efforts. Presence alone of tolerant taxa, on the other hand, says little about biological condition since tolerant groups inhabit a wide range of places and conditions. As conditions deteriorate, however, their relative abundance rises (see Figure 22). In general, we recommend that only about 10% (no fewer than 5% or more than 15%) of taxa in a region should be classed as intolerant or tolerant. The point of these metrics is to highlight the strong signal coming from presence of the most intolerant or most tolerant taxa, the two ends of a continuum. We avoid the average tolerance value calculated in biotic indexes because the strong signals of tolerants and intolerants are swamped by the remaining 70% to 90% of taxa with intermediate tolerances.

(For a more statistical rationale for choosing taxa richness and relative abundance, see Premise 20, Figure 34.)

Multimetric indexes are built from proven metrics and a scoring system

Across taxonomic groups, many of the same biological attributes indicate human-induced disturbance (see Premise 14; Table 6). Numerous studies have helped define the most broadly applicable metrics (Karr 1981; Miller et al. 1988; Kerans and Karr 1994; Fore et al. 1996; Scott and Hall 1997; Voshell et al. 1997; see Barbour et al. 1996b for summary table of metrics). After testing in a series of independent studies, 10 attributes of stream invertebrates and 10 to 12 attributes of stream fishes consistently emerge as reliable indicators of biological condition at sites influenced by different human activities in different geographic areas (Tables 7 and 8; see also Table 5).[6]

Consistently reliable metrics include the total number of taxa present in a sample (total taxa richness), the number of particular taxa or ecological groups (e.g., taxa richness of darters or mayflies), the number of intolerant taxa, and the percentage of all sampled individuals (relative abundance) belonging to stress-tolerant taxa (e.g., tubificid worms). Among fishes, a high percentage of individual fish with disease, fin erosion, lesions, or tumors indicates toxic chemicals in a stream. Increased frequency of hybrids seemed a useful metric in early IBI studies (Karr 1981; Karr et al. 1986), although relatively few studies since then have used it successfully. Increased hybridization could indicate a loss of habitat variety and consequent mixing of gametes from different species spawning in a homogenized environment (Hubbs 1961; Greenfield et al. 1973).

[6] The number of metrics in the fish IBI is somewhat smaller in relatively simple systems such as cold-water streams (Lyons et al. 1996). Wetlands may be most appropriately assessed with multiple taxa (e.g., plants, insects, fish, birds) with fewer metrics for each IBI based on a given taxon or assemblage.

Table 6. Regardless of taxon used or habitat sampled, similar metrics respond predictably (✓) to human influence. As human influence increases, taxa richness declines, the relative abundance of generally tolerant organisms increases, and generally sensitive taxa disappear. (Sources: See pages 71–72, Premise 14.)

Taxon	Habitat	Taxa richness	Relative abundance of tolerants	Number of sensitive or intolerant taxa
Fish	River	✓	✓	✓
Fish	Lake	✓	✓ (generalists)	✓
Fish	Estuary	✓		✓ (nursery specialists)
Benthic invertebrates	River	✓	✓	✓
Periphyton	River	✓	✓	✓

The values of metrics such as these provide the best and most complete assessment of a site's condition, but to compare sites and communicate their relative condition to the widest possible audience, metric values at a site are summarized in the form of an aggregate index—the index of biological integrity. Because human actions affect biological resources in multiple ways and at multiple scales, 10 to 12 metrics from four broad categories (see Tables 4 and 5) are selected and then scored using standardized scoring criteria; these metrics are the building blocks of the multimetric index (Karr 1981, 1991; Karr et al. 1986).

Because we now know a great deal about which metrics respond consistently to different levels of human effect, agency biologists with limited budgets do not have to test all attributes to begin using a multimetric index; instead, they can build on studies that have been done before. Nevertheless, whenever more than five sites with different human influences can be sampled, we encourage testing of metric responses in particular locales to see whether the patterns observed elsewhere can be generalized.

Before building a multimetric index, you must convert metric data into a common scoring base. Typically, metrics are quantified with different units and have different absolute numerical values (e.g., numbers of taxa may range from 0 to a few dozen; relative abundances of certain groups may range from 0% to 100%). Also, some metrics increase in response to human disturbance (e.g., percentage of omnivores) while others decrease

Table 7. Potential metrics for benthic stream invertebrates. Metrics that responded to human-induced disturbance as predicted are indicated by a check (✓); those marked with a dash (—) were not tested. Percent sign (%) denotes relative abundance of individuals belonging to the listed taxon or group(s). Metrics marked with an asterisk (*) have been included in a 10-metric multiregional B-IBI (Karr 1998a; see also Table 11). Human influence in the Tennessee Valley consisted primarily of mining and agriculture; in southwestern Oregon, logging and road building; in north-central Oregon, altering riparian condition; in the Puget Sound lowlands, urbanization (measured by percentage of impervious surface); in Japan, multiple human influences; and in Wyoming, recreation.

Metric	Predicted response	Tenn. Valley	SW Ore.	NC Ore.	Puget Sound	Japan	NW Wyo.
Taxa richness and composition							
Total number of taxa*	Decrease	✓	✓		✓	✓	
Ephemeroptera taxa*	Decrease	✓	✓		✓	✓	✓
Plecoptera taxa*	Decrease	✓	✓	✓	✓		✓
Trichoptera taxa*	Decrease	✓	✓	✓	✓	✓	
Long-lived taxa*	Decrease	—	✓		✓	—	
Diptera taxa	Decrease						
Chironomidae taxa	Increase			—	—		
Tolerants and intolerants							
Intolerant taxa*	Decrease	✓	✓	✓	✓	✓	✓
Sediment-intolerant taxa	Decrease	✓[a]	✓	✓	—	—	—
% tolerant*	Increase	✓	✓		✓	✓	✓
% sediment-tolerant	increase	—	✓	✓	—	—	—
% planaria + amphipods	Increase	—	—	—	✓	—	—
% oligochaetes	Increase	✓				✓	
% chironomids	Increase						
% very tolerant	Increase	—	—	—	—	✓	—
% "legless" organisms	Increase	—	—	—	—	✓	
Feeding and other habits							
% predators*	Decrease	✓		✓			✓
% scrapers	Variable	✓		✓			✓
% gatherers	Variable			✓			
% filterers	Variable	✓					
% omnivores	Increase	✓					
% shredders	Decrease			✓			✓
% mud burrowers	Increase	—	—	—	—	✓	—
"Clinger" taxa richness*	Decrease	—	—	—	✓	✓	—
Population attributes							
Abundance	Variable	✓					✓
Dominance*	Increase	✓	✓			✓	✓

[a] Taxa richness at sediment surface

Table 8. Metrics used in the original fish index of biological integrity (IBI) for midwestern U.S. streams and equivalents for more general application.

Original fish IBI	General fish IBI[a]
Number of fish species	Number of native fish species
Number of darter species	Number of riffle-benthic insectivores
Number of sunfish species	Number of water column insectivores
Number of sucker species	Number of pool-benthic insectivores
Number of intolerant species	Number of intolerant species
Relative abundance of green sunfish	Relative abundance of individuals of tolerant species
Relative abundance of omnivores	Relative abundance of omnivores
Relative abundance of insectivorous cyprinids	Relative abundance of insectivores (specialized insectivores)
Relative abundance of top carnivores	Relative abundance of top carnivores
Number of individuals	*Not a reliable metric*
Relative abundance of hybrids	*Not often used successfully*
Relative abundance of diseased individuals	Relative abundance of diseased individuals

[a] Metrics chosen vary as a function of stream size, temperature class (warm-, cool-, cold-water), and ecological factors to reflect biogeographic and other patterns, including sensitivity to different human influences.

(e.g., overall taxa richness). To resolve such differences, each metric is assigned a score based on expectations for that metric at minimally disturbed site(s) for that region and stream size. Metrics that approximate what biologists would expect at minimally disturbed sites are assigned a score of 5; those that deviate somewhat from such sites receive a score of 3; those that deviate strongly are scored 1 (Karr 1981, 1991; Karr et al. 1986). The final index is the sum of all the metrics' scores (Figure 21).

In all cases, the basis for assigning scores is "reference condition," that is, the condition at sites able to support and maintain a balanced, integrated, and adaptive biological system having the full range of elements and processes expected for a region; thus IBI explicitly incorporates biogeographic variation into its assessment of biological condition. In some regions, biologists can actually find and sample from sites that have not been influenced, or have been influenced only minimally, by humans. In other regions, where pristine sites are unavailable, biologists may have to infer reference condition based on knowledge of the evolutionary and biogeographic processes operating in the region (see Premise 31). In still other cases (Fausch et al. 1984; Hughes 1995; Hughes et al. 1998), researchers must depend on historical data, collected when human activity was less, to define reference condition.

Simple, uniform rules for setting scoring criteria—the range of numerical

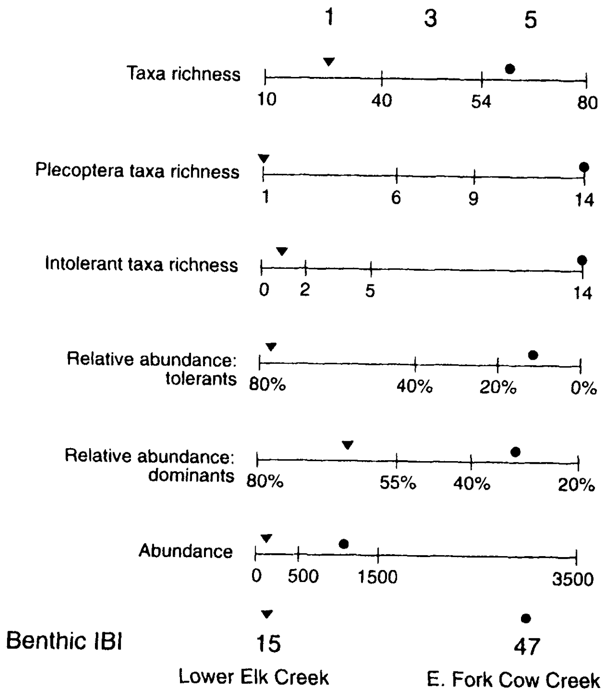

Figure 21. Range and numeric values for six invertebrate metrics from a severely disturbed site (lower Elk Creek, ▼) and a less disturbed site (East Fork Cow Creek, ●) in southwestern Oregon. Because the metrics have different quantitative values, they are given scores (5, 3, 1) to put them on the same scale: 5 indicates little or no deviation from expected, or reference, condition; 3 indicates moderate deviation from expected condition; and 1 indicates strong deviation from expected condition. Vertical lines in the figure represent the cutoff points for assigning these metric scores. Total benthic IBI (B-IBI) value for these two sites equals the sum of these metric scores and five others (from Fore et al. 1996).

values that qualify a metric for a score of 5, 3, or 1—are thus difficult to specify because they depend in part on the sampling design that generated the data. In a hypothetical watershed where one-third of sampled sites were pristine, one-third moderately disturbed, and one-third highly disturbed, the values for each metric could simply be divided at the 33d and 67th percentiles. But human activities tend to homogenize landscapes and living systems so that a majority of sites in a given watershed are likely to be moderately or even severely degraded, such as in the Japanese study illustrated in Figure 22. In the real world, therefore, it makes sense to err on the conservative side by expanding the middle score (3) or even the low score (1) to include more sites rather than fewer, thus making it more difficult for a site to attain a high score.

Natural shifts or breaks in the distribution of metric values can guide the setting of scoring criteria; indeed, scoring criteria should be adjusted to fall at these points, because the points often reflect a biological response. Where metric values increase or decrease linearly across the gradient of human influence (Figure 22, top), as in total taxa richness, the values are typically trisected into three equal divisions, each representing the criteria for assigning a score of 1, 3, or 5. Other metrics, such as relative abundance of

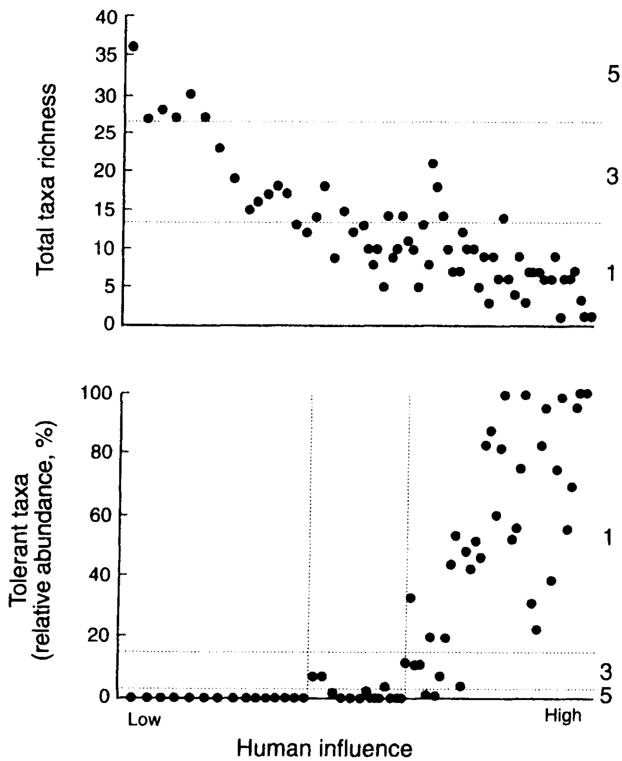

Figure 22. Plots of two sample metrics showing different ways to set the criteria for assigning metric scores of 1, 3, and 5. For metrics with a monotonic, or linear, distribution (e.g., total taxa richness: *top*), one divides into roughly equal thirds the range from 0 to the highest value. For metrics that are not distributed monotonically, one uses natural breaks in the distribution to define score boundaries (shown in the bottom plot by vertical dotted lines). Metric values and classification scheme for human influence come from Rossano (1995) (see also Figures 4 and 5).

tolerant organisms or particular trophic groups, respond in a more skewed pattern (Figure 22, bottom; Figure 23); for these metrics, natural break points suggest setting scoring criteria in unequal divisions. Setting scoring criteria is an iterative process and should be revisited as regional databases and biological knowledge expand.

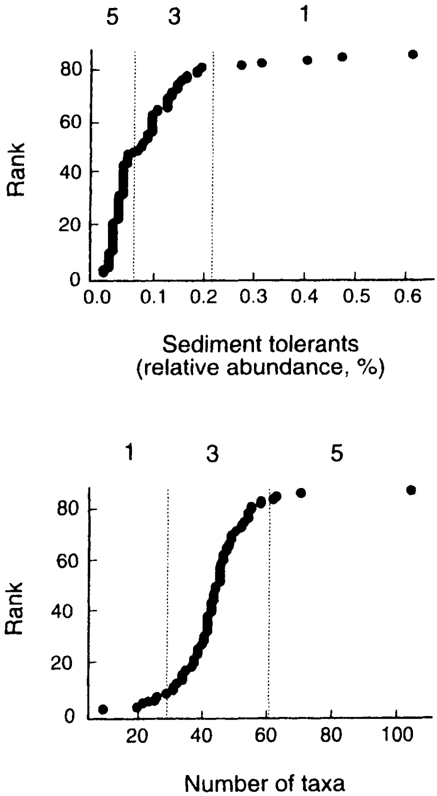

Figure 23. Relative abundance (percentage of sediment-tolerant individuals) and taxa richness (number of taxa) plotted against the rank order of that metric value for 86 stream sites sampled in southwestern Oregon. Dotted vertical lines mark the range of values (scoring criteria) for scoring metrics as 5, 3, or 1. Most sites have near 0% sediment-tolerant individuals; only very degraded sites show higher values of this metric. In other words, the distribution pattern for this metric is skewed. Taxa richness, in contrast, is less skewed. Scoring criteria are divided into unequal divisions for skewed metrics, reflecting a biological response in the data (*top*); the divisions are more equal for unskewed metrics (*bottom*). In both cases, most sites receive a score of 3, the most conservative interpretation of condition.