United States
Environmental
Protection Agency

Science Advisory Board
1400F
Washington, DC

EPA-SAB-EPEC-94-003
November 1993

EPA

AN SAB REPORT: EVALUATION OF DRAFT TECHNICAL GUIDANCE ON BIOLOGICAL CRITERIA FOR STREAMS AND SMALL RIVERS

PREPARED BY THE BIOLOGICAL CRITERIA SUBCOMMITTEE OF THE ECOLOGICAL PROCESSES AND EFFECTS COMMITTEE

U.S. Environmental Protection Agency

NOTICE

This report has been written as part of the activities of the Science Advisory Board, a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

ABSTRACT

On May 13-14, 1993; the Biological Criteria Subcommittee of the Ecological Processes and Effects Committee reviewed the draft document, "Biological Criteria: Technical Guidance for Streams and Small Rivers." Biological criteria (biocriteria) are numeric or narrative expressions that describe the biotic integrity (health) of aquatic communities in minimally impaired reference areas. The Subcommittee concluded that the options presented for selecting reference conditions (i.e., use of reference sites in concert with historical data, empirical models, and expert opinion/consensus) were appropriate. The Subcommittee also supported the use of multiple metrics to evaluate the integrity of aquatic communities, but felt that seasonal variability requires that sampling be conducted at various times of the year. The report stresses the importance of consistent taxonomic identification of biological specimens, use of established museum repositories for curation of voucher specimens, and the importance of developing diagnostic tools to differentiate probable causes of impairment. The Subcommittee also highlights the important linkages between the biocriteria program and other Agency efforts, including the Environmental Monitoring and Assessment Program (EMAP), the Framework for Ecological Risk Assessment, and the Ecoregion Research Program.

KEYWORDS: biological criteria; reference condition; ecological risk assessment; monitoring

U.S. ENVIRONMENTAL PROTECTION AGENCY
SCIENCE ADVISORY BOARD
ECOLOGICAL PROCESSES AND EFFECTS COMMITTEE
BIOLOGICAL CRITERIA SUBCOMMITTEE

CHAIR

Dr. William E. Cooper, Zoology Department, Michigan State University, East Lansing, Michigan

MEMBERS/CONSULTANTS

Dr. Sterling L. Burks, Oklahoma State University, Water Quality Research Laboratory, Stillwater, Oklahoma

Dr. Kenneth W. Cummins, University of Pittsburgh, Pymatuning Laboratory of Ecology, Linesville, Pennsylvania

Dr. Kenneth L. Dickson, University of North Texas, Institute of Applied Sciences, Denton, Texas

Dr. Charles P. Hawkins, College of Natural Resources, Utah State University, Logan, Utah

Dr. Robert J. Huggett, Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, Virginia

Dr. Richard Kimerle, Monsanto Corporation, St. Louis, Missouri

Dr. Alan W. Maki, Exxon Company, USA, Houston, Texas

Dr. Dean B. Premo, White Water Associates, Inc., Amasa, Michigan

Dr. Stephen T. Ross, Dept. of Biological Sciences, University of Southern Mississippi, Hattiesburg, Mississippi

Dr. Joseph Shapiro, Limnological Research Center, University of Minnesota, Minneapolis, Minnesota

<u>SCIENCE ADVISORY BOARD STAFF:</u>

Stephanie Sanzone, Designated Federal Officer, Science Advisory Board, U.S. EPA, 401 M Street, S.W. (1400F), Washington, D.C. 20460

Marcia K. Jolly, Staff Secretary, Science Advisory Board, U.S. EPA, 401 M Street, S.W. (1400F), Washington, D.C. 20460

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY

The Biological Criteria Subcommittee of the Ecological Processes and Effects Committee has reviewed the draft document, "Biological Criteria: Technical Guidance for Streams and Small Rivers." The draft guidance is an important step in the effort to develop scientifically-credible biological criteria (biocriteria) as a tool for assessing the biotic integrity, or health, of stream communities. Biocriteria offer the opportunity for EPA and the states to evaluate and demonstrate the adequacy of current regulatory and management activities in protecting aquatic ecosystems. The current state of the science is sufficient to support application of biocriteria for assessment of site-specific impacts and regional trends, but not for establishing point source criteria or permit limits. In order for biocriteria to be more broadly useful, the Agency should support the development of diagnostic tools to differentiate probable causes of observed changes in biological metrics. These tools would be analogous to the Toxicity Identification Evaluation procedures associated with the Whole Effluent Toxicity program.

We agree that the assessment of biological integrity should rely on multiple metrics, and multiple assemblages of organisms. However, we discourage the use of a single aggregate criterion for any site, since important changes in individual metrics may be masked by aggregation into a single index. To assist states with analysis and aggregation of multiple metrics, the discussion of statistical methods in the guidance document should be expanded. In addition, the guidance should place greater emphasis on the statistical design of the bioassessment sampling program, rather than focusing primarily on post-monitoring data analysis.

The definition of reference condition using reference sites is appropriate when used in conjunction with historical data, empirical models, and expert opinion/consensus. Definition of the reference condition, and detection of impacts at test sites, will require sampling at a variety of temporal and spatial scales in order to account for the natural variability of biological systems.

Taxonomic identification of collected specimens will be a critical task. The guidance should establish minimum levels of taxonomic resolution to be achieved for various assemblages, and the Agency should take an active role in ensuring that there is an adequate, consistent level of taxonomic training among various state biocriteria programs. With regard to voucher specimens from the biological monitoring, the Agency should strongly encourage states to use the established network of federal, state, university and other museums for regionally-centralized curation.

The guidance should also include several real case studies of the application of biocriteria for assessment on a watershed scale, additional information on costs associated with development and implementation of biocriteria, and cost-saving measures which have been adopted by state biocriteria programs.

2. INTRODUCTION

At the request of EPA's Office of Science and Technology in the Office of Water, the Biological Criteria Subcommittee of the Ecological Processes and Effects Committee met on May 13-14, 1993, to review the draft document "Biological Criteria: Technical Guidance for Streams and Small Rivers" (April, 1993). This document builds on the Agency's past efforts to assist states to develop and apply narrative and numeric biological criteria (biocriteria), as tools for water resource management which complement existing chemical and physical criteria (see, for example: Plafkin et al., 1989; EPA, 1990; 1992b). The Agency is currently developing additional biocriteria guidance documents for lakes/reservoirs and estuaries, and plans to develop similar guidance for large rivers and wetlands in the future.

In reviewing the draft guidance on biocriteria for streams and small rivers, we were asked to consider four questions:

a) Are the options presented for selecting reference conditions scientifically sound?

b) Does the recommended approach to sampling frequency adequately account for seasonal variability?

c) Will the recommended multimetric approach to evaluating aquatic communities adequately characterize the resource?

d) Does the recommended approach to selection and aggregation of biological information provide a sound basis for the development of biocriteria?

3. GENERAL COMMENTS

3.1 The Importance of Biocriteria

Maintaining the integrity of community and ecosystem structure and function should be the ultimate goal of environmental regulation. Stream and small river habitats are integral components of the terrestrial ecosystem (watershed) and must be managed within that context. The development and implementation of biocriteria for streams and other aquatic systems will allow EPA and state agencies to evaluate and to demonstrate the adequacy of current regulatory and management activities in protecting aquatic ecosystems. We applaud the Agency's efforts in this direction.

Traditional end-of-pipe, command and control regulation has tended to focus on single stressors at specific sites. In contrast, biocriteria are response variables that provide an integrated picture of the effects from multiple stressors (chemicals, sedimentation, exotic species, etc.) arising from point sources, nonpoint sources, habitat alteration and hydrological modification. Ecological responses observed at the community level of organization, for example, offer dependable and readily-observable indicators that integrate the impacts of multiple, and often subtle, stressors. In addition, the community responses reflect both the adaptive ability of populations and the resilience of the community to perturbations, and thus can provide insight into the ecological significance of observed changes in habitat and water quality.

In a real sense, the application of biocriteria to streams and small rivers is an assessment of risks to stream communities. The Agency has developed a guidance document, "Framework for Ecological Risk Assessment," that is intended to serve as the conceptual framework for currently evolving initiatives throughout the Agency (EPA, 1992a). Within that framework, the biocriteria guidance should more fully identify the linkages, both conceptual and procedural, between the two documents to highlight the risk assessment nature of the use of biocriteria for streams. The ecological risk assessment approach will be particularly important in diagnosing the causes of observed changes in biological metrics, and separating anthropogenic impacts from natural fluctuations in stream conditions (e.g., dissolved oxygen, water flow, and temperature). There are too many possible permutations and combinations of stress/response pathways to allow trial and error risk management. Rather, an adaptive management approach based on hierarchical diagnostic analysis needs to be developed.

3.2 Applications of Biocriteria

The guidance document needs a clear up-front definition of the Agency's goals for the biocriteria program. Potential applications of biocriteria range from use as an assessment and screening tool, to use in defining and evaluating permit limits for point source dischargers, to inclusion of the criteria in permits themselves. We feel strongly that the current state of the science underlying the identification of reference sites and the selection of suitable biological measures of stream communities limits the utility of biocriteria at this time to two critical applications. First, the site-specific assessment of ecological degradation using a valid reference site as a baseline is justifiable. This would include the use of biocriteria as a site-specific measure of ecosystem response to remediation or mitigation activities. Convergence of the biocriteria at the test site with the reference site will indicate when the recovery is sufficient and the corrective activity has been successful. Second, biocriteria may be used to assess biological resource trends in well characterized watersheds. Representative watersheds from different regions could be used to conduct trend analyses similar to those incorporated into the Agency's Environmental Monitoring and Assessment Program (EMAP).

With regard to the application of biocriteria in a regulatory context, we see several factors as limiting at this time: the degree to which we can detect subtle impacts; the current lack of diagnostic tools to determine the causes of observed impacts; and the state-of-the-science in defining ecoregions and reference areas. We address the need for diagnostic tools and improved classification schemes in subsequent sections of this report. In general, however, we feel that the methodologies have not yet undergone a sufficiently rigorous scientific evaluation, from a toxicological "cause and effect" perspective, to be used for establishing point source criteria or permit limits. This does not preclude the use of biocriteria as a way of judging the combined effectiveness of current point and nonpoint source controls (e.g., numerical chemical limits, whole effluent toxicity tests, and total maximum daily load--TMDL--assessments) in protecting biological integrity.

3.3 Diagnosing Causes of Impairment

While the draft document provides detailed guidance on how to establish biocriteria, it includes very little discussion of how to determine the probable causes of impacts on biological communities (stress/response relationships). Observed biological conditions at assessment sites can have multiple causes. Thus, it is essential that guidance on identifying and prioritizing the probable causes be developed. We suggest that an approach analogous to the Toxicity Identification Evaluation (TIE) procedures associated with the Whole Effluent Toxicity program be developed for the biocriteria program. Such an approach could include activities that enhance the interpretation of data collected in the biological assessment (e.g., use of biomarkers) and guidance on sampling design required for diagnosing probable causes of impairment. We recommend that the section "Identification of Impact Types" in Chapter 8 of the draft document be augmented to include guidance on diagnosing probable causes. However, the Agency should consider developing a separate (companion) guidance document on diagnosing probable causes of impairment in streams and small rivers. The diagnostic approach must embody the elements of the Agency's Ecological Risk Assessment paradigm, including evaluation of background chemical monitoring data, habitat alteration, and introduction of exotic species, to ensure that observed biological changes can be related to specific stressors vs. natural fluctuations.

3.4 The Need for Real Case Studies

We agree with the need to include case studies that illustrate the application of biocriteria on a watershed basis. However, the draft guidance contains only a hypothetical example. The inclusion of one or more "real world" case studies would be most effective in underscoring the utility of the approach and demonstrating to potential users that it has been successfully applied in the field. The case studies selected should represent a geographic balance (the current document has an "Eastern bias" in terms of examples and literature cited), and should include detailed information on implementation costs. In the current era of ever-dwindling budgets faced by most state agencies, one of the most important obstacles facing the adoption of a new program is the incremental cost, or cost savings vs. the status quo. Therefore, the guidance should include information on start-up and unit operating costs, incremental costs to attain increasingly refined levels of taxonomic accuracy, and cost-saving measures that have been adopted by states where a biocriteria program has been developed.

4. DEFINING THE REFERENCE CONDITION

4.1 Ecoregions and Other Classification Schemes

Effective classification and pairing of reference sites with test sites is essential to define reference condition and to detect impairment. Past research by the Agency to define ecoregions based on land use, land form, natural vegetation, and soil type (Omernik, 1987) has advanced our ability to identify areas having similar biotic assemblages and communities. Such approaches have application not only to the development of biocriteria, but also to broader Agency concerns such as remediation and restoration of damaged aquatic and terrestrial sites, watershed protection programs, and ecosystem approaches to environmental management.

However, we are concerned that the Agency is no longer supporting research to refine ecoregional classification techniques. With regard to streams and small rivers, more attention should be paid to factors that vary longitudinally along stream ecosystems (e.g., riparian corridors, temperature, hydrology, and channel geomorphology) and that may cross ecoregion boundaries and have more direct effects on stream communities than general ecoregion characteristics (Cummins et al., 1989; Statzrer and Higler, 1986; Vannote et al., 1980). In some parts of the country, the distribution of aquatic organisms, particularly invertebrates, tends to transcend traditional ecoregion boundaries. Therefore, classification of stream reference sites based on currently used ecoregion criteria may be inappropriate. We recommend that the Agency explore other ways of classifying reference sites that explicitly recognize biogeographic and distributional patterns of stream and riparian biota. In addition, the guidance should encourage states to overlay airsheds, since many pollutants are transported into watersheds by air.

4.2 Approaches for Defining Reference Condition

Defining the reference condition is one of the most critical aspects of biocriteria development since the reference condition describes the baseline against which test sites will be evaluated. The draft guidance document identifies four approaches to establishing the reference condition: use of reference sites; use of historical data; use of empirical models; and use of expert opinion/consensus. The draft guidance has a detailed discussion of the use of reference sites, describes criteria for their selection, and reviews approaches for classifying resources. However, the document gives only minimal guidance on how to use or apply the other approaches for establishing the reference condition. This issue will be particularly important when biocriteria are developed for large rivers, lakes and reservoirs, and estuaries, because it will be increasingly difficult to identify minimally impacted reference sites. In such cases, greater reliance will have to be placed on historical data, empirical models and expert opinion to define the reference condition. In addition, we suggest that the four approaches be used in combination where possible to provide convergent evidence of the reference condition. As currently written, the guidance suggests independent application of the approaches.

When using reference sites to establish the reference condition, the natural variability of communities (over seasonal, annual and longer time scales) must be considered. Such variation can be of high magnitude and often appears to be stochastic (e.g., McElravey et al., 1989). Thus, the final guidance document should discuss the feasibility and costs of establishing the long term collection records which may be needed to establish the range of conditions characteristic of an unperturbed reference site.

While we support the use of historical data to help define biological expectations for an area, the discussion on limitations of this approach should be strengthened. Data sets often contain species biases depending on the purpose for which the information was originally collected. For example, fish data sets compiled by agencies interested in managing sport or game fisheries often do not identify forage and other non-game fish, museum collections of fish often have a bias against game or larger non-game fish, and natural heritage inventories are often biased toward rare species. In addition, sampling efficacy and thoroughness are not always known for these data sets. Quite often, historical distributions do not accurately reflect where organisms occur, but rather where people collect (e.g., at bridges). This means that historical data may actually be from more disturbed (i.e., easily accessible) sites of the time.

The authors of the guidance document may find it desirable to address "natural biodiversity" of the reference condition and adopt the concept as the ideal reference condition. This condition can be estimated by using our best data on what organisms were historically present in or near the reference area. This historical information may be best for fish and is often available in archival collections. Our "best historical data," combined with more recent survey data, and scientific judgement, may provide the most appropriate and firm baseline for a reference condition. It should be acknowledged, however, that past societal decisions (e.g., introduction of "desirable exotic species" or land use/urban development) may make the goal of achieving the natural biodiversity ideal condition impossible.

The issue of the introduction (intentional or unintentional) of exotic species is controversial, and should be dealt with more specifically in the guidance. The guidance currently excludes exotic species from the definition of reference condition. Defining the reference condition based on naturally occurring species is one way of avoiding a potential downward spiral of reference sites over time. On the other hand, the reality in many regions is that the dominant species in the community are in fact non-native introduced species. To discount such dominant, pervasive species in the definition of reference condition may be to ignore the reality of past introductions. We agree, however, that decisions to introduce exotic species may be taken as the result of societal decisions, but cannot be justified on ecological grounds.

In addition, most biocriteria indices do not incorporate a measure of the subjective aesthetic or perceived value of populations of specific organisms. Whereas the concept of ecologically equivalent species (e.g., with respect to trophic level or niche) may reflect a scientifically objective approach to community structure/function, it may be unrealistic from a socio/political perspective. For example, recreational fishermen may not view ecologically equivalent species as equally desirable. This concept of the aesthetic value of populations may need to be addressed in the selection of biological metrics.

4.3 Chemical Characterization of Sites

Another consideration when selecting reference sites is the extent to which the sites may be adversely impacted by anthropogenic chemicals transported by water or air and concentrated primarily in the substrate. Chemical contamination could seriously skew the baseline definition of biotic integrity reflected in the reference condition. The final document should include guidance on how to determine whether or not a potential reference site is "chemically appropriate." Some minimum set of chemical data (perhaps a full scan of halogenated organic compounds and a suite of petroleum hydrocarbons and trace elements) should be required from all sites, both reference and test locations. The guidance document should also discuss the influence of total hardness and pH upon the relative toxicity of various metals, pH upon chemical species of ammonia, and other attenuating factors (presence of organic chelating agents, suspended solids, etc.) It may be possible to use chemical data collected by the Agency's Environmental Monitoring and Assessment Program (EMAP) to satisfy some of these data requirements.

4.4 Protection of Designated Reference Areas

Once reference areas have been selected that are representative of a region, the biotic integrity of these areas should be protected. This means, in part, that management of the areas should be coordinated with other government agencies (both federal and state). For example, while the draft guidance document states that forested ridge tops are usually the least impacted sites in the Appalachian region, the U.S. Forest Service specifically targets these areas for chemical application (e.g., Dimilin spraying for gypsy moth control). In addition, the natural variability of biological communities (over seasonal, annual, and longer time scales) may require long-term collection records to establish the range of conditions characteristic of reference sites. For these reasons, reference sites should be selected, wherever possible, from dedicated research areas (e.g., National Science Foundation's Long Term Ecological Research (LTER) sites, Department of Energy facilities, U.S. Forest Service research areas, and national parks) where the integrity of the watershed and the availability of baseline data will be supported over the long term. In addition, states should notify other agencies and academic institutions of the location of reference sites, perhaps through an Agency-sponsored Register of Biocriteria Reference Sites. Even with attempts to protect designated reference areas, there will inevitably be "drift" in the baseline reference condition. It will be important, therefore, to monitor trends at reference sites.

5. SAMPLING DESIGN AND IMPLEMENTATION

5.1 Design and Statistical Considerations

The draft guidance document does not adequately address the up-front planning steps necessary to design an appropriate biological sampling program. These steps include a decision on the goal of the program (site-specific impact assessment vs. trends assessment), definition of the universe to be assessed, and the level of precision needed for management decisions. The answers to these questions will affect the intensity and pattern of sampling, the level of quality assurance and quality control (QA/QC) required, the type of sample analysis, and ultimately, the cost of implementation.

The discussion on conducting biosurveys (Chapter 4) does not provide sufficient guidance on: a) how to identify the universe of concern (i.e., the test site or reference location); b) the a priori determination of the level of difference between a reference and test site one wants to detect, or can expect to detect with the use of biocriteria; and, c) the determination of sampling sufficiency. These issues have both a spatial and temporal component. While the temporal component is discussed in some detail in the Technical Issues section, issues related to spatial variability are less well developed. Much of the discussion focuses on choice of a particular habitat or substrate type, and avoids the larger issue of sample representativeness or the units of sampling.

Given that there will always be species biases of different sampling gears, some basic criteria need to be included to determine when a stream section has been sufficiently sampled. Because species richness and representation of trophic categories (e.g., herbivores, carnivores) are critical to the determination of the Index of Biotic Integrity (IBI), this is not a trivial issue. While the intent of sampling should not be the total documentation of species richness, which would require many samples over a long period of time, sampling should be sufficient to reveal the majority of species. Due to the annual dynamics of streams, characterization of both reference and test sites should be based on sampling at various times of the year to include the major components of the fall-winter and spring-summer (or wet season-dry season) communities.

Ideally, sampling should be based on random, stratified, or fixed interval arrays that will generate an unbiased estimate of the assemblage structure that occurs in the universe in question. A sufficient number of samples should be taken to ensure that within-site variance is sufficiently low to allow detection of whatever a priori defined level of between-site difference is required. Units of sampling could be linear stream sections or time blocks of sampling.

The guidance document should refer to available methods for determining the appropriate sample size. For example, simple approaches to determining sampling sufficiency include: a) fitting of species abundance data to lognormal distributions (Ludwig and Reynolds, 1988); b) construction of a species sample curve that plots cumulative number of species against cumulative number of samples (Brower et al., 1990); and, c) use of Elliott's precision estimate (e.g., number of samples required to achieve a variance of ± 40% of the mean) (Elliott, 1977).

Clearly, if sampling is restricted to one (or only a few) habitat types and to one time of the year, the resultant biased representation of the community will compromise the ability of any analysis to reveal true differences (or lack thereof) between test and reference sites. Some examples illustrating the consequence of selective sampling would be useful. For example, if human use has resulted in a shift in the abundance and frequency of major channel geomorphic forms (e.g., riffles and pools), sampling a single habitat type (e.g., riffles) would reveal little of the actual system-level differences in biotic structure between sites and little of the real change in structure that would occur over time with either future degradation or site improvement. These issues are especially problematic in western and/or mountainous streams that are geomorphically and hydrologically complex (chaotic according to geomorphologists), and in which the primary stressors are associated with landscape alterations. For specific guidance on sampling design, consult Elliott (1971), Green (1979), Hurlbert (1984), and Waters and Erman (1990).

A related question is that of determining when statistically significant changes in biocriteria are ecologically significant with respect to long-term stability of the ecosystem. It is well understood that biological communities are always in a state of flux, oscillating around some general community mean level of abundance and organization. Care must be taken to fully describe these natural oscillations and delineate them from those changes brought about by degradation of water quality, habitat, etc. The guidance document should reference the series of papers being developed by the Agency's Risk Assessment Forum which address the issue of ecological significance.

The guidance document should also reference recent advances in the ecological impact assessment field to provide more meaningful statistical descriptors of populations and communities than have traditionally been available through normal theory statistics. The application of binomial theory tests and packages such as GLIM (Generalized Linear Interactive Model) have significantly advanced our ability to test for effects of confounding variables and covariates that have traditionally plagued ecologists attempting to compare structural and functional measures between impacted and unimpacted communities. A state-of-the-science review of these techniques, and examples of their application, should be considered for addition to the document.

5.2 Quality Assurance and Quality Control

We appreciated the inclusion of the quality assurance/quality control (QA/QC) section in the guidance and the discussion of elements that should be included in all field biological assessments. However, a few issues are presented here for consideration. First, many of the biologists that will be implementing the biocriteria program at the state level may not be familiar with a fully-documented field assessment QA program. Therefore the guidance should include some specific examples of a biological sampling QA program, including cost estimates. The QA Plan developed for biological data by EMAP may provide relevant examples.

Second, it is critically important to ensure consistent quality in the taxonomic identifications in all field biological assessment programs. We agree with the statement on page 63 of the draft guidance that "a major factor influencing the comparability of field ecological projects is the type and intensity of appropriate training and professional experience for all personnel." The frequent reference in the guidance to the use of "volunteer" or "lay" personnel, however, seems to contradict the focus on adequate training. We urge that references to the use of lay personnel be removed from the guidance, or that a full discussion be added on the impacts that using nonprofessional personnel will have on the QA/QC of data collected.

Lastly, we agree with the need to tailor the depth of the QA program to the objectives of the bioassessment program. We recommend that the guidance offer suggestions on tiers of effort, as can be found in the EPA's Good Laboratory Practices (GLP) program, which would allow QA/QC Plans to be customized according to the individual needs of the state programs. For example, bioassessment programs that will support significant economic and regulatory decisions may need a more rigorous QA program than ones used as an assessment or screening tool.

5.3 Taxonomy and Reference Collections

The guidance needs further discussion of the training requirements for personnel conducting field sampling, and guidelines for sample identification and handling. Successful implementation of the biocriteria program is dependent upon accurate taxonomic identification. Personnel involved in field sampling and work-up of collections should have training in the systematics of the group(s) for which they are responsible. It is unlikely that resource agency personnel with skills in identifying macroinvertebrates or fish will have competence in identifying birds by visual or auditory cues, or mammals by the common signs of tracks and scat. (The statement on page 51 of the draft document that "tracks and droppings also provide easily attainable survey data" is misleading; in fact, very few field staff can reliably identify these mammal signs).

Although field sorting and observations are important to record behavioral and nonstructural color pattern characters that are lost after preservation, major groups (aquatic plants, periphyton, invertebrates and fish) should generally be identified in the laboratory, following field collection. Exceptions might include some game fish, many mammal species, and most birds. Collection of protected species (threatened, endangered, or otherwise protected) of vertebrates where field identification is possible should be avoided and conducted only with appropriate collecting permits. Many protected plant species would have to be properly collected for identification by botanists (again, only with appropriate permits).

The guidance should also recommend varying levels of taxonomic identification for different levels of intensity (different applications of the biosurvey results), for example:

> Level 1--Screening Assessment Protocol: field identification to the generic level for fish and familial or generic level for macroinvertebrates, as in the rapid bioassessment protocol for streams (Plafkin et al., 1989).
>
> Level 2--Site Assessment Protocol: identification of juvenile/adult fish to species, and benthic macroinvertebrates at least to the species level taxonomic levels indicated in Table 1.

In addition, the guidance should include recommendations on reasonable levels of taxonomic resolution for various categories of organisms. Field collection and systematics should be driven by protocols, established in the peer-reviewed literature, that are realistic. That is, the taxonomic resolution should be set at a level achievable by appropriately trained state personnel. Stream ecology research over the last decade indicates that a specific minimal level of resolution should be set ("lowest achievable taxonomic level" is not a helpful criterion) and that additional refinement should be left to individual state groups as their capabilities permit. Proposed levels of intensity and taxonomic resolution must receive a thorough evaluation by the scientific research community.

In most cases, adult and juvenile fish should be identifiable to species. While identification of larval fish may also provide useful information, it may only be feasible to identify to the generic or familial levels. Reasonable candidate levels for stream macroinvertebrates are given in Table 1.

Table 1. Proposed Minimal Levels of Taxonomic Resolution for Stream Macroinvertebrates

Taxonomic Level	Groups
Genus	Plecoptera (in part), Ephemeroptera, Odonata, Trichoptera, Megaloptera, Neuroptera, Lepidoptera, Coleoptera (in part, larvae and adults), Hemiptera, Diptera (Tipulidae and Simuliidae), Crustacea, Mollusca
Tribe	Chironominae
Subfamily	Chironomidae
Family	Diptera (other than Tipulidae and Simulidae), Oligochaeta, Plecoptera (in part), Coleoptera (in part)
Order	Other non-insect groups

The discussion of sample collection and processing should be expanded. As noted in the guidance, all specimens of a single species from each collection area, from a single date, should be assigned a unique field number. The number should be placed inside the container with the specimens, rather than being written or attached to the outside of the container. Information for each collection should be recorded in a field notebook containing, at a minimum, the following: field number, date, collectors, site location (state, county, township, range, section, and a verbal description of sampling location), sampling gear used, time of sample, and habitat characteristics.

Once the samples have been analyzed (identifications, enumerations, measurements, etc.), reference (voucher) material should be placed in the well established network of federal, state and university museums for regionally centralized curation. This ensures a second level of quality control in terms of specimen identification. Preferably, collection and identification of voucher specimens would be coordinated with taxonomic experts in regional museums. These repositories, which have always been the centers for systematics, should continue to be used for this function. Funds must be made available to these institutions to insure their participation in the process. The Agency should work with the U.S. Department of Interior as the new National Biological Survey is established to ensure that samples collected by state biocriteria programs are deposited in established museum repositories.

Once the information on the samples has been entered into a data base and verified, the repository institutions should be encouraged to conduct additional systematic studies on the material. Information from these additional analyses can then be made available to state biocriteria programs.

6. CHARACTERIZING BIOLOGICAL COMMUNITIES: MULTIMETRIC APPROACH

6.1 Use and Misuse of Indices

We agree that the use of multiple metrics in biocriteria will ensure a more balanced assessment of biotic integrity than any single measure. However, we caution the Agency about the loss of information that can result from aggregation of metrics, and the possible misuse of indices. One danger is the use of judgmental labels to characterize different value ranges for a metric or index. For example, the long-needed relationship between nutrient loading and trophic state of lakes provided by Vollenweider (1966) was grossly misused because of his use of the terms "permissible" and "dangerous." Because of this terminology, arguments ensued both in and out of court over whether additional nutrient loading would be "permissible" (i.e., would not cause a body of water to become eutrophic), or would be "dangerous" (i.e., would lead to a eutrophic condition). In fact, all nutrient additions contribute to eutrophication, and there is no threshold effect.

A second possible misuse of indices is for users to presume too high a precision. For example, the Carlson Trophic Scale Index (TSI) for lakes (Carlson, 1977) was originally conceived as a 1-10 scale, but was converted to a 10-100+ scale when it was pointed out that an index of 5.4 implied a greater degree of precision than 54 (i.e., once a decimal is introduced, people suspect greater precision than may exist).

A third misuse can occur when individual metrics are aggregated into a single index value. Metrics derived for different assemblages (e.g., fish, macroinvertebrates and periphyton) may respond in different ways to the same stressor. Therefore, important changes in individual metrics may be masked by aggregation into a single index. For these reasons, we recommend that the guidance discourage attempts to develop a single aggregate biocriterion for any particular site.

6.2 Selection of Metrics

In the discussion on selecting assemblages to measure, the draft guidance document appears to focus primarily on fish and benthic macroinvertebrate community structure and function as indicators of water quality. The case studies should emphasize the importance of including the periphyton and decomposer communities as well. For example, the potential impact of nutrient enrichment, a major problem affecting streams (as well as lakes/reservoirs and estuaries), may be underestimated if the periphyton community is not considered. In order to assess the temporal and spatial effects of nutrient enrichment, all trophic levels should be evaluated. In addition, the physical-chemical features of the habitat (i.e., substrate type, flow, depth, stream order, shading, water hardness, alkalinity, pH, temperature, conductivity, and dissolved oxygen) must also be considered when interpreting metrics or indices.

6.3 The Role of Biomarkers

The draft guidance document refers to the potential value of biomarkers when comparing the ecological health of potentially impacted sites with that of reference areas. Biomarkers, in this context, are defined as biochemical, physiological or histological markers of anthropogenic stress. We agree that biomarkers (e.g., protein induction or tissue lesions in finfish) may provide useful data to augment more classical measures of ecosystem health such as diversity, species richness, IBI, etc. However, care must be exercised to ensure that biomarker assay results are not misinterpreted or misused. The current state of the science is not advanced enough that biomarkers can be routinely used in biocriteria, or as a diagnostic tool.

Biomarkers can be divided into two broad categories: those that reflect or indicate exposure to stress, and those that denote an effect of exposure(s). For instance, induction of the enzyme P_{450IA1} in the liver of a finfish can be indicative of exposure to polynuclear aromatic hydrocarbons, PCB, etc. However, induction of P_{450IA1} does not necessarily mean that the organism has been adversely impacted. On the other hand, biomarkers of immune function, such as macrophage phagocytosis, may indicate a harmful biological response to anthropogenic stress since a reduced phagocytic activity relates, at some level, to the animal's ability to cope with infectious particles, etc.

We encourage the Agency to pursue, with appropriate caution, the feasibility of using biomarkers in the establishment of biocriteria. Additional research will likely enhance the utility of biomarkers in the biocriteria program.

7. TRAINING AND TECHNICAL ASSISTANCE

In order for the implementation of biocriteria programs to be successful, the Agency must play an active role in providing training and technical assistance to state agencies. At the current time, there is a wide spectrum of abilities and expertise in the state agencies that will be responsible for developing biocriteria. In recognition of this fact, the Agency should prepare a reference document on how to: a) assess the educational needs of state agency staff; b) develop in-house educational programs; and, c) fund outside educators (universities, consultants, or other agencies) for training programs. This companion document could also be a source book of museums, reference collections, and people with scientific expertise relevant to specific taxonomic or ecological needs. Without some education component and source of information and help for resource managers and regulators, development and implementation of biocriteria could overwhelm state agency technical resources.

We also recommend that the Agency consider the following steps to support states in the development of biocriteria programs:

a) continue development of the BIOS and STORET data management systems to provide centralized storage and access for state biological data;

b) encourage states to use the existing repository network of academic, state and federal museums for curation of type specimens of macroinvertebrates and fish;

c) establish a network list of taxonomic experts that could "verify" identifications (these experts should be readily available to confirm identifications, and therefore an equitable fee should be established for appropriate level of effort, i.e., graduated fee depending upon degree of difficulty in identifying species); and,

d) provide workshops to assist training of state personnel.

8. CONCLUSIONS AND RECOMMENDATIONS

We compliment the Agency on the quality of this initial guidance on developing biocriteria for streams and small rivers. As the latest in a series of documents from the Office of Water on this topic, the document clearly shows the continuing maturity of the Agency's approach to assessment of biological integrity as a measure of the effectiveness of water quality regulation. Our response to the four questions posed by the Agency in the charge to the Subcommittee are as follows:

a) **Are the options presented for selecting reference conditions scientifically sound?**

Yes, although the document may overemphasize the role of reference sites as the primary basis for defining the reference condition. The final document should include greater detail on the use of other approaches (historical data, empirical models, and expert opinion/consensus) for affirming reference condition based on reference sites, as well as for establishing the reference condition in degraded areas.

b) **Does the recommended approach to sampling frequency adequately account for seasonal variability?**

No. Because of the dynamic nature of streams, characterization of both reference and test sites should be based on sampling at various times of the year to include the major components of the fall-winter and spring-summer (or wet season-dry season) communities.

c) **Will the recommended multimetric approach to evaluating aquatic communities adequately characterize the resource?**

Yes. Assuming that the appropriate assemblages and metrics are chosen, we agree that the multimetric approach is an appropriate way to characterize the variety of responses at the individual, population, and community level.

d) **Does the recommended approach to selection and aggregation of biological information provide a sound basis for the development of biocriteria?**

Yes. However, we discourage the use of a single aggregate criterion for any site, since important changes in individual metrics may be masked by aggregation into a single index.

In addition to the questions in the charge, we have the following conclusions and recommendations:

a) The Agency should ensure an adequate, consistent level of taxonomic training among various state biocriteria programs. Within the academic community and state and federal agencies, there exists a wealth of expertise and data on the ecology and systematics of lotic organisms (e.g., there are 200 inland field stations in the U.S. and 15 LTER sites, to name a few). We strongly believe that the primary role of the Agency should not be to recreate this expertise within the EPA, but rather to ensure that this expertise is made available to the states through collaboration with existing centers of expertise in stream ecology, taxonomy and systematics. This collaboration may take the form of support for systematic training in universities and museums, on-site state training workshops, expert networks to verify specimen identifications, etc.

b) Development of Diagnostic Techniques: current state of the science is sufficient to support application of biocriteria for assessment of site-specific impacts and regional trends, but not for establishing point source criteria or permit limits. In order for biocriteria to be more broadly useful, the Agency should support the development of diagnostic tools to differentiate probable causes of observed changes in biological metrics. This may include research on biomarkers and indicator species.

c) The final guidance should strengthen the linkages, both conceptual and procedural, between the Ecorisk Framework and the application of biocriteria.

d) The Agency should refine existing classification schemes to include biogeographic and distributional patterns of stream and riparian biota.

e) The final guidance should include several real case studies of the application of biocriteria on a watershed scale, including cost estimates for development and implementation of the program.

f) Chemical characterization should be required at all sites, both reference and test locations.

g) Reference areas, once designated, should be managed in cooperation with other government agencies to protect the baseline reference condition. One option for communication among states and academic institutions would be an Agency-sponsored Register of Biocriteria Reference Sites.

h) The final guidance should place greater emphasis on statistical design of the bioassessment sampling program, rather than focusing primarily on post-monitoring data analysis. Important issues include identifying the universe to be sampled, defining the level of desired precision, and determining sampling sufficiency.

i) The final guidance should suggest reasonable levels of taxonomic resolution for various assemblages that a) are achievable by appropriately trained personnel, and b) provide an adequate level of information for the intended applications of the biosurvey results.

j) The Agency should strongly encourage states to use the established network of federal, state and university museums for regionally centralized curation of voucher specimens.

9. REFERENCES CITED

Brower, J.E., J. H. Zar, and C. N. von Ende. 1990. Field and laboratory methods for general ecology. Wm. C. Brown, Dubuque, Iowa. 237 p.

Carlson, R. 1977. A trophic state index for lakes. Limnol. and Oceanog. 22:361.

Cummins, K. W., M. A. Wilzbach, D. M. Gates, J. B. Perry, and W. B. Taliaferro. 1989. Shredders and Riparian Vegetation. BioScience 39(1):24-30.

Elliott, J. M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Sci. Publ. No. 25, Freshw. Biol. Assoc., U.K.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. John Wiley and Sons, New York. 257 p.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monographs 54:187-211.

Ludwig, J. A., and J. F. Reynolds. 1988. Statistical Ecology. John Wiley and Sons, New York. 337 p.

McElravy, E. P., G. A. Lamberti, and V. H. Resh. 1989. Year to year variation in the aquatic macroinvertebrate fauna of a Northern California stream. J. North Amer. Bentholog. Soc. 8:51-63.

Omernik, J. M. 1987. Ecoregions of the Coterminous United States. Ann. Assoc. Amer. Geogr. 77(1):118-125.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrates and Fish. EPA/444/4-89-001.

Statzrer, B. and B. Higler. 1986. Stream hydraulics as a major determinant of benthic invertebrate zonation patterns. Freshwater Biol.:16:127-139.

U.S. Environmental Protection Agency. 1990. Biological Criteria: National Program Guidance for Surface Waters. EPA-440/5-90-004.

U.S. Environmental Protection Agency. 1992a. Procedures for Initiating Narrative Biological Criteria. EPA-822-B-92-002.

U.S. Environmental Protection Agency. 1992b. Framework for Ecological Risk Assessment. EPA/630/R-92/001.

Vannote, R. L., G. W. Minshall, K. W. Cummins, J. R. Sedell, and C. E. Cushing. 1980. The river continuum concept. Can. J. Fish. Aquat. Sci. 37:130-137.

Vollenweider, R. A. 1968. Scientific fundamentals of the eutrophication of lakes and flowing waters, with particular reference to nitrogen and phosphorus as factors in eutrophication. Tech. Rpt. to OECD, Paris. DAS/CSI/68:27, 192 p. mimeo.

Waters, W. E. and D. C. Erman. 1990. Research methods: concept and design. In, C. B. Schreck and P. B. Moyle (eds.). Methods for fish biology. Amer. Fish. Soc., Bethesda, MD 684 p.

DISTRIBUTION LIST

Administrator
Deputy Administrator
Assistant Administrators
Deputy Assistant Administrator for Research and Development
Deputy Assistant Administrator for Water
EPA Regional Administrators
EPA Laboratory Directors
EPA Regional Libraries
EPA Laboratory Libraries