DATE TO BE INSERTED

The Honorable Lisa P. Jackson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: SAB Review of Empirical Approaches for Nutrient Criteria Derivation

Dear Administrator Jackson:

EPA's Office of Water (OW) requested that the Science Advisory Board (SAB) review the Agency's draft guidance document titled *Empirical Approaches for Nutrient Criteria Derivation* ("Guidance"). The Guidance is one of a series of technical documents developed by OW to describe approaches and methods for developing numeric criteria for nutrients. The Guidance specifically focuses on empirical approaches for determining stressor-response relationships to derive numeric nutrient criteria. In response to the Agency's advisory request, the Science Advisory Board Ecological Processes and Effects Committee, augmented with additional experts, met on September 9 – 11, 2009 to conduct a peer review of the Guidance. OW requested that the SAB: 1) comment on the technical merit of the methods and approaches described in the Guidance; 2) suggest approaches that might be considered to improve the Guidance; and 3) offer suggestions to improve the utility of the Guidance for state and tribal water quality scientists and resource managers. The enclosed advisory report provides the advice and recommendations of the Committee.

The SAB commends EPA for addressing nutrient issues. Nutrients (nitrogen and phosphorus) are a major cause of water quality impairment in the Nation's waters, and the SAB recognizes the importance of EPA's efforts to develop numeric nutrient criteria. The stressor-response approach is a legitimate, scientifically based method for developing numeric nutrient criteria if the approach is appropriately applied. We encourage the Agency to continue this important work. EPA's draft Guidance provides a primer on a limited set of statistical methods that could be used in deriving nutrient criteria based on stressor-response relationships. However, in its present form, the Guidance does not present a complete or balanced view of using the statistical methods to develop criteria. Improvements in the Guidance are needed prior to implementation to enable development of technically defensible criteria and to make the document more useful to state and tribal water quality scientists and resource managers.

In general, we find that the scope and intended use of the Guidance should be more clearly identified. The empirical stressor-response framework described in the Guidance is one possible approach for deriving numeric nutrient criteria, but the uncertainty associated with estimated stressor-response relationships would be problematic if this approach were used as a "stand alone" method because statistical associations do not prove cause-and-effect. We therefore suggest that the stressor-response approach be considered for use with other available methodologies in the context of a tiered approach where uncertainties in different approaches are recognized, and weight of evidence is used to establish the likelihood of causal relationships between nutrients and their effects for criteria derivation. In this regard, we recommend that EPA more clearly articulate how this particular guidance fits within the Agency's decision-making and regulatory processes and, specifically, how it relates to and complements EPA's other nutrient criteria approaches, technical guidance manuals, and documents. The SAB also recognizes that methods in the Guidance do not address downstream impacts of excess nutrients, and suggests ways in which these could be addressed.

The SAB has provided many recommendations to improve the Guidance and strongly recommends that they be incorporated into the final document. These recommendations focus on revising the document to address: cause-and-effect; the utility and limitations of the statistical methods and approaches in the document; the supporting analyses and data needed to correctly identify predictive relationships; the need for more guidance and examples to describe when and how to use various methods and approaches; linkages among designated uses and stressors; and the need for a more specific and descriptive framework outlining the steps in the criteria development process. Finally, the SAB strongly recommends that EPA invest in providing the technical support and training needed to make the approaches and methods in the final Guidance more useful to state and tribal water resource managers.

Thank you for the opportunity to review this important guidance document. The SAB looks forward to receiving the Agency's response to this advisory report and stands ready to provide additional advice as EPA continues to develop nutrient criteria guidance.

Sincerely,

U.S. Environmental Protection Agency
Science Advisory Board
Ecological Processes and Effects Committee (FY 2009) Augmented for Review of Nutrient Criteria Guidance

CHAIR

Dr. Judith L. Meyer, Distinguished Research Professor Emeritus, University of Georgia, Lopez Island, WA

MEMBERS

Dr. Richelle Allen-King, Professor and Chair, Department of Geology, University at Buffalo, Buffalo, NY

Dr. Ernest F Benfield, Professor of Ecology, Department of Biological Sciences, Virginia Tech, Blacksburg, VA

*** Dr. Ingrid Burke**, Director, Haub School and Ruckelshaus Institute of Environment and Natural Resources, University of Wyoming, Laramie, WY

Dr. G. Allen Burton, Professor and Director, Cooperative Institute for Limnology and Ecosystems Research, School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI

Dr. Peter M. Chapman, Principal and Senior Environmental Scientist, Environmental Sciences Group, Golder Associates Ltd, Burnaby, BC, Canada

Dr. Loveday Conquest, Professor, School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA

Dr. Wayne Landis, Professor and Director, Department of Environmental Toxicology, Institute of Environmental Toxicology, Huxley College of the Environment , Western Washington University, Bellingham, WA

Dr. James Oris, Professor, Department of Zoology, Miami University, Oxford, OH

*** Dr. Charles Rabeni**, Research Professor, Department of Fisheries & Wildlife, University of Missouri, Columbia, MO

Dr. Amanda Rodewald, Associate Professor of Wildlife Ecology, School of Environment and Natural Resources, The Ohio State University, Columbus, OH

Dr. James Sanders, Director and Professor, Skidaway Institute of Oceanography, Savannah, GA

Mr. Timothy Thompson, Senior Environmental Scientist, Science and Engineering for the Environment, LLC, Seattle, WA

*** Dr. Ivor van Heerden**, Director, Center for the Study of Public Health Impacts of Hurricanes, Louisiana State University, Baton Rouge, LA

CONSULTANTS

Dr. Victor Bierman, Senior Scientist, LimnoTech, Oak Ridge, NC

Dr. Elizabeth Boyer, Associate Professor, School of Forest Resources and Assistant Director, Pennsylvania State Institutes of Energy & the Environment, and Director, Pennsylvania Water Resources Research Center, Pennsylvania State University, University Park, PA

Dr. Mark David, Professor, Natural Resources & Environmental Sciences, University of Illinois, Urbana, IL

Dr. Douglas McLaughlin, Principal Research Scientist, National Council for Air and Stream Improvement, Inc., Western Michigan University, Kalamazoo, MI

Dr. Patrick J. Mulholland., Distinguished Research Staff Member, Carbon & Nutrient Biogeochemistry Group, Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN

Dr. Andrew N. Sharpley, Professor, Department of Crop, Soil and Environmental Sciences, Division of Agriculture, University of Arkansas, Fayetteville, AR

SCIENCE ADVISORY BOARD STAFF

Dr. Thomas Armitage, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

* Did not participate in this advisory activity.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY vii

2. INTRODUCTION 1

3. RESPONSE TO CHARGE QUESTIONS 3

3.1. Charge Question 1. Improving the utility of the Guidance 3

3.2. Charge Question 2. Selecting stressor and response variables 9

3.3. Charge Question 3. Approaches to demonstrate the distribution of and relationships among variables 13

3.4. Charge Question 4. Methods for assessing the strength of the cause-effect relationship 18

3.5. Charge Question 5. Statistical methods to analyze the data 20

3.6. Charge Question 6. Evaluating the predictive accuracy of stressor-response relationships 29

3.7. Charge Question 7. Evaluating candidate stressor-response criteria 36

4. REFERENCES 42

1. EXECUTIVE SUMMARY

EPA's Office of Water (OW) requested that the Science Advisory Board (SAB) conduct a peer review of Agency's draft guidance document, *Empirical Approaches for Nutrient Criteria Derivation* (the "Guidance"). The Guidance was developed by OW to provide information for state and tribal water resource managers on empirical stressor-response approaches for developing numeric nutrient criteria. To augment the expertise on the Committee for this advisory activity, several additional scientists with specific knowledge and expertise in assessing the effects of nutrient enrichment in aquatic systems also participated in the review.

EPA's Office of Water develops ambient water quality criteria that serve as guidance to states and tribes for adoption of water quality standards. The water quality standards include designated uses, such as aquatic life protection and recreation, and criteria that define levels of water quality variables protective of the designated uses. Because nutrients (nitrogen and phosphorus) are a major cause of water quality impairment in the Nation's waters, state adoption of numeric nutrient criteria in water quality standards has been a high priority for OW. To assist the states and tribes in developing numeric nutrient criteria, OW published technical guidance manuals for developing nutrient criteria for lakes and reservoirs (U.S. EPA, 2000a), rivers and streams (U.S. EPA 2000b), estuaries and coastal marine waters (U.S. EPA, 2001), and wetlands (U.S. EPA, 2008). These technical guidance manuals focus primarily on describing a reference condition approach for deriving criteria from distributions of nutrient concentrations and biological responses in minimally disturbed reference waterbodies. Other basic analytical approaches for nutrient criteria derivation recognized in the manuals include mechanistic modeling (i.e., predicting the effects of changes in nutrient concentrations using site-specific parameters and equations that represent ecological processes), which EPA intends as the subject of a later document, the stressor-response approach (discussed in the Guidance and considered in this advisory report), and the application and/or modification of established nutrient/algal thresholds. The stressor-response approach involves quantifying the relationship between nutrient concentrations and biological response measures related to the designated use of a waterbody.

The Guidance outlines a five step process for developing numeric nutrient criteria. It describes data analysis methods and approaches that could be used in each of these steps. Step1 involves the use of exploratory analysis and data visualization tools to select variables that appropriately quantify the stressor (i.e., excess nutrients) and the response. Step 2 involves the use of conceptual models, existing literature, and other methods to assess the strength of the relationship represented in the stressor-response linkage. Step 3 involves the use of various statistical methods to analyze data, estimate stressor-response relationships, and identify thresholds that may be used to derive water quality criteria. Step 4 involves the evaluation of estimated stressor-response relationships (including validation of predictive performance for a stressor-response model, and selecting a model that best represents the data). Step 5 involves evaluating candidate nutrient criteria by predicting conditions that might be expected after implementing different criteria. The Guidance contains five sections, each addressing one of the proposed steps in the criteria development process. In its charge to the SAB, EPA requested that the Committee comment on the methods and approaches described in each section of the

Guidance, recommend other approaches that might be considered, and offer suggestions to improve the utility of the Guidance for state and tribal water quality scientists and resource managers. In its responses to the charge questions, the Committee provides comments and recommendations to improve the Guidance and assist EPA in its efforts to support the development of numeric nutrient criteria.

General comments on the Guidance

The Committee recognizes the importance of EPA's efforts to support numeric nutrient criteria development and encourages the Agency to continue this important work. In addition, we recognize the stressor-response approach as a legitimate, scientifically based method for developing numeric nutrient criteria if it is appropriately applied. The draft Guidance provides a primer on a limited set of statistical methods that could be used in deriving numeric nutrient criteria. However, we find that improvements in the Guidance are needed prior to implementation to enable development of technically defensible nutrient criteria and to make the document more useful to state and tribal water quality scientists and resource managers.

In general, we find that the scope, limitations, and intended use of the Guidance should be more clearly described. The Guidance addresses only one type of "empirical" approach for derivation of numeric nutrient criteria (i.e., the stressor-response framework). As illustrated in many of the examples in the Guidance, considerable unexplained variation can be encountered when attempting to use empirical stressor-response approaches to develop nutrient criteria. The final Guidance should clearly indicate that such unexplained variation presents significant problems in the use of this approach. Further, the final document should clearly state that statistical associations may not be biologically relevant and do not prove cause and effect. However, when properly developed, biologically relevant statistical associations can be useful arguments as part of a weight of evidence approach to criteria derivation. Therefore, the final Guidance should provide more information on the supporting analyses needed to improve the basis for conclusions that specific stressor-response associations can predict nutrient responses with an acceptable degree of uncertainty. Such predictive relationships can then be used with mechanistic or other approaches in a tiered weight of evidence assessment including cause-and-effect relationships to develop nutrient criteria. In this regard, we also recommend that EPA more clearly articulate how the Guidance fits within the Agency's decision-making and regulatory processes and, specifically, how it relates to and complements EPA's other nutrient criteria technical guidance manuals and documents. As further discussed in the response to Charge Question 1, numeric nutrient concentration criteria and load-response models should be considered as two different approaches for accomplishing the goal of controlling excessive nutrient loadings. In addition, the Committee notes that the methods in the Guidance do not address the problem of excess nutrient enrichment downstream from waters for which the criteria are being developed. There is a need for methods to address this problem (one of which could be load-response modeling) and it should be clearly stated that this is beyond the scope of the current guidance document.

Charge Question 1. Improving the utility of the Guidance for state and tribal water quality scientists and resource managers

What suggestions do you have that will improve the utility of the draft document, Empirical Approaches for Nutrient Criteria Derivation, for State water quality scientists and resource managers to derive numeric nutrient criteria based on stressor-response relationships?

The Committee finds that improvements in EPA's Guidance are needed to make the document more useful to state and tribal water quality scientists and resource managers and to ensure against inadvertent misuse. In this regard, as previously mentioned, the scope, limitations, and intended use of the document should be more clearly identified.

- In the Guidance, and the Agency's related technical manuals, EPA should more clearly address the importance of: 1) establishing linkages among designated uses and measured responses, stressors and measures of stressors; and 2) relating measures of stressors directly to deleterious effects on designated uses.

- The Committee finds that the Guidance: 1) should provide a more specific and descriptive framework outlining the steps in the criteria development process. Figure 1 below illustrates EPA's proposed framework for developing nutrient criteria and the SAB recommendations for revision of the framework; 2) must be detailed and sophisticated enough to ensure statistical rigor, but additional support must also be provided from EPA to help users meet the technical demands of the methods; 3) should more clearly express the caveats and limitations of the statistical methods and approaches in the document including the fact that statistical correlations do not establish cause-and-effect; 4) should contain more technical guidance and examples to describe when and how to use various methods and approaches; and 5) should provide additional guidance on data requirements for application of the statistical methods and approaches.

- *Charge Question 2. Selecting stressor and response variables*

Section 1 of the draft guidance document reviews how to select the variables that appropriately quantify the stressor (i.e., excess nutrients) and the response (e.g., chlorophyll a, dissolved oxygen, or a biological index). Please comment on whether the factors to consider described in Section 1 of the draft document are appropriate for selecting response variables that are sensitive to nutrients and related to measures of designated uses.

In Section 1 of the Guidance, EPA discusses factors to consider when selecting the stressor and response variables. The Committee finds that EPA should strengthen the Guidance by including additional material.

- The examples in the Guidance rely heavily on taxa richness as a response variable. Some rationale as to how this variable relates to a designated use should accompany these examples. The coupling of response variables to designated uses must be clear and the rationale explained. Further, the Guidance could be strengthened considerably by presentation of examples showing strong nutrient-response relationships with response variables that are clearly linked to designated uses.

- The Committee notes that co-limitation by both nitrogen and phosphorus may be common in many systems and regions. Therefore, the use of multivariate or data stratification approaches may be needed to identify nutrient-response relationships.

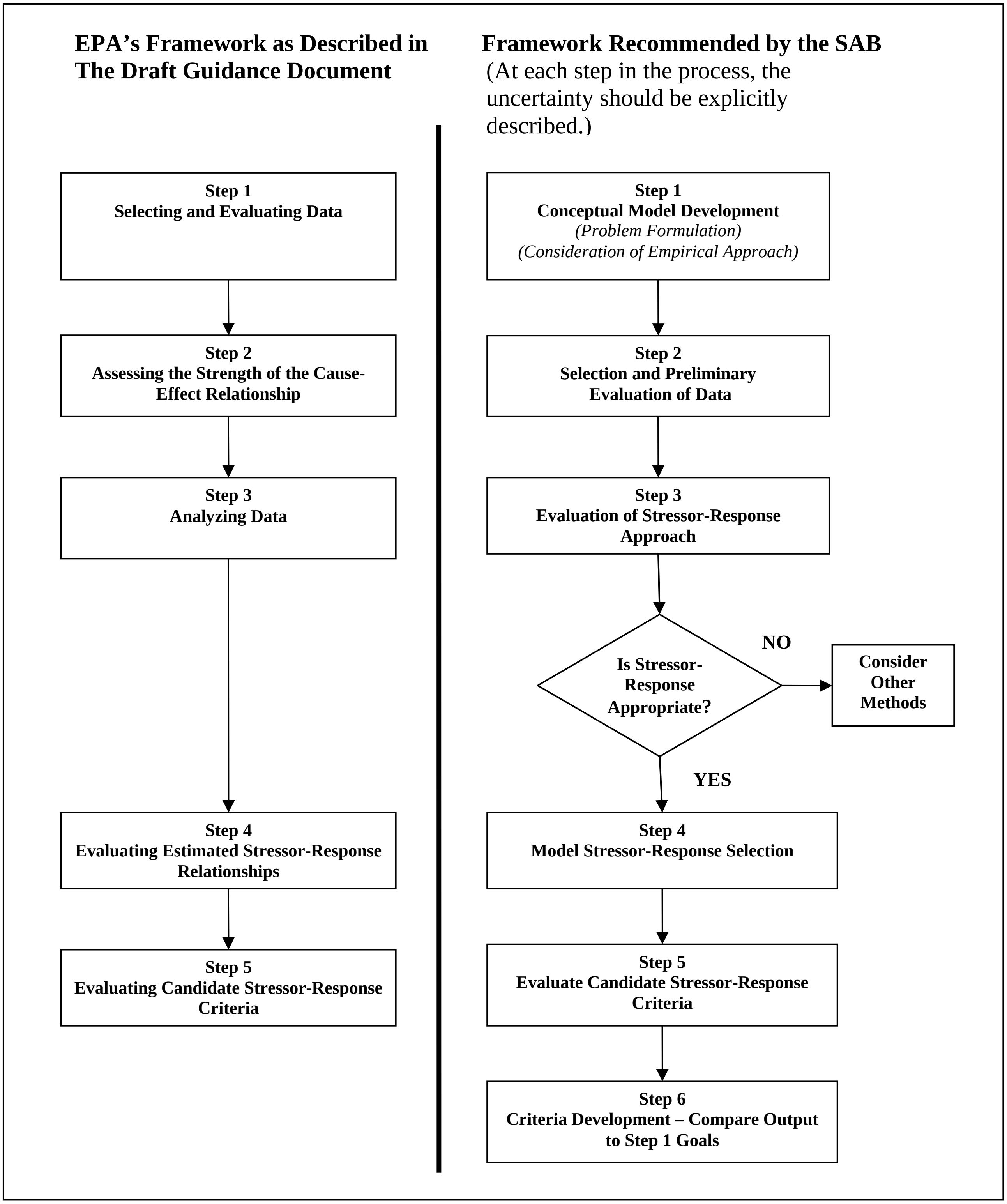

Figure 1. EPA's Framework for Developing Nutrient Criteria Based on Stressor-Response Relationships and SAB Recommendations for Revision

- The Guidance should provide more information on the data needed to characterize other stressor and constraint variables (e.g., high dissolved organic carbon versus low dissolved organic carbon lakes, shaded versus unshaded streams) which are critical for applying multivariate techniques or for stratification/classification of univariate nutrient-response relationships.

- The Guidance focuses on total nitrogen and total phosphorus as the primary nutrient stressor variables. Additional consideration should be given to inorganic forms of these nutrients because these forms are the most immediately biologically available.

- The Guidance focuses on nutrient-response pathways driven by autotrophic processes (nutrients directly control algal growth and excessive amounts of algae impair systems through indirect effects on dissolved oxygen, food web changes, and aesthetics). The Committee notes that nutrients can also directly control heterotrophic microbes (bacteria and fungi) and indirectly control decomposition of organic matter. This should be more fully discussed in the Guidance.

- The Guidance provides inadequate discussion of the temporal/spatial aspects of data needed to develop relevant stressor-response relationships. The Guidance should discuss the conditions under which mean/median or maximum/minimum values of stressor and response variables might be more appropriate than discrete instantaneous measurements for developing stressor-response relationships. The use of time series data to describe specific systems should also be addressed. Although such guidance may be provided in various system-specific technical manuals (e.g., U.S. EPA 2000a, b), a summary synthesis of the major points in these earlier documents should be included in the Guidance.

Charge Question 3. Approaches to demonstrate the distribution of and relationships among variables

Section 1 outlines methods to visualize available data. Please comment on the effectiveness of the following approaches described in the document (listed below) to demonstrate the distribution of and relationships among variables.

a) Basic data visualization techniques
b) Maps
c) Conditional probability
d) Classifications

Section 1 of the Guidance discusses exploratory data analysis, and presents several methods for demonstrating the distribution of and relationships among variables. In Subsections 1.2 – 1.6 several basic plotting techniques are presented. This is followed by a description of conditional probability analysis (a statistical approach for summarizing how changes in nutrient concentrations are associated with the probability of waterbodies attaining their designated uses).

The Committee finds that the discussion of exploratory data analysis would be more effective if Section 1 of the Guidance were reorganized and expanded.

- As further discussed in the response to Charge Question 3, Subsections 1.2 - 1.6 of the Guidance should be reframed as a separate major section on exploratory data analysis, which should follow another separate major section on problem formulation. The material in Subsection 1.1 (selection of stressor-response variables) should be moved to a later section of the document.

- The Guidance should stress that exploratory data analysis, including data visualization, should be conducted prior to inferential statistical analyses of potential stressors and responses. The objectives of exploratory data analysis should be to better understand the system of interest and to maximize the accuracy and minimize the variability of the subsequently derived stressor-response relationships.

- Additional methods for exploratory data analysis should be discussed in the Guidance. These additional methods should include: the use of summary statistics, time series plots at fixed points in space; longitudinal plots at fixed points in time; bubble plots; Pearson and other correlation analyses; and maps that show temporal (monthly, seasonal, inter-annual) as well as spatial patterns.

- Clear guidance is needed on when and how the statistical methods and visualization techniques presented in the document should be used. The strengths and limitations of the methods should also be clearly identified. It would be useful to show several case examples that range from state-wide to local, data-rich to data-poor; and exemplify different types of aquatic ecosystems (e.g., headwater streams, large rivers, lakes and estuaries). Examples should note the strengths, limitations, assumptions and uncertainties that must be considered when using the methods to explore and visualize the data, and subsequently develop criteria. These examples should demonstrate how nutrients can be identified as significant stressors in the presence of multiple stressors and habitat factors that may affect the resident communities.

- Subsection 1.6 of the Guidance (examination of stressor-response distributions across different classes, e.g., ecoregions) should be expanded. The subsection should discuss additional data analysis and examples for different spatial classifications (e.g., ecoregions, states, watersheds, systems of interest), different waterbody types (e.g., streams, rivers, lakes, estuaries) and other important physical and chemical characteristics of systems that could affect the applicability of the nutrient criteria.

Charge Question 4. Methods for assessing the strength of the cause-effect relationship

Section 2 of the draft guidance document describes methods for assessing the strength of the cause-effect relationship represented in the stressor-response linkage. Please comment on whether the draft guidance document adequately describes how conceptual models, existing *literature, and empirical models can be used to assess how changes in nutrient concentration are likely to cause changes in the chosen response variable.*

Section 2 of the Guidance provides a summary of how the strength of candidate stressor-response pairings from step 1 can be assessed. A number of improvements in this section are recommended.

- It is appropriate to use conceptual models and existing literature as the scientific basis to assess how changes in nutrient concentrations might affect response variables. However additional discussion of conceptual model selection, with specific examples, would be helpful. As illustrated in Figure 1 and further discussed on in the response to Charge Question 7, the Committee recommends that development of the conceptual model should occur in the problem formulation step (Figure 1), early in the process of criteria development.

- Structural Equation Modeling (SEM) is discussed in the Guidance as a method for exploring nutrient-ecosystem response. The Committee finds that SEM should be more fully explained and clear examples of its use should be provided.

- The Guidance discusses the use of Propensity Score Analysis (PSA) to estimate stressor-response relationships. PSA seems to be useful for sorting out groups that share covariates but may have unique nutrient characteristics. Such sorting could lead to a clearer understanding of how nutrients function amid multiple covariates. The example of PSA in the Guidance appendix is helpful, but further explanation of how to interpret the results of the analysis is needed. An analysis such as PSA should be discussed in a later section of the Guidance because it is a tool for analyzing data (Section 3 of the Guidance) rather than supporting potential relationships.

- It is not clear why EPA did not include information obtained from mechanistic models in Section 2 of the Guidance. Because mechanistic models can integrate information on the interactions of major ecosystem processes to derive quantitative estimates, they should be discussed as a means to interpret the stressor-response relationship.

Charge Question 5. Statistical methods to analyze the data

Section 3 of the draft guidance document outlines statistical methods to analyze the data to estimate stressor-response relationships. Please comment on the appropriateness of the methods outlined in the document (listed below) for describing stressor-response relationships associated with nutrient pollution. What approaches would you recommend that could effectively address indirect pathways of adverse effects? What recommendations do you have to address the effects of confounding variables and uncertainty in the estimated relationships?

a) Simple linear regression
b) Quantile regression
c) Logistic regression

d) Multiple linear regression
e) Non-parametric changepoint analysis
f) Discontinuous regression models

Section 3 of the Guidance describes a number of statistical methods for analyzing data to estimate stressor-response relationships. The Committee provides comments addressing the appropriateness of the statistical methods for estimating stressor-response relationships.

- The statistical methods described in the Guidance are generally appropriate for estimating stressor-response relationships in support of conceptual models. However, as further discussed in the response to Charge Question 5, more careful consideration of confounding variables is necessary to maximize the potential for stressor-response relationships to reflect cause and effect between nutrient concentrations and ecological responses. The Guidance should be revised to state this more definitively and better assist its users in achieving this objective.

- Those charged with using stressor-response methodology will likely require additional technical support to use the methods in the Guidance. Support should include references (articles, books, and websites), recommended software, and access to technical expertise so that the methodology can be used appropriately and the results interpreted properly.

- EPA should provide guidance regarding what the Agency would consider to be a sufficiently strong stressor-response relationship for criteria development, and how to interpret statistical significance with “low” R^2 values. The Committee also notes that uncertainty must be identified and quantified for all methods and at all stages of the process.

Charge Question 6. Evaluating the predictive accuracy of stressor-response relationships

Section 4 of the draft guidance document describes how to evaluate the predictive accuracy of estimated stressor-response relationships. Please comment on the appropriateness of approaches in Section 4 of the guidance document and factors to consider in evaluating and comparing different estimates of the stressor-response relationships and selecting those most appropriate for criteria derivation.

The Committee provides comments on the appropriateness of approaches discussed in Section 4 of the Guidance and the factors to consider in evaluating and comparing different estimates of stressor-response relationships in order to select those most appropriate for criteria development. Overall, the Committee finds that this section of the Guidance lacks the detail provided in other sections and needs improvement.

- A clear framework for statistical model selection is needed. This framework should include: 1) an assessment of whether analyses indicate that the stressor-response approach is appropriate; 2) selection criteria to establish models of cause/effect and direct/indirect relationships between stressors and responses; 3) consideration of model relevance to known mechanisms and existing conditions; 4) establishment of biological relevance; and 5) ability to predict probability of meeting designated use categories.

- The concept of "validation" as presented in Subsection 4.1 of the Guidance is inconsistent with other EPA guidance (U.S. EPA, 2009b) on development, evaluation, and application of models. Model corroboration (sensu "validation") and uncertainty analysis should both be part of model evaluation and selection. These activities should be directed and informed by pre-established data quality objectives. Additional guidance is also needed on: data set specification and stratification; a suite of validation techniques (e.g., random or non-random held-out data, independent data, resampling/Monte Carlo); and appropriate quantitative levels of goodness-of-fit and uncertainty measures.

- With regard to validation, the Committee suggests that nutrient criteria should result from a tiered weight of evidence approach based on the application of multiple empirical approaches and consideration of multiple response variables as appropriate. The nutrient criteria values that may be determined, after considering validation and uncertainty, may vary significantly from technique to technique or from response variable to response variable. EPA should provide greater guidance on how to assign numeric criteria when a range of responses among analyses/models results in different values.

Charge Question 7. Evaluating candidate stressor-response criteria

Section 5 of the draft guidance document describes how to evaluate the candidate stressor-response criteria. An approach is outlined for predicting conditions that might result after implementing different nutrient criteria. Please comment on uncertainties that would remain if water quality criteria for nutrients were based solely on estimated stressor-response relationships and in what ways other information/analysis would help address and possibly reduce this uncertainty?

Section 5 of the Guidance describes how to evaluate candidate numeric nutrient criteria. The Committee provides comments on uncertainties associated with deriving candidate water quality criteria. We also recommend improvements in the Guidance to help address and reduce uncertainty.

- The Guidance describes approaches that use a data-mining exercise to demonstrate a possible cause-effect relationship for the nutrient-ecosystem response. However, as further discussed in the response to Charge Question 7, the document does not address or partition the inherent critical uncertainties associated with the stressor-response approach. We note that these uncertainties can be extremely large (e.g., several orders of magnitude). To address these uncertainties, the Guidance should better document the physical, chemical and biological variables comprising the morphological relationships (e.g., habitat, spatial, and temporal) that define the aquatic system of interest, and which may be important in modifying the relationship between nutrient concentrations (both nitrogen and phosphorus) and observed endpoints. These factors may dominate the cause-effect pathway and should be documented so that the uncertainty in the relationship between nutrient concentrations and measured endpoints can be reduced.

- The Guidance should indicate that, at the start of the initial problem formulation exercise, a realistic cause-effect conceptual model must be developed, and that the model should include the factors that are likely to contribute most to the change in the response variable for the specific region / system of interest. Then data analyses can be used to evaluate which of the factors, or combination of factors, caused the observed change in the response variable.

- There is considerable uncertainty in linkage of the response variables discussed in the Guidance to the Clean Water Act goals of drinkable, swimmable, and fishable waters. The recommended response variables in the Guidance should be directly linked to these Clean Water Act Goals.

2. INTRODUCTION

EPA's Office of Water (OW) requested that the Science Advisory Board (SAB) conduct a peer review of the Agency's draft guidance document, *Empirical Approaches for Nutrient Criteria Derivation* (the "Guidance"). The Guidance was developed by EPA's Office of Water to provide information for water resource managers on the scientific foundation for using empirical approaches to describe stressor-response relationships for developing numeric nutrient water quality criteria. The SAB Ecological Processes and Effects Committee (Committee) met on September 9th-11th, 2009 to review the Guidance. To augment the expertise on the Committee for this advisory activity, several additional scientists with specific knowledge and expertise in assessing the effects of nutrient enrichment in aquatic systems also participated in the review. This report transmits the advice of the Committee.

EPA's Office of Water is charged with protecting aquatic life, wildlife, and human health from adverse water-mediated effects of anthropogenic pollutants. In support of this mission, OW develops ambient water quality criteria that serve as guidance to states and tribes for adoption of water quality standards. State and tribal water quality standards include designated uses, such as aquatic life protection and recreation, and criteria that define levels of water quality variables protective of the designated uses. Because nutrients (nitrogen and phosphorus) are a major cause of water quality impairment in the Nation's waters, state adoption of numeric nutrient criteria into water quality standards has been a high priority for OW. The Office of Water has stated that numeric nutrient water quality standards are important because they can: support development of nutrient related Total Maximum Daily Loads (TMDLs); provide targets for nutrient trading programs; and make it easier to write NPDES permits, evaluate the success of nutrient runoff minimization programs, and measure environmental progress.

To assist the states and tribes in developing numeric nutrient criteria, OW published peer reviewed technical guidance for developing such criteria for lakes and reservoirs (U.S. EPA, 2000a), rivers and streams (U.S. EPA 2000b), estuaries and coastal marine waters (U.S. EPA, 2001), and wetlands (U.S. EPA, 2008). These technical guidance documents focus primarily on a reference condition approach for deriving nutrient criteria from distributions of nutrient concentrations and biological responses in minimally disturbed reference waterbodies. Other basic analytical approaches for nutrient criteria derivation identified in prior guidance documents include mechanistic modeling (i.e., predicting the effects of changes in nutrient concentrations using site-specific parameters and equations that represent ecological processes), the stressor-response approach, and the application and/or modification of established nutrient/algal thresholds. The stressor-response approach involves quantifying a relationship between nutrient concentrations and biological response measures related to the designated use of a waterbody. In the Guidance, EPA states that, when developing nutrient criteria, the strengths and characteristics of each analytical approach should be carefully considered with respect to data availability and designated use protection needs.

The Guidance outlines a five step process for developing numeric nutrient criteria. Step1 involves selecting variables that appropriately quantify the stressor (i.e., excess nutrients) and the response. The Guidance describes various techniques for exploratory data analysis to understand the properties of different variables and visualize data. These techniques include histograms, box and whisker plots, quantile-quantile plots, cumulative distribution plots, scatter diagrams, and spatial mapping. Step 2 involves assessing the strength of the relationship represented in the stressor-response linkage. The Guidance discusses the use of conceptual models, existing literature, and empirical models to assess the degree to which changes in nutrient concentration are likely to cause changes in a chosen response variable. Step 3 involves analysis of data to estimate stressor-response relationships and identify thresholds that may be used to derive criteria. The Guidance describes a number of statistical methods for analyzing data to estimate stressor-response relationships. These methods include linear regression, logistic regression, quantile regression, non parametric changepoint analysis, and discontinuous regression modeling. Step 4 involves evaluating the stressor-response relationships (including validation of predictive performance for a stressor-response model and selecting a model that best represents the data). Step 5 involves evaluating candidate stressor-response criteria. The Guidance outlines an approach for predicting conditions that might be expected after implementing different nutrient criteria and selecting a value to optimize resource protection. The Committee was asked to comment on the scientific and technical merit of the methods and approaches discussed in the Guidance and to offer suggestions to improve the usefulness of the document to state and tribal water quality scientists and resource managers.

The Committee recognizes the importance of EPA's efforts to support numeric nutrient criteria development and encourages the Agency to continue this important work. In addition, we recognize the stressor-response approach as a legitimate, scientifically based method for developing numeric nutrient criteria if it is appropriately applied. The draft Guidance provides a primer on a limited set of statistical methods that could be used in deriving nutrient criteria based on stressor-response relationships. However, the Committee finds that improvements in the Guidance are needed prior to implementation to enable development of technically defensible criteria and to make the document more useful to state and tribal water quality scientists and resource managers.

In general, we find that the scope, limitations, and intended use of the Guidance need to be more clearly described. The Guidance addresses only one type of "empirical" approach for derivation of numeric nutrient criteria (i.e., the stressor-response framework). In this regard, we strongly recommend that EPA more clearly articulate how the Guidance fits within the decision-making and regulatory processes and, specifically, how it relates to and complements EPA's other nutrient criteria technical guidance manuals and documents. As illustrated in the data analysis examples in the Guidance, a large degree of unexplained variation can be encountered when attempting to use empirical stressor-response approaches to develop nutrient criteria. The final Guidance should clearly indicate that such unexplained variation can present significant problems in the use of this approach. Further, *the final document should clearly state that statistical associations may not be biologically relevant and do not prove cause and effect.* When properly developed, statistical associations can be useful in supporting cause-and-effect arguments as part of a weight of evidence approach to criteria development. Therefore, the final Guidance should provide more information on the supporting analyses needed to improve the basis for conclusions that specific stressor-response associations can predict nutrient responses with an acceptable degree of uncertainty. Such predictive relationships can then be used, with mechanistic or other approaches, in a tiered weight of evidence assessment to develop nutrient criteria. In this regard, we also recommend that EPA more clearly articulate how the Guidance fits within the decision-making and regulatory processes and, specifically, how it relates to and complements EPA's other nutrient criteria technical guidance manuals and documents.

In our responses to the charge questions we have recommended specific revisions to improve various sections of the Guidance before it is published. These recommendations focus on: modifying the framework of the Guidance to make it more specific and descriptive (as illustrated in Figure 1 of this report); providing additional information on conditions under which the stressor-response framework may apply; more clearly expressing the caveats, limitations, and data requirements associated with the approaches presented in the Guidance; providing additional information and examples showing when and how to use methods and approaches described in the document; and providing more detailed and descriptive guidance on the use of statistical methods and additional support from EPA to help users meet the technical demands of the methods.

3. RESPONSE TO CHARGE QUESTIONS

In the responses to each of the charge questions, the Committee has listed key findings and comments as bullets. These findings are followed by the Committee's key recommendations.

3.1. Charge Question 1. Improving the utility of the Guidance

What suggestions do you have that will improve the utility of the draft document, *Empirical Approaches for Nutrient Criteria Derivation,* for State water quality scientists and resource managers to derive numeric nutrient criteria based on stressor-response relationships?

The Committee was asked to offer suggestions to improve the usefulness of the Guidance to state and tribal water quality scientists and resource managers for deriving numeric nutrient criteria based on stressor-response relationships. In this regard, we find that the following improvements in EPA's Guidance are needed. A number of these findings are further discussed in responses to the other charge questions.

Key findings concerning improving the utility of the Guidance

- <u>The scope, limitations, and intended use of the Guidance should be more clearly identified.</u> The Guidance addresses only one possible approach (i.e., the stressor-response framework) for derivation of numeric nutrient criteria. The Guidance should be more useful if it: 1) expanded upon the utility of the mechanistic modeling and reference condition approaches for criteria derivation; 2) more clearly articulated how it relates to EPA's other published nutrient criteria guidance; 3) explained the linkages among designated uses, stressors, measures of stressors, and the deleterious effects of the stressors on designated uses; 4) explained that the Guidance does not address far field "downstream" effects of nutrients; and 5) acknowledged other factors that have appeared to limit state progress toward the developing nutrient criteria (e.g., lack of data, lack of technical expertise, and insufficient resources).

- Revision of the document is needed to facilitate identification of the most scientifically defensible approaches to deriving numeric nutrient criteria. The Committee emphasizes that understanding the causative link between nutrient levels and impairment is necessary in order to assure that managing for particular nutrient levels will lead to desired outcomes. As further discussed below, the stressor-response framework in the Guidance may often not be the most appropriate approach for deriving numeric nutrient criteria.

- Revision of the document is needed to increase its usability while reducing the likelihood of misuse. The Committee finds that the Guidance would be more useful if it: 1) provided a more specific and descriptive framework outlining the steps in the criteria development process (a specific example is illustrated in Figure 1 of this report); 2) contained more technical guidance and examples to describe when and how to use various methods and approaches in the document and ensure statistical rigor (with additional support provided from EPA to help users meet the technical demands of the methods); 3) more clearly expressed the caveats and limitations of the statistical methods and approaches in the document; and 4) provided additional guidance on data requirements for application of the statistical methods and approaches.

- The absence of a direct causative relationship between stressor and response is one of the most serious issues raised by the Committee. Without a mechanistic understanding and a clear causative link between nutrient levels and impairment, there is no assurance that managing for particular nutrient levels will lead to the desired outcome. There are numerous empirical examples where a given nutrient level is associated with a wide range of response values due to the influence of habitat, light levels, grazer populations and other factors. If the numeric criteria are not based upon well-established causative relationships, the scientific basis of the *water quality* standards will be seriously undermined.

- Numeric nutrient concentration criteria and load-response models should be considered as two different approaches for accomplishing the goal of controlling excessive nutrient loadings. EPA has put forth the reference condition approach, the empirical stressor-response approach, and mechanistic modeling as basic analytic approaches for development of numeric nutrient criteria. However, the way in which EPA used results from mechanistic models to develop nutrient load reduction goals for the Gulf of Mexico (Mississippi River/Gulf of Mexico Watershed Nutrient Task force, 2008), and the way in which it is currently using mechanistic models for nutrient and sediment TMDLs for Chesapeake Bay, does not involve development or use of numeric nutrient criteria. The reason is that these mechanistic models (Scavia et al., 2004; Cerco and Noel, 2004) are load-response models, not empirical stressor-response models, and hence they obviate the need for numeric nutrient criteria because they directly link nutrient loads to response variables that represent water quality impairments (e.g., dissolved oxygen, chlorophyll, water clarity and acreage of submerged aquatic vegetation). This reasoning applies not only to mechanistic models but can also apply to empirical models. Turner et al. (2008) and Hagy et al. (2004) developed empirical statistical models for hypoxia in the Gulf of Mexico and Chesapeake Bay, respectively. Both of these models were load-response models and neither involved numeric nutrient concentrations. Further support for this reasoning can be found in Carleton et al. (2005).

Key recommendations concerning identification of the scope, limitations, and intended use of the document

As a consequence of the Committee's discussion and the findings listed above, we provide the following recommendations for revising the Guidance

1. EPA should specify how the Guidance is to be used in combination with other EPA nutrient criteria technical guidance manuals. In the preamble, the Guidance should clearly state that the contents represent one of several possible approaches (i.e., the stressor-response framework in the Guidance, mechanistic modeling, reference condition, and the application and/or modification of established nutrient/algal thresholds) that should be considered when deriving numeric nutrient criteria, and expand upon the utility of considering all approaches in a tiered weight of evidence approach before deciding on a particular course of action. In this regard, the guidance should indicate that <u>numeric nutrient concentration criteria and load-response models should be considered as two different approaches for accomplishing the goal of controlling excessive nutrient loadings.</u> To provide additional information on other approaches, EPA should consider appending to the document relevant portions from earlier guidance manuals.

2. EPA should more clearly articulate how the Guidance fits within the decision-making and regulatory processes and, specifically, how it relates to and complements EPA's nutrient criteria technical guidance manuals and other EPA technical documents. Outlining the fundamental principles that underlie the use of stressor-response relationships and providing background information on water quality impairments (e.g., causes and types of impairments, types of designated uses) might provide a useful context. Including a clearer description of how water use designations influence the derivation of empirically-derived nutrient criteria might be considered as well. Considering the number and usefulness of other EPA-developed processes and recommendations, the authors should consider how they might improve the integration of this document with other EPA efforts. For example, the Guidance would benefit by incorporating the problem formulation stage that is part of the Ecological Risk Assessment process (see Figure 1).

3. In the Guidance EPA should address the importance of: 1) establishing linkages among designated uses, measured responses, stressors, and measures of stressors; and 2) relating measures of responses directly to deleterious effects on designated uses. We agree with the statement in the Florida Department of Environmental Protection's letter of September 4, 2009 indicating that the "most scientifically defensible strategy for

managing nutrients within the range of uncertainty is to verify a biological response prior to taking a management action." This risk/performance-based approach to setting nutrient criteria is evident not only in Florida's program, but also in those developed by California, Maine, and Ohio (Florida Department of Environmental Protection, 2009; Maine Department of Department of Environmental Protection, 2009; McLaughlin and Sutula, 2007). Those risk-based linkages are not addressed in either the Guidance or the Nutrient Criteria Technical Guidance documents for Rivers (2000), Lakes/Reservoirs (2000), and Estuaries (2001).

4. In the Guidance, EPA should emphasize that the document does not address downstream effects of nutrient enrichment, which are intended to be the focus of a separate future document. Load-response models may prove useful in addressing downstream effects. The Committee has some reservations about addressing downstream effects in a separate document because fragmentation of the guidance documents will increase the likelihood that each will be used in isolation and potentially provide misleading results.

5. In the Guidance, EPA should acknowledge key factors that have appeared to limit state progress toward developing nutrient criteria. It is the Committee's understanding that one of the key aims of the Guidance is to accelerate State progress toward adopting numeric nutrient criteria. Because a variety of issues (such as limited availability of data and technical expertise, insufficient resources, and expense) are responsible for slow progress, the Guidance may not sufficiently remedy the underlying problems and therefore not facilitate state numeric nutrient criteria adoption. A more thorough exploration of the underlying reasons for slow progress would help EPA more directly address specific issues that impede progress.

Key recommendations concerning identification of the most scientifically defensible approaches to deriving numeric nutrient criteria

6. In the Guidance, EPA should recommend that users consider alternative conceptual and methodological approaches in cases where such approaches may be needed to account for complex problems associated with nutrients. The problem of eutrophication is complex, involving multiple causal variables, multiple response variables, and feedbacks among the variables (e.g., plants increase in response to nutrients then, in turn, those nutrients are provided a second time as plants decay). Moreover, response variables can be at multiple levels - primary response variables (e.g., plants), secondary response variables (e.g., dissolved oxygen [DO], pH), and tertiary response variables (e.g., fish, macroinvertebrates). A change in a response variable is unlikely to be satisfactorily described by changes in a single "causal" variable (e.g., total nitrogen [TN] or total phosphorus [TP]). The Committee suggests that developing conceptual models/diagrams (more detailed/accurate than shown in Figure 10 of the Guidance) to illustrate linkages and feedbacks between nutrients and response variables would be a useful approach to capture ecological complexity and better construct the conceptual framework.

7. The Guidance EPA should explicitly acknowledge the conditions under which the stressor-response relationship applies. For example, the stressor-response relationship is relatively strong and well-established in lakes and reservoirs as opposed to in streams and rivers where the relationship is more complex and influenced by many factors (e.g., shading, sediment, flow regime). In cases where the relationship is not the most appropriate lens through which the problem should be viewed the user could be directed to other approaches that might better fit the problem. Several other documents referenced above (e.g., Florida, Ohio, and California nutrient guidance documents) provide useful examples *<<Reference is needed for Ohio>>*. The addition of an inset "red-flag" text box that lists circumstances or system characteristics that would alert the user to the need to consider approaches other than stressor-response. This box also might caution the user about mixed systems that have been highly modified and are not easily classified. Likewise, these caveats should also include explicit recognition that the most appropriate criteria may depend upon contexts of the waterbody (e.g., shaded versus open canopy streams), as was done in Florida's guidance document. Searching for a single statewide criterion might obscure important relationships.

8. The Committee suggests that EPA consider the following two key questions as they select variables to develop numeric criteria: 1) which measures will allow detection of impairment of designated uses? and 2) is the relationship sufficiently strong to determine a management or regulatory target (i.e., a criterion) to ensure that the designated use is protected? In certain cases, the most appropriate numeric criterion may not be a particular concentration level of a nutrient. Moreover, the stressor-response framework is but one approach for developing numeric nutrient criteria, and often it may not be the most appropriate. Because this concern cuts across all recommendations and approaches included in the Guidance, and also cuts across all charge questions, it should be seriously addressed.

Key recommendations to increase the usability of the Guidance and reduce the likelihood of misuse

9. EPA should consider modifying the steps that provide the framework of the Guidance. The Committee suggests that the steps in the framework should be more specific and descriptive. An example is provided in Figure 1 of this advisory report. Two important aspects of the example in Figure 1 currently are missing from the Guidance: conceptual model development should be the first step in the process, and the Framework should contain an explicit step to determine whether the stressor-response relationship is appropriate.

10. EPA should revise the Guidance to include more detailed and descriptive information on the use of the statistical methods in the document. In addition, EPA should provide additional support to help users meet the technical demands of the methods. The Committee finds that that the current draft of the Guidance is written for a user with far more expertise than is likely possessed by water managers. This mismatch has two serious potential consequences. First, the Guidance will not be helpful if it cannot be

easily used by state/tribal water scientists, and second, the recommended methods are likely to be misused and/or misapplied if not sufficiently understood by the user. As a corollary, the Guidance could specify the level of expertise needed by potential users. Correctly identifying the level of expertise of the anticipated users and writing the document for them is perhaps the most critical step in the continued development and refinement of the Guidance. As part of this process, EPA needs to outline a relatively straightforward process that the users can follow to employ the methods described and provide technical support for their use.

11. In the Guidance, EPA should more clearly express the caveats and limitations of the approaches presented. In this regard, the following issues are of greatest concern to the Committee: a) The approaches presented in the Guidance are correlative and do not demonstrate causation. b) Many water quality criteria are site-specific and confounding variables likely exist. c) As further discussed in the responses to charge questions 2, 3 and 5, there are limitations associated with the retrospective approaches that are the primary focus of the Guidance, and also shortcomings associated with the multivariate techniques presented in the document. In particular, EPA should better identify potential confounding variables and other latent variables that may affect the response.

12. The Guidance should be revised to include additional information (i.e., technical guidance) and more examples showing when and how to use different approaches presented in the document, the advantages and limitations of each approach, the underlying assumptions and data requirements, appropriate interpretations of statistical results, and how to best parameterize the statistical models. This “how-to” information could take a number of forms, including keys, inset boxes, and appendices. Users must be given additional information that provides a clear understanding of why and under which conditions they should consider any particular approach. Related to this, the Committee recommends that the Guidance contain additional examples of the methods described in the document. Specific topics that might be included in this technical guidance include: how to modify the approaches in order to derive site-specific criteria, how to identify thresholds, use of weight-of-evidence approaches, and how to handle censored values. EPA also could include an appendix that lists other sources of assistance (e.g., Regional Technical Assistance Groups [RTAGs]), statistical packages, and methodological resources). The organization of the document and current section headings also could more clearly identify the steps involved in the suggested empirical approaches. It would also be helpful to incorporate case studies that apply datasets typical of what most states have. These case studies could highlight decision points in the process of criteria derivation. Particularly helpful would be the use of a single case study across all the various approaches suggested in the document.

13. The document should better address data requirements, including data acquisition and data quality. Without providing guidelines on data requirements, the potential for applying techniques to inappropriate or inadequate datasets is great. The Committee recommends casting this discussion in terms of data quality objectives (DQOs), which suggests the following process: 1) state the problem; 2) identify the decision; 3) identify

inputs to the decision; 4) define the study boundaries; 5) develop a decision rule; 6) specify tolerable limits on decision errors; and 7) optimize the design for obtaining data.

3.2. Charge Question 2. Selecting stressor and response variables

Section 1 of the draft guidance document reviews how to select the variables that appropriately quantify the stressor (i.e., excess nutrients) and the response (e.g., chlorophyll a, dissolved oxygen, or a biological index). Please comment on whether the factors to consider described in Section 1 of the draft document are appropriate for selecting response variables that are sensitive to nutrients and related to measures of designated uses.

Section 1 of the EPA Guidance reviews factors to consider when selecting stressor and response variables for empirical derivation of numeric nutrient criteria. The Committee finds that EPA should strengthen the Guidance by including additional material to address the points discussed below. Although the current version of the Guidance addresses some of these points, we recommend including additional examples and revisions to further develop various parts of the text as discussed below.

Findings on selecting response variables

- <u>Although the Guidance states that response variables should be coupled to designated uses, the Committee finds that this point needs additional elaboration.</u> Some response variables described in the Guidance are clearly related to designated uses (e.g., DO) but the linkage of other responses to designated uses is less obvious or not as well supported scientifically (e.g., macroinvertebrate species richness). Despite the importance of DO and the fact that a large number of water bodies are impaired due to low DO concentrations, none of the examples in the draft Guidance includes DO This is a glaring omission.. The Committee notes that appropriate response variables are also highly ecosystem specific. For example, chlorophyll concentrations are often more clearly related to designated uses for lakes than streams. While response variables for single taxa (e.g., salmon) may be tightly related to designated use, multimetric variables (macroinvertebrate indices, index of biotic integrity [IBI]) may be more powerful for integrating the response to nutrients at the community or ecosystem level. The Guidance would be strengthened by including more discussion relating ecosystem type and potential response variables to the designated uses (a table with some accompanying text might be an effective way to do this).

- <u>Conceptual models are an important component in selection of response variables. Any stressor-response relationship used in criteria development must have ecological relevance (based on ecological understanding of the system) that can be readily explained and defended as discussed in step 2 in the Guidance.</u> Conceptual models based on past empirical and experimental studies are important for identifying the mechanisms responsible for responses and effectively communicating this linkage. In the framework suggested by the Committee (Figure 1), developing the conceptual model is the first step in the process.

- The Guidance would be strengthened considerably by presentation of examples illustrating a strong nutrient-response relationship and, as previously mentioned, clear linkage of the response variable to a designated use. It is important to clearly present the rationale for such linkage. Some of the examples in the Guidance illustrate relationships with very low R^2 and response variables that are not clearly related to designated use.

- In the Guidance, further discussion of potential response variables appropriate for nutrient effects on detritus-based systems is warranted (e.g., how macroinvertebrate populations dependent on detritus may respond). The Guidance focuses on nutrient-response relationships driven by autotrophic processes (nutrients directly control algal growth, excessive amounts of which impair systems through indirect effects on DO, food web changes, and aesthetics). However, nutrients can also directly control heterotrophic microbes (bacteria, fungi) and indirectly control decomposition of organic matter. Excessive nutrient levels could produce large microbial growths or alter food webs in detritus-based ecosystems (e.g., many streams). Studies in the literature are cited, but examples using relevant response variables (e.g., shredder macroinvertebrate biomass, leaf breakdown rate) would be useful.

Findings on stressor and related variables

- In the Guidance, more discussion is needed to provide advice and outline the rationale for selecting variables that should be included in data collection in order to allow: 1) classification/stratification of data prior to evaluation of stressor-response relationships (e.g., development of different criteria for different strata of systems); and 2) use of multivariate approaches to separate the influence of nutrients from other stressors (e.g., sediments, light regime, toxics). Stratification/classification is a particularly important issue for defining nutrient stressor-response relationships for streams where other factors can impose significant constraints on the effects of excess nutrients on designated uses. For example, nutrient-chlorophyll relationships may not be observed in highly shaded (forested) streams, but may be significant in open-canopy streams. Similarly, nutrient-chlorophyll relationships may be weak in high gradient streams but much stronger in low-gradient streams. For lakes, nutrient-chlorophyll relationships may be much different for highly-colored (high dissolved organic carbon [DOC]) versus clear (low DOC) systems.

- Single variable stressor-response relationships (e.g., those derived using the simple linear regression approach discussed in the Guidance) that explain a substantial amount of variation are likely to be uncommon for most aquatic ecosystems (in particular, streams). Multivariate approaches (multiple regression, structural equation modeling [SEM], etc.) may be needed to identify nutrient effects. These approaches require data on other potential stressors or constraining variables. Multivariate approaches may also be useful early in the analysis to determine whether nutrient effects are significant relative to other stressors/constraints and whether/how to pursue the nutrient effects using simple univariate regressions, perhaps after stratification of systems.

- The Guidance focuses primarily on TN and TP as the primary nutrient stressor variables. Some consideration should be given to inorganic N and inorganic P because these forms are the most immediately biologically available. In addition, it is easier to measure the inorganic forms of N and P and more and/or better data may be available for these forms. This is particularly true for ammonium and nitrate versus TN, but perhaps less so for P.

- In many regions nitrogen (N) and phosphorus (P) are often co-limiting to plants and microbes and stressor-response relationships based on only one nutrient are weak. Nevertheless, nutrients (N and P) may be the primary factor controlling productivity/biomass. There have been several recent papers arguing for management of N and P in combination rather than singularly (Lewis and Wurtsbaugh, 2008; Conley et al., 2009; Paerl, 2009). This would suggest development of multivariate stressor-response relationships (e.g., multiple regression) that include both N and P as independent variables.

- A basic conceptual problem concerning selection of nutrient concentrations as stressor variables (as illustrated in the Guidance) is that nutrient concentrations directly control only point-in-time, point-in-space kinetics, not peak or standing stock plant biomass. Plant biomass is driven by nutrient supply rates (i.e., nutrient mass loads). Ambient nutrient concentrations are not necessarily good surrogates for nutrient mass loads. Relationships between nutrient mass loads and ambient nutrient concentrations are highly system-specific and depend on many factors including inflows, hydrology, bathymetry, sediment-water exchanges and chemical-biological processes. Consequently, there may be many systems for which nutrient concentrations will not be appropriate stressor variables. For such systems it may be more appropriate, and scientifically defensible, to use site-specific mechanistic models incorporating loading to determine the nutrient controls required to attain designated uses.

Findings on temporal/spatial aspects of data

- The Guidance provides little discussion regarding the temporal/spatial aspects of data needed to develop relevant stressor-response relationships. For example, the document could be strengthened by providing additional material to address the following questions. “Under what conditions might the use of mean/median or maximum/minimum values of stressor and response variables be more appropriate than discrete instantaneous measurements?” “Are there instances when the use of temporally out-of-phase stressor and response data are most appropriate (e.g., the widely recognized relationship between spring nutrient concentration and summer maximum chlorophyll concentration in lakes)?” “How can time series or longitudinal data in specific systems be used to develop more generalized stressor-response relationships?” Although such guidance may be covered in the various system specific technical manuals (USEPA 2000a, 2000b, 2001), a summary/synthesis of the major points of these earlier documents should be included in the empirical approaches document.

- The Guidance could be strengthened by including a discussion of the importance of considering "data bias" in interpreting the stressor-response relationships. This discussion should focus on how "data bias" (i.e., limits on data representativeness) might affect predictive performance and uncertainty in stressor-response relationships. Uncertainty imposed by model assumptions should also be discussed. Specifically, additional guidance is needed with regard to interpretation of data from particular environments (e.g., a set of lake data from a particular region) and its appropriateness (or lack thereof) for describing conditions more broadly. It would be helpful to include in the Guidance examples illustrating databases that would be "ideal" or appropriate for each empirical model presented. For example, information could be provided to indicate whether a conceptual model for considering nutrient criteria might be best approached using: seasonal data; data from shaded versus unshaded streams; data from wadeable streams versus big rivers; and/or long versus short term averages of data describing the stressor or the response.

- It would be useful to include in the Guidance some discussion of how nutrient recycling and other feedbacks influence stressor-response relationships. For example, the Guidance could be strengthened by addressing the following questions. "How does recycling contribute to variability and uncertainty in stressor-response relationships?" "Are there variables that can be used in stressor-response relationships to account for recycling?"

Key recommendations concerning selection of variables to appropriately quantify the stressor and response

The Committee provides the following key recommendations to address the findings above and strengthen Section 1 of the Guidance.

1. The Guidance should be revised to elaborate upon the coupling of response variables to designated uses and the importance of ecological relevance of the stressor-response relationship. Examples should be included to further illustrate this important point. The examples should show strong nutrient-response relationships. The Guidance should be revised to include at least one example for DO. Ideally, each method should include an example for streams/rivers and an example for lakes. If empirical stressor-response relationships are not appropriate or workable for DO in lakes, then the Guidance should state this specifically and recommend other approaches, for example, site-specific mechanistic models. There are a large number of waterbodies that are impaired by low DO and the draft Guidance is silent on this important nutrient-related problem.

2. The Guidance should be revised to include discussion of potential response variables appropriate for assessing nutrient effects on detritus-based systems.

3. The Guidance should be revised to include more discussion and advice concerning selection of variables and data needed to allow:

- Classification/stratification of data prior to evaluation of stressor-response relationships (e.g., development of different criteria for different strata of systems).

- Use of multivariate approaches to separate the influence of nutrients from other stressors (e.g., sediments, light regime, toxics). In general, the importance of multivariate stressor-response relationships and tools for multivariate approaches should be further discussed in the final Guidance.

4. The Guidance should consider including inorganic N and inorganic P as nutrient stressor variables because these forms are the most immediately biologically available.

5. The basic conceptual problem associated with selecting nutrient concentrations as stressor variables should be addressed in the Guidance (i.e., nutrient concentrations directly control only point-in-time, point-in-space kinetics, not peak or standing stock plant biomass).

6. The Guidance should be revised to include discussion of:

   - The temporal/spatial aspects of data needed to develop relevant stressor-response relationships. (e.g., are there instances when the use of temporally out-of-phase stressor and response data are most appropriate).

   - How "data bias" (e.g., data from different types of systems) might affect predictive performance and uncertainty in stressor-response relationships.

   - How nutrient recycling and other feedbacks influence stressor-response relationships.

3.3. Charge Question 3. Approaches to demonstrate the distribution of and relationships among variables

Section 1 outlines methods to visualize available data. Please comment on the effectiveness of the following approaches described in the document (listed below) to demonstrate the distribution of and relationships among variables.

a) Basic data visualization techniques
b) Maps
c) Conditional probability
d) Classifications

Section 1 of EPA's Guidance discusses exploratory data analysis and presents several methods for demonstrating the distribution of and relationships among variables. Several basic plotting techniques are presented in Subsections 1.2 – 1.6 of the document. This is followed by a description of conditional probability analysis, a statistical approach for summarizing how changes in nutrient concentrations are associated with the probability of waterbodies attaining their designated uses. The Committee was asked to comment on the effectiveness of the methods presented in this section of the Guidance.

The Committee notes that the response to Charge Question 3 necessarily overlaps with responses to other charge questions, particularly those that focus on identifying stressor-response relationships and conducting statistical analyses. We emphasize that visualization of data is of secondary importance if the data and statistical methods being visualized are inappropriate, because the visualization in itself suggests authenticity. Furthermore, the exploratory data analysis, including visualization, should be conducted prior to inferential statistical analyses of potential stressors and responses. The objectives of exploratory data analysis should be to better understand the system of interest and to maximize the accuracy and minimize the variability of the subsequent stressor-response relationships. The Committee finds that discussion of exploratory data analysis in the Guidance would be more effective if the document were reorganized and expanded to address the following points.

- The Guidance would be more effective if exploratory data analysis were included by itself in a separate section of the document following a major section on problem formulation (as indicated in Figure 1).

- Additional methods for exploratory data analysis should be described in the Guidance. These additional methods should include: the use of summary statistics; time series plots at fixed points in space; longitudinal plots at fixed points in time; bubble plots; Pearson and other types of non-parametric correlation analyses; and maps that show temporal (monthly, seasonal, inter-annual) as well as spatial patterns.

- Clear guidance is needed for identifying when and how the statistical methods and visualization techniques should be used. The strengths and limitations of the methods should also be identified. It would be useful to show several case examples that range from state-wide to local, data-rich to data-poor, and exemplify different types of aquatic ecosystems (e.g., headwater streams, large rivers, lakes and estuaries). Examples should note the strengths, limitations, assumptions and uncertainties that must be considered when using the methods to explore and visualize the data, and subsequently develop the criteria. These examples should demonstrate how nutrients can be identified as significant stressors when multiple stressors and habitat factors are present that may affect the resident communities.

- The discussion in Subsection 1.6 of the Guidance (examination of stressor-response distributions across different classes, e.g., ecoregions) should be expanded. The subsection should discuss additional data analysis and contain examples for different spatial classifications (e.g., ecoregions, states, watersheds, systems of interest), different waterbody types (e.g., streams, rivers, lakes, estuaries) and other important physical and chemical characteristics that could affect the applicability of the nutrient criteria.

- The examples provided in the Guidance generally do not demonstrate a strong nutrient stressor linkage to beneficial use impairment. The stream examples show very weak correlations that have high levels of uncertainty, and lump data from distinctly different ecosystems where multiple factors in addition to nutrients will contribute to biotic responses.

- All of the statistical and visualization methods discussed in Subsections 1.2 -1.6 of the Guidance can be effective but they should be presented and used in a combined, weight of evidence approach because they each involve exploring the data in different ways.

- The Committee emphasizes the importance of choosing the biological endpoints (i.e., response variables) that respond specifically to nutrients. We note that responses of benthic indices can be related to many types of stress. We question why periphyton would not be a better receptor to measure.

- The Committee suggests that field-based species sensitivity distributions (SSDs) may be useful for nutrient criteria development. We note that SSDs have been used effectively in recent publications for establishing guidelines (or refuting them) for contaminants, temperature and salinity (Hickey, 2008; Leung et al., 2005).

The Committee also notes the following technical edits and corrections needed in the Guidance.

a. Clarify that macroinvertebrate richness is plotted.
b. The Guidance (p. 7) states that "variables are equally weighted" yet only one variable is plotted in each box plot. A better statement would be: "One limitation for boxplots is that all of the samples are equally weighted."
c. Spell out Q-Q plots and CDF when first used.
d. Explain probability survey design and data smoothers or reference.
e. Figure 7 is very confusing to those unfamiliar with scatterplot matrices; some additional explanation regarding how to "read" the horizontal and vertical axes of each graph in the matrix would help. Suggested wording: "For each scatterplot, its x-axis is the variable stated in the column in which the graph appears. Its y-axis is the variable stated in the row in which the graph appears."

Key recommendations regarding methods for demonstrating the distribution of and relationships among variables

As discussed above, the Committee recommends that EPA restructure and revise the Guidance to strengthen the discussion of the methods for demonstrating the distribution of and relationships among variables. The following key recommendations are provided.

1. The Committee recommends that the Guidance be clarified by reframing Subsections 1.2 through 1.6 as a separate major section on exploratory data analysis. These subsections should follow another separate major section on problem formulation (see Figure 1), and the material in Subsection 1.1 (selecting stressor and response variables) should be moved to later section(s) of the document.

2. The Guidance should be revised to include additional methods for exploratory data analysis. These additional methods should include: the use of summary statistics; time series plots at fixed points in space; longitudinal plots at fixed points in time; bubble plots; Pearson and nonparametric correlation analyses; and maps that show temporal (monthly, seasonal, inter-annual) as well as spatial patterns.

3. Subsection 1.6 of the Guidance should be expanded to include additional examples of different spatial classifications. Specifically, the classification subsection of the Guidance (Subsection 1.6) should be expanded with data analysis examples for different spatial classifications (e.g., ecoregions, states, watersheds, systems of interest), different waterbody types (e.g., streams, rivers, lakes, estuaries) and other important characteristics that will affect the applicability of the nutrient criteria. These characteristics could include, but should not be limited to, stream order, flow, velocity, canopy cover, dissolved oxygen, reference condition trophic status, channel width and depth.

4. The Guidance should be revised to clarify, early in the document, that there are many useful statistical and visualization methods that are not presented and which may be useful. The more common/well accepted methods could be listed in a table with references. It may also be useful to mention methods that are inappropriate. With each method the associated strengths, limitations, assumptions and uncertainties should be noted to better guide the user.

5. Several case examples of exploratory data analysis should be included in the Guidance. These examples should illustrate cases ranging from national to local in scope, and data-rich to data-poor, with guidance on how best to explore and visualize the data.

6. The Guidance should contain additional information concerning statistical assumptions associated with various methods. Some guidance should be presented, as in other EPA documents (e.g., U.S. EPA, 2006a; U.S. EPA 2006b), to address the importance of ensuring that statistical assumptions are not violated and that adequately trained statisticians, in concert with experienced aquatic ecologists and environmental modelers, evaluate the data. An example could be included to show how overly simplistic statistical analysis could not identify a relationship that became evident after complex/advanced analysis. The Committee notes that CProb 1.0, EPA's tool for conditional probability analysis was developed with the R language and environment for statistical computing. The Committee questions whether R, an open-source freeware product that is becoming very popular, is completely acceptable, in the sense that there are many R-macros in use that remain to be properly "vetted." There should be some level of assurance that the recommended R-products have been properly vetted (e.g., CProb 1.0).

7. The Guidance should contain a quantitatively based "weight-of-evidence" (WoE) framework using multiple methods and then combining them into figures and tables for visualization. Multiple statistical methods on one dataset do not equate to a reasonable WoE that significantly reduces uncertainty. Rather, the WoE should involve different

assessment methods (e.g., different datasets, different biological endpoints, measures of habitat, etc.). This premise has been embraced by other EPA programs and the scientific community. *<<<Reference(s) needed>>>*

8. The Guidance should contain a discussion of how the stressor/response variables to be used are linked to one another in space and time for further analysis. There is no mention of this in Subsection 1.1 of the Guidance. The Committee questions whether it should be assumed that stressor/response measurements always occur at the exact same time and locations. It is also important to ensure that high flow events have been measured. It is well established that most nutrient loading occurs during high flows. Therefore, the influence of seasonality and smaller-scale temporal dynamics (e.g., storm events) and the importance of linking stressor and response variables with these factors should be at least noted in the Guidance.

9. The Guidance should discuss the use of modeled data (e.g., land use characterization, hydrology, surface runoff, receiving water quality) for estimating nutrient concentrations/exposures. The pros and cons associated with the use of such data should be briefly mentioned. There are a number of EPA-supported models that have been widely used and documented in recent years (e.g., HSPF, QUAL2K, WASP, AQUATOX, Chesapeake Bay WQSTM). Some of these are integrated watershed models designed to represent inflows and non-point source runoff loads. Typically, they are used as a "loading engine" for a receiving water quality model. Receiving water quality models describe load-response relationships for exposures (ambient nutrient concentrations) and effects (e.g., plant biomass, zooplankton, dissolved oxygen), and response parameters that represent use impairment. Some receiving water quality models can address multiple stressors. For example, they can include nitrogen, phosphorus and silicon as potentially limiting nutrients, sediment (suspended solids) and its influence on underwater light attenuation, incident solar radiation, temperature, and grazing pressure. It is possible to use these water quality models to describe exposure (in terms of ambient nutrient concentrations) but in the absence of empirical data, this would not be scientifically defensible.

10. The Committee recommends that EPA re-evaluate many of the figures in the Guidance (e.g., 4-8, 13-16, 21, 25, and 26). These figures show widely varying data that demonstrate weak relationships.

11. The Committee recommends that the Guidance be revised to clearly indicate the statistical assumptions and uncertainties that should be taken into consideration when using methods described in the document. Some of the methods are complex and the descriptions of the methods lack transparency. Guidance should be provided to ensure that States and other users have an understanding of the data requirements and limitations, the associated statistical assumptions, and uncertainties.

12. The Guidance should contain a discussion of how to examine the independent and interactive effects of the variables to be considered in deriving numeric nutrient criteria.

Statistically, there are several well known ways to address additional contributing variables, such as total suspended solids (TSS). One way would be to use a multiple regression model or analysis of covariance (ANCOVA). This is a good approach, as the additional variables are to be treated as continuous variables, and interaction terms could be added to see if the effects of TN/TP were dependent on levels of TSS, which would be expected, particularly for TP. If one treats the additional variables as factors then an analysis of covariance (ANCOVA) model would be most appropriate. For example, if there were a TSS threshold of interest, a relationship could be established between an invertebrate endpoint and nutrient levels above and below a critical TSS threshold. This would allow one to examine independent and interactive effects.

13. The Guidance should mention the potential benefits of using proxy variables in an initial approach for exploratory analysis of data trends. For example, variable data sets that are easier and more practical to obtain, such as more generic point/nonpoint source loadings or commonly sampled stressor/response variables, might be used as proxy variables for exploratory analysis of data trends. This is briefly mentioned in Subsection 3.1 of the Guidance (auxiliary model), but such an approach could also be useful for selecting stressor/response variables early in the process (Section 1).

3.4. Charge Question 4. Methods for assessing the strength of the cause-effect relationship

Section 2 of the draft guidance document describes methods for assessing the strength of the cause-effect relationship represented in the stressor-response linkage. Please comment on whether the draft guidance document adequately describes how conceptual models, existing literature, and empirical models can be used to assess how changes in nutrient concentration are likely to cause changes in the chosen response variable.

Section 2 of the Guidance provides a summary of how the strength of tentative stressor-response pairings from step 1 can be assessed. Certainly, as indicated in the Guidance, conceptual models and existing literature can be used to support relationships that will be explored with the statistical analysis that follows. At this stage of the analysis, stressor-response relationships for which there is no reasonable conceptual model or literature to explain the underlying mechanisms would be of limited value for setting criteria. Such relationships should be set aside. The Committee finds that the Guidance should be improved by incorporating revisions to address the following points.

- <u>Section 2 of the Guidance does not address the strength of the stressor response relationship, but rather support for the stressor-response relationship that is to be explored statistically.</u> “Support” for the stressor response relationship, rather than “strength” of the relationship, would be a better term to use in this section of the Guidance, because strength refers to the “tightness” of the statistical association between stressor and response. Use of the term “support” would therefore be less confusing to the user.

- It is not clear why information from mechanistic models was not included in Section 2 of the Guidance. Because mechanistic models can integrate information on the interactions of major ecosystem processes to derive quantitative estimates of effects, they too should be discussed as a possible way of supporting the stressor-response relationship.

- Additional discussion of conceptual model selection with specific examples would be helpful. There are many ways to select a conceptual model and various model selection criteria that could be applied. An expanded discussion of these issues could help provide further background for a user of the document. Specific examples could be followed in later sections with discussion of statistical approaches to analyze the strength of the potential cause-effect relationships. In other words, EPA could provide an example from beginning to end that a user could follow from step to step.

- One important aspect of finding support for stressor-response pairings is that it is difficult to fully understand the complex relationships that may be identified without formal training and practical experience in the sciences, especially the biological and ecological sciences. The Guidance should state the level of statistical and ecological expertise needed to use the document.

- Structural equation modeling (SEM) and Propensity Score Analysis (PSA) are techniques that can be used to organize and evaluate relationships between nutrients and response variables when extensive data are available. SEM might be more useful in tracing pathways (it is also called path analysis) of cascades that are initiated by excess nutrients than in defining criteria candidates. A relevant example of SEM is really needed in the Guidance if this approach is to be considered by users. PSA, on the other hand, seems to be useful for sorting out groups that share covariates but may have unique nutrient characteristics. Such sorting could lead to a clearer understanding of how nutrients function amid multiple covariates. The example of PSA in the Guidance appendix is helpful, but further explanation of how to interpret the results of the analysis is needed. An analysis such as PSA might really belong in a later section of the document, as it is used for data analysis rather than supporting potential relationships.

- A reasonable way to assess nutrient effects might be to split datasets (through PSA, principal components analysis, and/or cluster analysis) to enable a system-specific analysis (or analysis of a small groups of sites). Given the many factors that affect streams and rivers, system-specific analysis really provides an assessment of whether altering nutrient concentrations would have the desired effect on the biotic communities present. Possible factors to consider in splitting data for streams and rivers might include, for example, stream order, flow, velocity, canopy cover, dissolved oxygen, bottom type, channel width, habitat, and depth .

- Experimental validation of causal relationships between nutrient and response variables should be approached with caution. The final method discussed on page 17 of the Guidance is experimental validation of causal relationships between selected nutrients and response variables. The Committee notes that this approach could be helpful *in situ* (and there are examples of this *<< References for these studies would be helpful>>*, but it is of more limited usefulness with mesocosm or laboratory experiments. For example, Hill and Fanta (2008) and Hill et al. (2009) showed in Oak Ridge National Laboratory artificial streams how P and light interact. This type of work provides fundamental data on how stream algae respond to P and light, and supports basic conceptual models of this relationship. These and previous studies have shown that, under controlled conditions it takes very little P to maximize algal growth given high light. This fundamental relationship could be applied to any stream in the U.S. However, the relationship is often not observed in datasets because other factors such as bottom substrate, turbidity, canopy cover, hydrology, or depth limit algal production. Therefore, caution must be used in applying a relationship from a subset of data to all data that do not have the same or similar conditions.

Specific recommendations concerning methods for assessing the strength of the cause-effect relationship represented in the stressor-response linkage

In light of the comments and findings discussed above, the Committee provides the following key recommendations to improve Section 2 of the guidance.

1. Section 2 of the Guidance would be more appropriately titled "Assessing Support for the Potential Cause-Effect Relationship."

2. Mechanistic models should be discussed in the Guidance as one way of supporting the stressor-response relationship.

3. The discussion of conceptual models should be expanded to address various criteria for model selection, and additional examples should be included.

4. The level of statistical and ecological expertise needed to use the Guidance should be stated.

5. Structural Equation Modeling (SEM), offered as an alternative model for exploring nutrient-ecosystem response, should be more fully explained with clear examples.

6. Further explanation of how to interpret the results of propensity score analysis (and additional examples) should be included in the Guidance.

7. Experimental validation of causal relationships between nutrient and response variables should be approached with caution because a number of factors can affect the response of a system to nutrient enrichment.

3.5. Charge Question 5. Statistical methods to analyze the data

Section 3 of the draft guidance document outlines statistical methods to analyze the data to estimate stressor-response relationships. Please comment on the appropriateness of the methods outlined in the document (listed below) for

describing stressor-response relationships associated with nutrient pollution. What approaches would you recommend that could effectively address indirect pathways of adverse effects? What recommendations do you have to address the effects of confounding variables and uncertainty in the estimated relationships?

a) Simple linear regression
b) Quantile regression
c) Logistic regression
d) Multiple linear regression
e) Non-parametric changepoint analysis
f) Discontinuous regression models

The Committee notes that EPA's draft Guidance appropriately states that numeric nutrient criteria should be based on predictive stressor-response relationships so that changes in the level of stressor variables will result in predictable ecosystem responses. However, based on examples presented in the draft document and elsewhere, a large degree of unexplained variation can be encountered when attempting to use empirical stressor-response approaches to establish criteria. The final Guidance needs to clearly indicate that such unexplained variation can present a significant problem to this method of developing numeric criteria. Further, the final document should emphasize that statistical associations may not be biologically relevant and do not prove cause and effect. However, when properly determined, statistical associations can be very useful in supporting a cause and effect argument as part of a weight of evidence approach to criteria development. To this end, the final document should provide greater detail on the implementation of statistical procedures and development of other supporting information to minimize the degree of unexplained variation and maximize the potential for the empirical stressor-response approach to result in useful numeric nutrient criteria. EPA should also provide guidance on the strength of stressor-response relationships needed to support criteria development using an empirical stressor-response approach. Further, because nutrients are essential elements, the application of statistical methods must consider both nutrient deficiency and excess. Clear links between response variables and designated uses are needed to ensure that both of these possible impairment types are addressed. The Committee provides the following specific findings and comments concerning the appropriateness of statistical methods in the Guidance, approaches to address indirect pathways of adverse effects, and ways to address the effects of confounding variables and uncertainty in the estimated relationships.

Findings on appropriateness of listed statistical methods

- The Guidance represents a substantial step forward in describing statistical methods that can be used in deriving nutrient criteria based on stressor-response relationships, but more information is needed to describe supporting analyses necessary for application of the methods. The six methods identified in the Guidance generally provide appropriate options for describing stressor-response relationships that may be sufficiently predictive to support setting numeric nutrient criteria. As many examples in the draft document illustrate, there is likely to be considerable variability in stressor-response nutrient relationships and thus in the predicted outcome or response to both target setting and response to mitigation efforts.

Therefore, the document must provide more information on the supporting analyses needed for each method to correctly identify useful predictive relationships, and acknowledge that the use of these statistical methods alone cannot provide sufficient evidence of a cause-effect relationship.

- The use of non-parametric change point analysis and discontinuous regression analysis must be associated with biological significance and the designated uses to be protected by numeric nutrient criteria. As stated previously, response variables must be associated with designated uses in all cases. This has implications for the use of non-parametric change point analysis (nCPA) and discontinuous regression in criteria development. The Guidance indicates that, because these procedures may identify breakpoints in nutrient responses that can serve as criteria thresholds, the methods may be used when designated use thresholds are not available. However, although these methods may be able to identify and characterize breakpoints, such breakpoints may not necessarily have any biological significance, nor will they necessarily be related to designated uses that are to be protected by numeric nutrient criteria. Use of these methods must be associated with designated uses.

- The statistical methods in the Guidance require careful consideration of confounding variables before being used as predictive tools. For example, the appropriate use of bivariate regression methods requires additional efforts through classification or other means to minimize the influence of other potential causal variables so that an acceptable level of confidence in the predictive power of the relationship can be achieved. Without such information, nutrient criteria developed using bivariate methods may be highly inaccurate. Multiple linear regression is an appropriate way to incorporate covariates into a single analysis, although predictive power using this procedure must also be evaluated carefully.

- As previously noted, because plant biomass is driven by nutrient supply rates (mass loads), a potential conceptual problem exists with the selection of nutrient concentration (often used in the Guidance) as a stressor variable. This problem illustrates the importance of careful characterization of confounding variables. Nutrient concentrations control only point-in-time, point-in-space kinetic rates, not peak or standing stock plant biomass. Plant biomass is driven by nutrient supply rates (mass loads). Furthermore, nutrient concentrations may not be direct surrogates for nutrient mass loads. Relationships between nutrient mass loads and ambient nutrient concentrations are highly system-specific and depend on many factors. Consequently, in some circumstances, statistical methods alone will not adequately account for the influence of confounding variables and reduce uncertainties. In other words, the Committee anticipates situations in which stressor-response statistical analysis may not lead to a scientifically justified endpoint.

- In order to be scientifically defensible, empirical methods must take into consideration the influence of other variables. On page 22 of the Guidance, the authors acknowledge that factors co-varying with TP concentrations may explain a portion of the 61% of the variation in log chlorophyll a concentrations apparently attributable to log TP concentrations. This presents a critical challenge in the use of empirical methods as a means of establishing numeric nutrient criteria because it means that controlling TP concentrations may have no potential to yield reductions in chlorophyll a concentrations. Thus, in order to be scientifically defensible, empirical methods must take into consideration the influence of other variables.

- It is important to discuss strength-of-relationship concerns and how results of empirical approaches should be interpreted in the context of criteria development. Figure 13 on page 24 of the Guidance provides an illustration of the challenges facing the users of simple linear regression (SLR) and other empirical approaches. In this case, total macroinvertebrate species richness was regressed against total nitrogen concentrations obtained from EPA Environmental Monitoring and Assessment Program (EMAP) West Xeric region streams. Overall, total species richness declines with increasing TN concentration in these stream data. Applying SLR to log-transformed data yields a statistically significant slope -3(log(TN)) at $p<0.001$. However, a large degree of scatter remains, as indicated by the R^2 value of 0.19. A TN "candidate criterion" of 320 ug/L is obtained by finding the point of intersection of an assumed designated use total species richness threshold of 40 and the mean regression line log(TN) = ~ 2.5. Unfortunately, the points where the lower and upper 90% prediction interval lines cross a species richness threshold of 40 cover a TN concentration range from about log (TN) = 1.25 to log(TN) = 4 based on inspection of Figure 13. This corresponds to a TN concentration range of 16 ug/L to 10,000 ug/L. It is important to understand the management consequences of this considerable uncertainty. Also, the fact that the relationship in Figure 13 is both statistically significant (i.e., some trend is evident) and has a low $R^2 = 0.19$ (much scatter also exists) presents an opportunity to discuss strength-of-relationship concerns and how such results should be interpreted in the context of criteria development.

- As previously discussed, relationships for streams may be more complex than for lakes and must account for multiple stressors/conditions and/or stream 'types' or conditions, and then be applied appropriately. For example, a stratified approach that considers attributes known to be important for a particular environment (lake, stream, estuary) such as canopy, habitat, etc., should be considered. It is also important to deal with both N and P simultaneously and to consider inorganic N and dissolved P. An exercise in Section 3 of the Guidance illustrates the relationship between chlorophyll a and total phosphorus in lake water. This is perhaps the easiest and most well known example of stressor-response in natural waters, and specifically in lakes. This relationship is less certain in streams because there is potentially a continuous new supply of nutrients as streams flow over relatively stationary algae and macrophytes. The Guidance also inappropriately assumes that only nutrients affect taxa. The functionality of aquatic food chains is not solely dependent on one type of biota, sediment type, or single nutrient concentration. There are multiple stressors affecting receptors in a number of ways, over the landscape and watershed in question. Confounding variables are not sufficiently addressed in the Guidance. As previously discussed, approaches that address multiple factors, such as a stratified (or hierarchical) approach that considers other attributes known to be important (e.g., canopy, habitat, multiple nutrients) should be considered.

- The Guidance could be improved by replacing many examples that provide low explanatory power. Concerns include examples with very low r^2 indicating low explanatory power and incomplete description of large uncertainty. These examples indicate that variables other than TP or TN have a greater impact on response, which implies that reducing TP or TN may not have the desired effect. Helpful examples could include: one with a response variable indirectly associated with a designated use; and one from a state where a Secchi depth is used as a criterion for water quality (otherwise Subsection 3.1, paragraph 2 sounds extremely vague).

- Parametric (e.g., Pearson) and nonparametric (e.g., Spearman's rank, Kendall's tau) correlation analyses can assist in identifying the influence of confounding variables, but these methods are not specifically mentioned in the Guidance. Both of these types of analyses would be helpful in exploratory data analysis.

- The Guidance lacks sufficient discussion of the importance of variable selection and data characteristics to ensure useful implementation of the statistical procedures. In addition to its incomplete treatment of confounding variables, the Guidance lacks sufficient discussion of the importance of variable selection and data characteristics to ensure useful implementation of the statistical procedures. Many of the nonparametric procedures rely upon bootstrap procedures to obtain confidence intervals. This underscores the importance of using a probability sampling procedure. The implications of different sample sizes should also be more fully discussed. The Guidance states that an advantage of using quantile regression (QR) is that it can provide direct estimates of percentiles of a distribution of Y values at given X values, which may be better estimates of these values than provided by SLR when the assumptions of SLR are not met. Uncertainty associated with estimating extreme quantiles from "small" sample sizes is appropriately identified in the Guidance as a concern for QR. However, small sample size is likely to present considerable challenges to any nutrient criteria development approach, and the Guidance should provide a discussion of how the amount of data may affect the utility of empirical stressor-response approaches.

- In the Guidance, more information must be provided regarding regression assumptions, limitations, and diagnostic procedures. Although the Guidance should not be expected to provide the same level of detail on the implementation of statistical procedures contained in a statistics textbook, more information must be provided regarding regression assumptions, limitations, and diagnostic procedures. The appropriateness of the regression methods will depend on the assumptions and use restrictions of each method. Although the document discusses many of the important assumptions, it would be helpful for this information to be clearly summarized in a table. The table could include headings for each method such as use, inherent assumptions, and specific remarks. In addition, the importance of regression diagnostic procedures should be emphasized. Examples and specific references to additional sources of information should be provided. This could include evaluating data with and without outliers or unusual values.

- More guidance is needed on the interpretation of results from the listed regression procedures. For example, how does one decide whether the results of quantile regression are adequate for criterion development? In the discussion of logistic regression, (p. 28, last paragraph), nothing is said about whether the coefficients in this analysis are significantly different from zero, or about the proportion of total deviance accounted for by the regression. For multiple linear regression (p. 31) a reference (e.g., Kutner et al. 2004) is needed for Akaike and the other methods listed in that sentence and the sentence following.

- The role of, and options for, data transformations should receive considerably more discussion in the Guidance. Data transformation may be appropriate in the development of stressor-response relationships using regression analysis, but this topic (including the associated back-transformation of slope estimates and confidence intervals to yield criteria) should be more carefully developed. In reading the document, one wonders when the log-transformation should be used to establish linear relationships or whether curvature that may be present in raw data (with no transformation) should be characterized. In addition, the document does not describe the range of data transformations that may be appropriate, instead focusing only on the log-transformation. For example, regarding the nCPA presented in Figure 24, would the analysis give the same result if it were based on TP data that were NOT log transformed? It is not clear in the Guidance when to apply a linear method to transformed data or a changepoint or discontinuous regression method to untransformed data. As a start, a table like Table 6.5, "Linearizing Transformations" in Weisberg (1985, p. 142) could be included in the Guidance, along with some explanation. Finally, "back-transformation" has the potential to introduce bias into the criterion value if done incorrectly, and this topic should be treated more completely to minimize that potential.

- The Guidance appropriately points out that regression relationships should generally not be used to project conditions beyond the range of conditions used to develop the relationships.

- The Guidance is silent on how and when the results of multiple statistical procedures may be integrated to support numeric criteria as an alternative to selecting "the best" model in situations where a clearly preferred model does not emerge from the analysis. Rather than presenting the statistical techniques strictly as alternatives, the document could describe how these procedures can complement each other and provide a more robust picture of what an appropriate criterion should be. For example, a linear regression whose residuals appear to show the presence of curvature might also be evaluated with nCPA to evaluate the range of stressor values over which the curved response occurs. Model averaging (Burnham and Anderson, 2002) is recommended for use with multiple regression when slight changes in the data lead to different final models.

- The Guidance provides a limited list of the statistical methods that could be explored to yield useful criteria. If a data set includes censored values, maximum likelihood estimation can provide an alternative to bivariate or multivariate linear regression that avoids the need to substitute values such as one-half the detection limit for nondetects. In addition, parametric multivariate methods including principal components analysis (PCA), discriminant function analysis, cluster analysis, and others may also provide a useful means of incorporating covariates in a stressor-response relationship. PCA may be used to describe a group of correlated variables through a single equation. A number of nonparametric linear regression approaches are also available, including the family of Kendall tests available from the U.S. Geological Survey (Helsel and Hirsch, 1992; Helsel et al., 2006)

- Users of the Guidance should be provided direction concerning the selection of good software packages for statistical analysis. This information could be provided in a table or appendix. The Committee notes that SAS is well accepted software. A URL is given on page 69 of the Guidance for the Excel add-in to CPA. In the Guidance, it would be helpful to actually show Figures 1-5 (dialogue boxes, demonstration data, and resulting CPA output) from the CProb 1.0 installation document referenced in Appendix C. Other available software includes the EPA ProUCL 4.0 software and "Tools for Trends" software (based on Helsel and Hirsch, 1992) from USGS for analysis of trends. The free and changeable software R is useful for methods that appear in R's established libraries. The Committee notes that any specialized macros need to be properly vetted.

- A key and an associated appendix of case studies should be included in the Guidance to explain the appropriate use of statistical methods and inherent assumptions and uncertainties. Since choice of method(s) will depend on the nature of the data being modeled and on the underlying assumptions, it would be useful to include in the Guidance some kind of key giving an explanation of "which method to use when", with the inherent required assumptions and uncertainties associated with each method. Better use of case studies (from lakes, streams, estuaries) in an appendix could help show "why this approach works in this situation and why that one does not." One case study should estimate the stressor-response relationship when the data form a "wedge-shaped" scatterplot, a pattern commonly observed in nutrient stressor-response relationships.

- Statistical rigor is essential to the development of scientifically defensible criteria. Simplistic application of approaches in the Guidance can lead to stressor-response relationships with poor predictive power and result in inappropriate numeric nutrient criteria. Therefore, EPA will need to provide support to states in the training and use of these statistical methods. As previously stated, the use of bivariate methods (including nCPA) must involve a careful examination of potentially confounding variables to develop support for a predictive relationship. In order to properly evaluate the predictive power of empirical stressor-response relationships, uncertainties associated with each method used must be identified and quantified. Simulated data sets designed to contain specific properties that may be encountered by users of this Guidance document could help to communicate how these statistical procedures behave over a variety of data set characteristics (e.g., a range of uncertainty in the regression slope).

- The need for statistical rigor applies to both the strength and the form of the relationship among variables (i.e., evaluating the presence of curvature in a stressor-response relationship). The Guidance should describe the goal of data analysis as one of characterizing not only the strength of relationship but also its form, and the evidence supporting conclusions about both. This is particularly relevant when deciding to use nCPA or discontinuous regression to characterize a relationship. A more complete approach should be presented to test the hypothesis that a true data threshold exists.

- Perhaps most importantly, EPA should provide guidance on the degree of relationship (indicated by r^2, residuals analysis, and other evidence) needed to establish sufficiently predictive stressor-response relationships. At a minimum, EPA should describe how to address the important question of "when is the evidence insufficient to support using a empirical stressor-response approach?" One suggestion is to better incorporate the EPA data quality objectives process into the Guidance (see U.S. EPA, 2009c).

Findings on indirect pathways

- The Committee notes that, with respect to approaches used to address indirect pathways of adverse effects, the Guidance currently does not contain a clear definition of the term "indirect pathway". One definition follows in part from the caption of Figure 10 in the Guidance:

> "Simplified diagram illustrating the causal pathway between nutrients and aquatic life use impacts. Nutrients enrich both plant/algal as well as microbial assemblages, which lead to changes in the physical/chemical habitat and food quality of streams. These effects directly impact the insect and fish assemblages. The effects of nutrients are influenced by a number of other confounding factors as well, such as light, flow, and temperature."

This description appropriately indicates that nutrient concentrations directly impact plant/algal and microbial communities and indirectly impact insect and fish assemblages through impacts on plant/algal and microbial communities. As discussed previously, a challenge in using empirical approaches is establishing sufficient evidence to support conclusions of cause and effect so that relationships with adequate predictive power can be developed. The farther removed the response variables are from immediate responses of variations in nutrient concentrations, the more difficult it may be to demonstrate a useful degree of predictive power. Guidance on the acceptable degree of uncertainty, and/or the desired level of predictive power, may help users of the Guidance identify useful relationships whether or not pathways are direct or indirect. On the other hand, empirical methods alone are unlikely to effectively address indirect pathways of adverse effects. This requires appropriate conceptual and mechanistic models, adequate site-specific data, and experienced professional judgment.

Findings on confounding variables and uncertainty

- As previously discussed, exploratory data analysis that includes classification of data by similarities in confounding variables prior to the evaluation of stressor-response relationships may improve the predictive power of the relationships if sufficient data are available. Incorporation of confounding variables in a multiple regression is also appropriate.

- Because uncertainty in the appropriate criterion value cannot be eliminated, it is prudent to evaluate the potential consequences of varying degrees of uncertainty in a stressor-response relationship on the resulting criteria and management objectives. This may be accomplished in part through the use of the EPA data quality objectives (DQO) process or a similar approach.

- References should be provided to direct the reader to more information on regression diagnostics including leverage statistics and information on influential points. This would assist the user in addressing uncertainties associated with these values. (One useful textbook is Kutner et al., 2004; there are many others.)

- The Guidance should emphasize the importance of careful pairing of potential stressor and response variables. Uncertainty in a stressor-response relationship may be increased if incompatible data types are paired. For example, combining a seasonal average chlorophyll a concentration calculated from multiple samples with a total phosphorus concentration obtained from a single grab sample could introduce considerably more uncertainty than if both variables represent seasonal averages. There are places in the guidance document where measured values are presented without a clear description of the spatial or temporal components that the value represents (on p. 22, for example, 15 ug/L chlorophyll a is presented as a threshold between mesotrophic and eutrophic conditions without indicating the applicable averaging period). The Guidance should consistently include such information in its descriptions of various components of the threshold identification and criteria-setting process.

Key recommendations concerning statistical methods in the Guidance

The Committee provides the following key recommendations to address the comments and findings presented above.

1. In the Guidance, EPA must provide more information on the supporting analyses needed for each statistical method to correctly identify useful predictive relationships, and acknowledge that the use of these statistical methods alone cannot provide sufficient evidence of a cause-effect relationship.

2. The Guidance should indicate that response variables must in all cases have biological relevance and be associated with designated uses.

3. The Guidance should emphasize that use of the statistical methods requires careful consideration of confounding variables before the methods can be used as predictive tools. As discussed above, further information on how to address confounding variables should be included in the document.

4. The Guidance should contain additional discussion of the potential consequences of varying degrees of uncertainty in a stressor-response relationship on the resulting criteria

and management objectives. This may be accomplished in part through the use of the EPA DQO process or a similar approach.

5. The Guidance should contain more information on approaches that address multiple factors, such as a stratified (or hierarchical) approach that considers other attributes known to be important such as canopy, habitat, multiple nutrients, etc.

6. EPA should consider replacing the examples in the Guidance that provide low explanatory power.

7. As discussed above, the Guidance should contain additional specific information (or guidance on where to find it) on:

   - The use of parametric (e.g., Pearson) and nonparametric (e.g., Spearman's rank, Kendall's tau) correlation analyses.
   - The importance of variable selection (including careful pairing of stressor and response variables) and data characteristics to ensure useful implementation of the statistical procedures.
   - Regression assumptions, limitations, and diagnostic procedures.
   - Interpretation of results from the listed regression procedures.
   - The role of, and options for, data transformations.
   - How and when the results of multiple statistical procedures may be integrated to support numeric criteria.
   - The selection of good software packages for statistical analysis.
   - An appendix of case studies to explain the appropriate use of statistical methods and inherent assumptions and uncertainties.

8. The Committee recommends that EPA consider providing technical support and training to states and tribes to assist them in the use of the statistical methods in the Guidance.

9. The Guidance should describe the goal of data analysis as one of characterizing not only the strength of relationship but also its form, and the evidence supporting conclusions about both.

10. The Committee emphasizes that EPA should provide guidance on the degree of relationship (indicated by r^2, residuals analysis, and other evidence) needed to establish sufficiently predictive stressor-response relationships for numeric nutrient criteria development.

3.6. Charge Question 6. Evaluating the predictive accuracy of stressor-response relationships

Section 4 of the draft guidance document describes how to evaluate the predictive accuracy of estimated stressor-response relationships. Please comment on the

appropriateness of approaches in Section 4 of the guidance document and factors to consider in evaluating and comparing different estimates of the stressor-response relationships and selecting those most appropriate for criteria derivation.

Overall, the Committee notes that Section 4 of the Guidance lacks the detail provided in other sections and, as discussed below, needs improvement. The Committee finds that this section is particularly important because it addresses the reliability or "validity" of the approaches considered. The Guidance should provide information to help managers decide which criteria derivation approach to use (e.g., analysis of best fit by regression or some other means). These are important decisions and additional guidance on how to select the best tools would be helpful. If the proposed methods yield inaccurate results, this could lead to inappropriate or ineffectual solutions to comply with Clean Water Act goals. The Committee provides the following specific findings and comments in response to Charge Question 6.

- <u>The Committee finds that a clear framework and criteria for statistical model selection is needed in the Guidance.</u> This framework should include a set of decision tools and criteria used not only to determine which model fits best, but also to decide whether the stressor-response approach to criteria development is appropriate. Model selection criteria should include:
  - Capability of model to consider cause/effect and direct/indirect relationships between stressor-response;
  - Biological relevance;
  - Relevance to known mechanisms and existing conditions; and
  - Capability of model to predict probability of meeting designated use categories.

Findings on model validation

- <u>More detail is needed in Subsection 4.1 of the Guidance to describe model validation techniques.</u> In the Guidance there is limited discussion of validation of empirically derived stressor-response relationships. This is a critical component. Validation can be defined as demonstrating the accuracy of the model for a specified use. Within this context, accuracy is the absence of systematic and random error - in ecology they are commonly known as trueness and precision, respectively. All models are, by their nature, incomplete representations of the system they are intended to model but, in spite of this limitation, models can be useful. Many discussions of mathematical modeling discriminate between model confirmation (i.e., plausible, worthy of belief) and model verification (i.e., shown to be true). Given the nature of the environmental stressor and response data, such stressor-response models cannot be fully validated. EPA should provide much more detailed validation guidance, including four components:

  - *Conceptual validation* concerns the question of whether the model accurately represents the environmental system. This is largely qualitative and requires consideration of the strength of the cause/effect relationships. To consider whether the empirical model assumptions are credible, a conceptual model of factors affecting the stressor-response relationship should be developed. For each of the proposed methods, guidance should be provided with examples showing the mechanistic reasoning behind the cause/effect assumptions and the direct/indirect responses of the stressor and response variables. This should be supported by some experimental evidence relevant to the context in which it is used (e.g., data needs appropriate for lakes may be different than for streams). For each application of the empirical model, experimental or observational data in support of the principles and assumptions should be presented and discussed.

- *Algorithm validation* concerns the translation of model concepts into mathematical formulae. It addresses questions such as: "Do the equations represent the conceptual model?" "Under which conditions can simplifying assumptions be justified?" "Is there agreement among the results from use of different methods (e.g., different response variables) to solve the model?" For ecological stressor-response models, these questions relate to the adequacy of the empirical models themselves for describing the effects of nutrient enrichment on aquatic life.

- *Functional validation* concerns checking the model against independently obtained observations. For this type of validation the Guidance recommends using additional empirical observations (an alternative experimental dataset). However, this requires more information than is usually available, and results may not be expected to be the same from one dataset to another, given the heterogeneity of environmental systems. Such data cannot truly validate the stressor-response model per se, but may produce valuable insights. Guidance is needed to answer questions such as: "what are the minimum data requirements for validation?" and "if one is working with a limited dataset, how does one consider the tradeoffs between using more data in the original analysis and reserving data for validation?"

- *Software validation* concerns the implementation of mathematical formulae in various computer software. This validation takes into consideration the possible effects of software-specific factors on the model output (e.g., with regard to precision). For example, problems have been documented with regard to performing statistical analyses with some spreadsheet programs or open source codes. Choices of appropriate software approved by federal agencies for such purposes (e.g., USGS and EPA software packages) should be explored and recommended.

- <u>The Committee finds that the concept of “validation” as presented in Subsection 4.1 of the Guidance is inconsistent with other EPA guidance (U.S. EPA, 2009a) on development, evaluation, and application of models.</u> In EPA’s other modeling guidance, model evaluation includes model corroboration, and sensitivity and uncertainty analyses. Model corroboration is defined as quantitative and qualitative methods for evaluating the degree to which a model corresponds to reality. In practical terms, this is the process of “confronting models with data.” In some disciplines, this process has been referred to as validation. EPA prefers the term “corroboration” because it implies a claim of usefulness and not truth. The Committee finds that this is not just a semantic distinction and we recommend that Subsection 4.1 of the Guidance be revised so that it is consistent with U.S. EPA (2009a).

- The use of data quality objectives (DQOs) should be discussed in Subsection 4.1 of the Guidance. The DQOs should be established at the beginning of the criteria development process (i.e., Guidance step 1) but they can also be used to evaluate the potential stressor-response models (Guidance step 4). The discussion of DQOs should address levels of uncertainty, Type I and Type II error rates, and the extent to which each model can predict the probability of meeting designated use categories .

- In Subsection 4.1, more detailed guidance should be provided on the use of randomly or non-randomly selected data sets to help address questions about how much data should be held out of the original analysis to adequately support the validation process. Subsection 4.1 is intended to describe how to validate "the predictive performance of different models." Recommended approaches include: a) collecting new samples; and b) holding out a subset of the original data from the analysis. Reserved samples may be selected randomly or non-randomly. The authors of the Guidance appropriately note that a potential problem with using random subsetting is that the covariance structure of the data is likely to be the same, so that this approach may not provide an independent test of the predictive power of a relationship. As stated in the Guidance, reserving a non-random subset may be a useful alternative. Some discussion of the relative size of calibration and validation data sets is warranted.

- The concept of "best fit" needs elaboration in the Guidance. Best fit is based on the assumptions made and the model developed and, as previously discussed, there may be considerable uncertainty even if a model is well and carefully developed. Assumptions that are incorrect or incomplete will lead to erroneous criteria. The authors of the Guidance understand this, and state that relationships can be confounded by unsampled or unmodeled factors. This statement is true and it should be more fully discussed, and perhaps given much greater weight in each section. EPA should consider whether each example in the Guidance should be accompanied by a discussion of possible confounding issues and what might be missing. The concept of uncertainty, its effect on model results, and ways to at least understand the level of uncertainty are not fully described in the Guidance.

- The Guidance should contain additional information to assess the closeness of root-mean-square predictive error (RMSPE). The RMSPE as presented on p. 42 of the Guidance is a well-recognized measure of how well a statistical model does in predicting response values from given stressor values. Figure 27 of the Guidance gives an example where the RMSPE for the calibration dataset was 0.28, while the RMSPE for the held-out validation data (from a particular State) was 0.27. Many would agree that those two RMSPEs are "close". But it is necessary to answer the question, "how close is close?" No further statements appear in the Guidance about how to assess the closeness of two RMSPEs. Comparing 0.28 with 0.27 in a single example does not help users of the Guidance extend this example to their own data sets. It might be possible to take a bootstrap approach with regard to the calibration data set, to derive an actual distribution of values for the calibration RMSPE, against which the RMSPE of the validation data set could be compared. The Guidance does not address this. In addition, it is appropriate to characterize fit quality using other information such as r^2, residuals analysis, regression results, etc.

- With regard to validation, nutrient criteria should result from WoE from the application of multiple empirical approaches considering multiple response variables and other approaches as appropriate. The nutrient criteria values determined, after considering validation and uncertainty, may vary significantly from technique to technique or from response variable to response variable. The Committee suggests that EPA consider the range of responses and concordance among analyses/models and, as stated previously, establish linkage between response variables and designated use categories. The Guidance should discuss model averaging and should recommend considering the range of responses as a measure of overall utility of the empirical approach. In addition, the Guidance should more strongly advocate decision making based on WoE from multiple empirical and other approaches.

Findings on qualitative assessment of the uncertainty of the estimated stressor-response relationship

- The Committee finds that Guidance Subsection 4.2 (addressing uncertainty) is too brief. Given the importance of this cross-cutting issue, a section on uncertainty is needed for each of the steps outlined in the Guidance, and uncertainty should be summarized at the end of the document.

- Subsection 4.2 of the Guidance should address both qualitative and quantitative estimates of uncertainty. Given reasonable expectations for data availability and inevitable limits on the conceptual understanding of complex environmental systems, the Guidance should discuss both qualitative and quantitative estimates of uncertainties. The Committee notes that an explicit accounting of uncertainty is critical.

- Validity of the space-for-time substitution assumption can be supported by analysis of long-term stressor-response data for selected, data-rich sites. Subsection 4.2 of the Guidance states that all stressor-response models estimated from cross-sectional or synoptic data must also invoke the assumption that spatial differences in sites can be substituted for temporal differences without a substantial degradation of model accuracy (i.e., the space-for-time substitution). As the Guidance states, a good way to provide support for the validity of this assumption is to analyze long-term stressor-response data for selected data-rich sites.

- As previously discussed, the Guidance should contain additional information about the importance of considering "data bias" in interpreting the stressor-response results with regard to predictive performance and uncertainty, and also the importance of uncertainty imposed by model assumptions. Additional guidance is needed on to how to interpret data from a particular environment (e.g., a data set based on lake data) and its appropriateness (or lack thereof) for describing conditions more broadly. It would be helpful to include in the

Guidance examples of databases that would be "ideal" or appropriate for each empirical model presented (e.g., would the conceptual model for considering nutrient criteria be ideally approached using seasonal data, data from shaded versus unshaded tributaries, data from wadeable streams versus big rivers, and/or long versus short term averages of data describing the stressor or the response?)

Findings on selection of the stressor-response model

- The Committee notes that Subsection 4.3 of the Guidance should discuss grounding models in reality through use of prior knowledge. A great deal is known about the effects of nutrients on aquatic systems, and the relationships between variables should reflect that knowledge. All models should be evaluated to determine whether they make sense biologically (e.g., is the range of data used appropriate?; are they mechanistically sound?).

- Subsection 4.3 of the Guidance could be improved by providing a more detailed discussion of how to decide when to use each method to model stressor-response relationships, and the advantages/disadvantages associated with each method. Table 1 on page 44 of the Guidance is not sufficient for this purpose. It would be beneficial to provide a case study using a single data set to demonstrate the comparison of a range of model choices.

- The Committee notes that the stated objective of Subsection 4.3 in the Guidance, "demonstrating how to select a stressor-response model using the response variable that best represents the data," is not the same as the goal of Section 4, "evaluating the predictive accuracy of estimated stressor-response relationships." Confidence in predictive accuracy should be the primary consideration in model selection. Further, while it may ultimately be necessary to select a single model, one should also understand the significance to criteria derivation of selecting among reasonable alternative models or the effect of model averaging when a single most appropriate model cannot clearly be identified.

- In Subsection 4.3 of the Guidance, more detail should be provided in the discussion of conditions under which the last two methods, non-parametric changepoint analysis (nCPA) and discontinuous regression, should be applied (other than simply stating that they should be used when a direct designated use impairment threshold is unavailable). In addition, the Committee notes that a curved response: 1) may or may not be real; 2) may or may not signal an impaired designated use; and 3) may or may not be indicated at all by the data. Further, a curved response may be modeled by one of the linear methods after transformation.

- The Committee notes that linear stressor-response functions may not provide high levels of accuracy for nutrient criteria development. Six different methods are summarized in Table 1 of Subsection 4.3. The first four methods all assume that the stressor-response function can be modeled sufficiently as a linear model or a generalized linear model. It is unlikely that linear stressor-response functions can ever achieve high levels of accuracy across the many different confounding variables and the many different physical, chemical and biological characteristics of specific sites.

Key recommendations concerning evaluating the predictive accuracy of estimated stressor-response relationships

As a consequence of the findings presented above, the Committee provides the following key recommendations.

1. The Guidance should be revised to provide a clear framework for statistical model selection. This framework should include a set of decision tools and criteria used not only to determine which model fits best, but also whether the stressor-response approach to criteria development is appropriate.

2. The Guidance should be revised to provide much more detailed model validation guidance.

3. Subsection 4.1 of the Guidance (Model validation) should be revised to:

   – Make it consistent with other EPA guidance (U.S. EPA, 2009a) on development, evaluation, and application of models.
   – Provide more detailed information on the use of randomly or non-randomly selected data sets to help address questions about how much data should be held out of the original analysis to adequately support the validation process.

   – Elaborate upon assumptions and uncertainties in "best fit" determinations, and in particular provide additional information to assess the closeness of root-mean-square predictive error (RMSPE).

   – State that nutrient criteria should result from a weight of evidence argument based on the application of multiple empirical approaches considering multiple response variables as appropriate.

4. Subsection 4.2 of the Guidance should be revised to provide an expanded discussion of uncertainty. This section should address both qualitative and quantitative estimates of uncertainty as well as data bias.

5. Subsection 4.3 of the Guidance should be revised to:

   – Address grounding models in reality through use of prior knowledge.

   – Provide a more detailed discussion on how to decide when to use each method for modeling stressor-response relationships, and the advantages/disadvantages associated with each method.

- Provide more detail regarding the conditions under which the last two methods, non-parametric changepoint analysis (nCPA) and discontinuous regression, should be applied.

- Address inaccuracies associated with linear stressor-response functions.

3.7. Charge Question 7. Evaluating candidate stressor-response criteria

Section 5 of the draft guidance document describes how to evaluate the candidate stressor-response criteria. An approach is outlined for predicting conditions that might result after implementing different nutrient criteria. Please comment on uncertainties that would remain if water quality criteria for nutrients were based solely on estimated stressor-response relationships and in what ways other information/analysis would help address and possibly reduce this uncertainty?

Section 5 of the Guidance is an important part of the document because selection of inappropriate criteria will result in negative environmental, social, and economic consequences. We provide the following comments and findings in response to Charge Question 7.

Findings on recognizing uncertainty

- As previously discussed, the Guidance does not address or partition inherent critical uncertainties in the stressor-response approach. The Guidance describes approaches that use a data-mining exercise to demonstrate a possible cause-effect relationship for the nutrient-ecosystem response. However, the document does not address or partition inherent critical uncertainties in the stressor-response approach which, as demonstrated in examples in the Guidance and in public presentations given to the Committee, can be extremely large (e.g., several orders of magnitude). Because of the demonstrated uncertainties, prediction from an empirical stressor-response model for a specific system of interest cannot always be interpreted as an accurate prediction of future conditions.

- Uncertainty also results from climatic or other environmental conditions under which studies were conducted. In addition to uncertainties documented in the Guidance and in the public presentations to the Committee, uncertainty also results from the climatic or other environmental conditions under which empirical studies were conducted and response models developed. Studies conducted over relatively limited conditions (e.g., wet or dry years) or short-term periods (e.g., base flows, summer) are unlikely to provide the robust response relationships required for criteria development.

Findings on reducing uncertainty

- A major uncertainty inherent in the Guidance is accounting for factors that influence biological responses to nutrient inputs. For criteria that meet EPA's stated goal of "protecting against environmental degradation by nutrients," the underlying causal models must be correct. Habitat condition is a crucial consideration in this regard (e.g., light [for example, canopy cover], hydrology, grazer abundance, velocity, sediment type) that is not adequately addressed in the Guidance. Thus, a major uncertainty inherent in the Guidance is accounting for factors that influence biological responses to nutrient inputs. Addressing this uncertainty requires adequately accounting for these factors in different types of water bodies.

- <u>Uncertainty in the water quality criteria for nutrients could be reduced by obtaining data from well-designed site-specific monitoring programs.</u> If "water quality criteria for nutrients were based solely on estimated stressor-response relationships," a critical overall uncertainty would be understanding where, within the range of probabilities, a single water body to which the criteria are applied will fall. This, in effect, is uncertainty in the space-for-time assumption discussed in the Guidance. That is, if the criterion nutrient concentration developed using an approach involving data from multiple locations is exceeded, will the predicted response and designated use impairment occur at a single location of interest? This type of uncertainty can be reduced by obtaining data from well-designed site-specific monitoring programs. Such monitoring would focus on obtaining specific information on the variability in stressor and response variables and important covariates with a goal of better defining the interactions of multiple variables on the attributes affecting the designated uses of a water body. Measurement of actual biological responses would be appropriate, emphasizing variables that respond most directly to changes in nutrient concentrations. These are typically measures of primary productivity or primary producers, or water chemistry changes such as DO and pH. Where necessary, such data may be used to develop computer simulation models specific to the system of interest that can facilitate forecasting of stressors and associated responses.

- <u>Numeric nutrient criteria developed and implemented without consideration of site specific conditions can lead to management actions that may have negative social and economic and unintended environmental consequences without additional environmental protection.</u> The Committee emphasizes the importance of not only recognizing but also making allowance in the Guidance for conditions specific to the system of interest so that the resulting science allows the best management decisions to be made.

- <u>The Guidance can be used to develop numeric nutrient criteria in a tiered, weight of evidence assessment using appropriately modified EPA approved procedures together with other approaches that address causation.</u> Large uncertainties in the stressor-response relationship and the fact that causation is neither directly addressed nor documented indicate that the stressor-response approach using empirical data cannot be used in isolation to develop technically defensible water quality criteria that will "protect against environmental degradation by nutrients." The Guidance can, however, be used in a tiered, weight of evidence assessment (using appropriately modified U.S. EPA-approved procedures, e.g., EPA's Causal Analysis/Diagnosis Decision Information System [CADDIS]), (U.S. EPA, 2009b).

- EPA should consider addressing the use of probabilistic modeling (using the distribution of data in the model and re-sampling or simulating a new distribution) to better determine significant stressor-response relationships. For instance, a statistically significant stressor-response relationship can be derived that may represent only a small portion of the variability in the data. Relying solely on this relationship would result in a tremendous amount of uncertainty in the final criterion developed. A good example of this is Figure 14 (p. 25) of the Guidance, which shows a statistically significant model that explains only 5% of the variation in the data – meaning that 95% of the variation is not explained by the model. Guidance on model selection is critical to reducing uncertainty. The selection of target numeric criteria as outlined in the Guidance is enhanced by the attempt to predict post-implementation conditions. However, the example used in Figures 29 and 30 of the Guidance is confusing as it appears that the values are re-projected using one criterion value (log TP=2) and the prediction analysis is made (i.e., that all 8 of the sites would still exceed the criterion) using a different value (log TP=1.6).

Findings on criteria application and monitoring for assessment

- Caution is urged in using overall regression to predict conditions that might result after implementing different nutrient criteria because the approach presented in Section 5 of the Guidance is highly sensitive to deviations of individual data points from the regression line at specific portions of the relationship. For example, in Figures 30 and 31 of the Guidance, near the intersection of TP and chlorophyll a targets and candidate criteria more than half of the data points fall above the regression line which reflects the best fit to all the data. Projecting back to lower TP concentrations for each of these individual data points would force a lower TP criterion than would be the case if the data were actually normally distributed around the regression line. In other cases, there may be a "cluster" of data points below the regression line, and the back-projected TP criterion would be higher than if all data points were distributed randomly about the regression line.

- The Guidance does not adequately address the important issue of continued monitoring and assessment for adaptive management. With regard to application of numeric nutrient criteria, Section 5 of the Guidance discusses comparison of predicted and observed data to evaluate response(s), along the lines of adaptive targets. This intrinsically implies that continued monitoring and assessment of concentration versus biological response is taking place. While this is a good idea in principle, it is not clear from the Guidance that this is to be done, how it is to be done, or at what scale it should be done. This is important because it relates to the issue of measuring changes in indicators of biological response as nutrient inputs are reduced to waterbodies. It is unclear how hereditary or legacy losses or inputs of N and P to water bodies will be considered and accounted for in such an empirical approach. This begs the next set of questions facing water resource managers who establish targets for nutrient loss reduction: “if no water quality improvement or indicator biological response is seen, are the targets / criteria too high or are legacy nutrient inputs increasingly significant contributors?”; and “how long does it take dynamic ecosystems and watersheds to respond to changing nutrient inputs?”

- The Guidance should address a number of questions to clarify how the evaluation of candidate stressor-response criteria will occur, presumably through monitoring. These questions include the following:
  - While a sound monitoring program will be essential, what form will this take?
  - At what level in time and space will monitoring be established to evaluate criteria?
  - Where, when, and how will samples be collected to establish a long-term monitoring program to clearly define and measure candidate response(s) to any changes in management and stressor inputs, as predicted by nutrient criteria?
  - How will monitoring be conducted to give a whole watershed assessment, considering all nutrient sources and stressors that are contributing spatially and temporally?
  - How will continued legacy stressor inputs (N and P) be distinguished from management change-related decreases? Internal recycling of nutrients can mask water quality improvements brought about by nutrient loss reductions resulting from land management changes.

- The direct and indirect effects of best management practices should be captured in setting numeric nutrient targets and evaluating responses to target reductions. Implementation of practices to decrease nutrient losses or inputs to surface waters (i.e., best or beneficial management practices [BMPs]) can influence other factors that will affect biological response to nutrient loadings. For instance, riparian buffers are effective at removing sediment and sediment-bound nutrients (particularly P), as well as removing N by uptake and denitrification. However, they also provide shade and will influence stream water temperature and thereby the stressor-response relationship. Such interactions should be addressed in nutrient criteria development. In addition, the use of buffers, for example, will influence the size of particulates or sediment in a stream or river that may affect the benthic population dynamics or species diversity. These direct and indirect effects and complexities should be captured in target setting and the evaluation of response to achieving target reductions.

Key Recommendations in response to Charge Question 7

The Committee provides the following key recommendations to address the comments and findings above.

Key Recommendations with regard to recognizing uncertainty

1. The Guidance needs to clearly indicate that the empirical stressor-response approach does not result in cause-effect relationships; it only indicates correlations that need to be explored further. For example, the words “cause-effect” should be removed from the title of Step 2.

2. The Guidance should address partitioning the uncertainty among the various factors that are involved in the stressor-response relationship for the specific region/system of interest. Some variables may be irrelevant to the hypothesized model for that system.

3. The Guidance should better document the physical, chemical and biological variables comprising the relationships (e.g., habitat, spatial, and temporal) that define the aquatic system, and which may be important in modifying the relationship between nutrient concentrations and observed endpoints. These factors need to be well documented, so that the uncertainty in the relationship between nutrient concentrations and measured endpoints can be reduced.

Key recommendations with regard to conceptual models and uncertainty description/analysis

4. The Guidance should caution users about potential problems associated with using the overall regression to predict conditions that might result after implementing different nutrient criteria.

5. EPA should consider addressing the use of probabilistic modeling to better determine significant stressor-response relationships.

6. The Guidance should address uncertainty resulting from climatic or other environmental conditions under which studies were conducted.

7. EPA should avoid using "biased" databases (i.e., that do not contain the range of data necessary to fully characterize a system of interest) to develop stressor-response relationships.

8. Ranges of values for stressors and responses in empirical models should fully encompass not only the current conditions in systems of interest, but also the predicted values for the stressors and responses corresponding to removal of the designated use impairment.

9. The Committee recommends predicting conditions that might result after implementing different nutrient criteria and testing these conditions on specific data-rich systems of interest.

10. The Committee recommends that EPA frame uncertainty according to the following key issues:

    – What are the goals of the decision makers (e.g., what are the designated uses and when are they impaired?), and what amount of certainty is required to make that decision?

    – Are the mechanisms of the cause-effect relationship understood and are they reflected in the types of measurements recommended?

- Do the variables measured reflect the goals of the Clean Water Act? In the examples presented in Section 5 of the Guidance species richness or chlorophyll a are not clearly linked to the stated goals (fishable, swimmable waters, etc).

- Does the analysis tool reflect a known cause-effect relationship and does it allow an understanding of the process?

- What are the *a priori* criteria to be met by the data? This must be established to make it possible to tell when the data cannot support the decision making process.

4. REFERENCES

Burnham K.P. and D.R. Anderson. *Model Selection and multimodel Inference: A Practical Information-Theoretic Approach.* 2002. Springer-Verlag, NY, 488 pp.

Carleton, J.N., M.C. Wellman, A.S. Donigian, J.C. Imhoff, J.T. Love, R.A. Park and J.S. Clough. 2005. Nutrient Criteria Development with a Linked Modeling System: Methodology Development and Demonstration Case Studies for Blue Earth, Rum and Crow Wing Rivers, Minnesota. EPA-823-R-05-003. U.S. Environmental Protection Agency, Office of Water and Office of Science and Technology, Washington, DC.

Cerco, C.F. and M.R. Noel. 2004. The 2002 Chesapeake Bay Eutrophication Model. EPA 903-R-04-004. U.S. Environmental Protection Agency, Region III, Chesapeake Bay Program Office, Annapolis, MD, and U.S. Army Corps of Engineers, Engineer Research and Development Center, Vicksburg, MS.

Conley, D.J., H.W. Paerl, R.W. Howarth, D.F. Boesch, S.P. Seitzinger, K.E. Havens, C. Lancelot and G.E. Likens. 2009. Controlling eutrophication: nitrogen and phosphorus. *Science* 323:1014-1015.

Florida Department of Environmental Protection. 2009. Draft Technical Support document: Development of Numeric Nutrient Criteria for Florida Lakes and Streams. Standards and Assessment Section, Tallahassee, FL [Available at: http://www.dep.state.fl.us/water/wqssp/nutrients]

Hagy, J.D., W.R. Boynton, C.W. Keefe and K.V. Wood. 2004. Hypoxia in Chesapeake Bay, 1950-2001: Long-term change in relation to nutrient loading and river flow. Estuaries. 27(4):634-658.

Helsel D.R. and R.M. Hirsch. 1992. *Statistical Methods in Water Resources*. Elsevier, NY, 522 pp. [Available online at: http://www.practicalstats.com/aes/aes/AESbook_files/HelselHirsch.PDF]

Helsel, D.R., Mueller, D.K., and Slack, J.R., 2006, Computer program for the Kendall family of trend tests: U.S. Geological Survey Scientific Investigations Report 2005–5275, 4 p.

Hickey, G.L. 2008. Making species salinity sensitivity distributions reflective of naturally occurring communities: using rapid testing and Bayesian statistics. *Environmental Toxicology and Chemistry* 22:2403-2411.

Hill, W.R. and S.E. Fanta. 2008. Phosphorus and light colimit periphyton growth at subsaturating irradiances. *Freshwater Biology* 53:215-225.

Hill, W.R., S.E. Fanta and B.J. Roberts. 2009. Quantifying phosphorus and light effects in stream algae. *Limnology and Oceanography* 54:368-380.

Kutner M.H., C.J. Nachtsheim, J. Neter and W. Li. 2004. *Applied Linear Statistical Models*, 5th Edition. McGraw-Hill Irwin, NY, 1396 pp.

Lewis, W.M., Jr. and W.A. Wurtsbaugh. 2008. Control of lacustrine phytoplankton by nutrients: erosion of the phosphorus paradigm. *International Review of Hydrobiology* 93:446-465.

Leung, K.M.Y., A. Bjorgesester, J. Gray, W.K. Li, G.C.S. Lui, Y. Wang, and P.K.S. Lam. 2005. Deriving sediment quality guidelines for field-based species sensitivity distributions. *Environmental Science and Technology* 39:5148-5156.

Maine Department of Department of Environmental Protection. 2009. Nutrient Criteria for Fresh Surface Waters. http://www.maine.gov/dep/blwq/rules/Other/nutrients_freshwater/index.htm . [Accessed on October 25, 2009]

McLaughlin, K. and M. Sutula. 2007. Developing Nutrient Numeric Endpoint and TMDL Tools for California Estuaries: An Implementation Plan. Southern California Coastal Water Research Project Technical Report 540. Costa Mesa, CA [Available at: ftp://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/540_CA_NNE_PhaseII.pdf]

Mississippi River/Gulf of Mexico Watershed Nutrient Task Force. 2008. Gulf Hypoxia Action Plan 2008 for Reducing, Mitigating, and Controlling Hypoxia in the Northern Gulf of Mexico and Improving Water Quality in the Mississippi River Basin. Washington, DC.

Paerl, H.W. 2009. Controlling eutrophication along the freshwater-marine continuum: dual nutrient (N and P) reductions are essential. *Estuaries and Coasts* 32:593-601.

Scavia, D., D. Justic and V.J. Bierman, Jr. 2004. Reducing hypoxia in the Gulf of Mexico: Advice from three models. Estuaries. 27(3):419-425.

Turner, R.E., N.N. Rabalais and D. Justic. 2008. Gulf of Mexico hypoxia: alternate states and a legacy. Environmental Science and Technology. 42:2323-2327.

U.S. EPA. 2000a. *Nutrient Criteria Technical Guidance Manual: Rivers and Streams.* EPA-822-B-00-001. U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA. 2000b. *Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs*. EPA-822-B00-001. U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA. 2001. *Nutrient Criteria Technical Guidance Manual: Estuarine and Coastal Marine Waters*. EPA-822-B-01-003. U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA. 2006a. *Data Quality Assessment: A Reviewers Guide (QA/G-9R)*. EPA/240/B-06/002. U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA. 2006b. *Data Quality Assessment: Statistical Tools for Practitioners (QA/G9s)*. EPA/B-06/003. U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA. 2008. *Nutrient Criteria Technical Guidance Manual: Wetlands*. EPA-822-B-08-001, U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA 2009a. *Guidance on the Development, Evaluation, and Application of Environmental Models*. EPA/100/K-09/003. Office of the Science Advisor, Council for Regulatory Environmental Modeling, U.S. Environmental Protection Agency, Washington, DC.

U.S. EPA. 2009b. *CADDIS: Helping Scientists Identify the Causes of Biological Impairments*. http://cfpub.epa.gov/caddis/ [Accessed September 15, 2009]

U.S. EPA. 2006c. Quality Management Tools - Systematic Planning. http://www.epa.gov/quality1/dqos.html [Accessed November 11, 2009]

Weisberg, S. 1985. *Applied Linear Regression*, 2nd Edition. John Wiley & Sons, New York, 324 pp.