EPA
United States
Environmental
Protection Agency

EPA/600/R-24/057 | May 2024 | www.epa.gov/research

Application of Weight-of-Evidence Methods for Transparent and Defensible Numeric Nutrient Criteria

Office of Research and Development
Center for Public Health and Environmental Assessment

Front Cover Photo Credits
Top: Lake Okoboji, IA; Sylvia S. Lee (U.S. EPA).
Bottom: Potomac River, Washington, DC; Eric Vance (U.S. EPA)

EPA/600/R-24/057

May 2024

www.epa.gov/research

Application of Weight-of-Evidence Methods for Transparent and Defensible Numeric Nutrient Criteria

By

Caroline E. Ridley[1], S. Douglas Kaylor[1], Sylvia S. Lee[1], Jesse N. Miller[2], Sam Penry[3], and Kate A. Schofield[4]

Center for Public Health and Environmental Assessment
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, NC

[1]U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment, Integrated Climate Sciences Division
[2]Oak Ridge Institute for Science and Education
[3]Oak Ridge Associated Universities
[4]U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment, Health and Environmental Effects Assessment Division

Notice and Disclaimer

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication.

This document is for informational purposes only and is not intended as legal advice. The contents are for general informational purposes, and should not be construed as legal advice concerning any specific circumstances. You are urged to consult legal counsel concerning any specific situation or legal issues. This document does not address all federal, state, and local regulations, and other rules may apply. This document does not substitute for any EPA regulation and is not an EPA rule.

Table of Contents

Figures

Tables

Boxes

Acronyms and Abbreviations

BCA	Bray-Curtis cluster analysis
CADDIS	Causal Analysis/Diagnosis Decision Information System
CARE	collective benefit, authority to control, responsibility, ethics
CEE	Collaboration for Environmental Evidence
Chl-a	chlorophyll *a*
CWA	Clean Water Act
DO	dissolved oxygen
EFSA	European Food Safety Authority
FAIR	finable, accessible, interoperable, reusable
GIS	geographic information systems
HABs	harmful algal blooms
HBI	Hilsenhoff Biotic Index
LAGOS-NE dataset	Lake Multi-Scaled Geospatial and Temporal Database
MMI	multi-metric index
NARS	National Aquatic Resource Surveys
NAWQA	National Water Quality Assessment
nCPA	non-parametric changepoint analysis
NNC	numeric nutrient criteria
NPDES	National Pollutant Discharge Elimination System
N-STEPS	Nutrient Scientific Exchange Partnership and Support
QA	quality assurance
QAPP	quality assurance project plan
r	correlation coefficient
SAV	submerged aquatic vegetation
S-R	stressor-response
TMDL	total maximum daily load
TN	total nitrogen
TP	total phosphorus
TV	tolerance value
USEPA OST/HECD	USEPA Office of Science and Technology/Health and Ecological Criteria Division
USEPA OW	USEPA Office of Water
USGS	Unites States Geological Survey
WHO	World Health Organization
WoE	weight of evidence
WQS	water quality standards

Executive Summary

Water quality standards are important for protecting and restoring the condition of lakes, rivers, estuaries, and other water bodies in the United States. Given that nutrient pollution continues to be a widespread problem in aquatic systems, the development of numeric nutrient criteria (NNC) as part of water quality standards is a priority to enhance prospects for managing excess nutrients and their effects. In this report, we complement existing NNC guidance and support to states by discussing weight-of-evidence (WoE) methods that enable rigorous and transparent development and integration of multiple lines of evidence.

The NNC development phases (Planning, Problem Formulation, Analysis, and Derivation) align with the Basic WoE Framework steps (Assemble Evidence, Weight Evidence, and Weigh the Body of Evidence) (Figure ES-1). The process is conducted within the context of the WoE core principles of transparency, documentation, and communication.

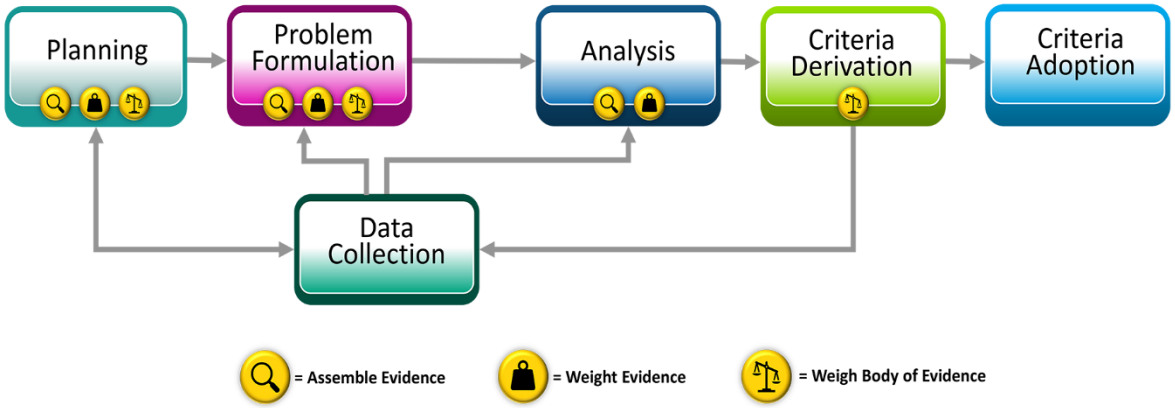


Figure ES-1. Criteria Development and the Weight-of-Evidence Steps Align

Criteria development phases (boxes) and steps of the *Basic Weight-of-Evidence Framework* (circles) align. Planning and Problem Formulation involve activities that span all steps of the framework. Analysis aligns with assembling and weighting evidence. Criteria Derivation aligns with weighing the body of evidence. Note that the similar sounding "Weight" and "Weigh" steps comprise distinct activities. Data Collection is shown outside of the criteria development phases, but it is closely associated with and both feeds into and is fed by various activities that happen during multiple criteria development phases.

The following are take-home messages for the role WoE can play in strengthening each phase of NNC development.

Planning Phase- Activities undertaken during Planning provide a transparent foundation for developing NNC; transparency is a core principle of WoE. Grouping water bodies during Planning is a process to which WoE could be applied when diverse evidence needs to be combined.

Problem Formulation Phase- Selecting endpoints during Problem Formulation is also a process to which WoE could be applied when diverse evidence needs to be combined. Conceptual models developed during Problem Formulation can help inform what evidence should be assembled in the Analysis Phase.

Analysis Phase- This phase includes assembling evidence and weighting evidence. Unbiased assembly of evidence is best practice and can ensure NNC are based on transparent data and information of sufficient amount and quality. Weighting evidence by establishing, objectively evaluating, and documenting qualities of that evidence shows how much influence individual evidence will have on overall NNC conclusions.

Criteria Derivation Phase- This phase includes weighing the body of evidence by integrating and interpreting evidence, as well as communicating conclusions. Methods for integrating evidence to derive criteria can range from simple to sophisticated; selected methods should be logical, informed by evidence availability and stakeholder needs, and communicated clearly.

Overall, there are sets of intended outcomes when WoE methods are applied to NNC development (Table ES-1). Those outcomes occur during the process of criteria development and ultimately result in improved water quality and the protection of designated uses once criteria are adopted and implemented.

Table ES- 1. Summary of Suggested Practices and Intended Outcomes

A summary of suggestions for how to carry out WoE and what can be achieved.

Criteria Development Phase	Basic WoE Framework Element	Key Suggested Practices	Intended Outcomes
Planning	Core principle	Planning is transparent, documented, and leverages collective expertise.	Decision-makers and stakeholders understand and trust the criteria development process. Planning minimizes bias, is realistic, and meets stakeholder needs.
	Assemble, weight, weigh	WoE methods are used to group water bodies.	When **Criteria Derivation Phase** is reached, candidate criteria for each water body grouping have acceptably low amounts of variation.
Problem Formulation	Assemble, weight, weigh	WoE methods are used to select endpoints.	Endpoints are relevant to management goals, measurable, ecologically relevant, sensitive to nutrients, and important to stakeholders.
Analysis	Assemble	Evidence is assembled in an unbiased way.	Conclusions reached in the **Criteria Derivation Phase** are objective and defensible because they are based on evidence that accurately and fairly represents what is known about nutrients and their effects in water bodies.
	Weight	Weighting criteria are established ahead of time; relevance, strength, and reliability of evidence are assessed and documented.	Each piece of evidence has influence on the conclusions in the **Criteria Derivation Phase** that appropriately corresponds to its objectively evaluated relevance, strength, and reliability.

Criteria Development Phase	Basic WoE Framework Element	Key Suggested Practices	Intended Outcomes
	Core principle	Processes for assembling and weighting evidence are documented and communicated clearly.	Decision-makers and stakeholders understand the pieces of evidence that make up the body of evidence and how they influence conclusions in the **Criteria Derivation Phase**.
Criteria Derivation	Weigh	If necessary, evidence is logically aggregated. Integration method is appropriate for the evidence.	Derived criteria are sound and defensible, because the method to either (a) select the weightiest evidence or (b) merge multiple lines of sufficiently weighty evidence is technically appropriate and justified to protect the designated use.
	Core principle	Conclusions are clearly communicated.	Decision-makers and stakeholders understand and trust the derived criteria.

Introduction

1.1. Background

Water quality standards are important for protecting and restoring the condition of lakes, rivers, estuaries, and other water bodies in the United States. Designated uses, water quality criteria to protect designated uses, and antidegradation requirements are the core components of state water quality standards (33 U.S.C. § 1313(c)).[1] Together, they function to protect the health of humans who enjoy these waters and the aquatic life that call these waters home. There are many successful examples of restoring water bodies to meet water quality standards, but challenges remain.[2]

Nutrients continue to be a widespread stressor in U.S. water bodies. Latest results from the National Aquatic Resources Surveys (NARS) indicate that 58% of streams and rivers have a phosphorus concentration at or above the 95th percentile of reference sites and 43% have a nitrogen concentration at or above the 95th percentile of reference sites (U.S. EPA, 2019). For lakes, 45% meet that same mark for phosphorus and 46% for nitrogen (U.S. EPA, 2022a). By states' own accounting (pursuant to 33 U.S.C. § 1313(d)), 55% of assessed stream and river miles and 70% of assessed lake acres were listed as impaired as of July 2016, with nutrients being a commonly identified cause of impairment (U.S. EPA, 2017b).

The development of numeric nutrient criteria (NNC) as part of water quality standards can enhance prospects for managing nutrient pollution and its effects. NNC enable effective monitoring and assessment of water bodies (33 U.S.C. § 1315(b)), facilitate formulation of national pollutant discharge elimination system (NPDES) permits (33 U.S.C. § 1342), and simplify development of total maximum daily loads (TMDLs) (33 U.S.C. § 1313(d)). Since 1998, states in the U.S. have made progress in adopting NNC for their waters; however, only Hawai'i has adopted a complete set of numeric criteria for total nitrogen (TN) and total phosphorus (TP) for all its applicable water body types (U.S. EPA, 2022c).

Considerable guidance and support exist for the development of NNC and other water quality benchmarks. The U.S. EPA has published technical documents broadly relevant across the country (e.g., (U.S. EPA, 2021a, 2013, 2010b, 2001a, b, 2000a, b, c)) and provides technical support to individual states and tribes upon request through the Nutrient Scientific Exchange Partnership and Support (N-STEPS) program (pursuant to 33 U.S.C. § 1314(a)). This guidance refers to multiple lines of evidence and that combining one or more of those lines using weight-of-evidence "will produce criteria of greater scientific validity" (U.S. EPA, 2000c).

Among many useful approaches for NNC development, considering multiple lines of evidence has emerged as valuable for several reasons. Like many common stressors (e.g., bedded sediment, conductivity), nutrients can have a variety of direct and indirect effects on the diverse taxa comprising biological communities. As a result, nutrients can affect these communities via a number of different pathways (Bennett et al., 2021; Ryan, 2021; Cook et al., 2018; Munn et al., 2018; Hilton et al., 2006; Carpenter et al., 1998). In addition, the evidence applicable to NNC development can be diverse in that it is generated using a variety of experimental designs and analysis methods. Further, evidence may be

[1] Under the Clean Water Act (CWA), 'state' refers to states, territories, and authorized tribes.

[2] https://www.epa.gov/nps/success-stories-about-restoring-water-bodies-impaired-nonpoint-source-pollution#read

associated with greater heterogeneity or uncertainty because much of it is field-based (although certain lab- and mesocosm-based evidence can still be relevant and useful) (Cormier et al., 2008). Finally, development of NNC can attract substantial and diverse stakeholder interest, which requires that both the process and conclusions are rigorous and transparent.

In this report, we complement existing NNC guidance and support by discussing weight-of-evidence (WoE) methods that improve rigorous and transparent development and integration of multiple lines of evidence. WoE methods are the specific operational details embedded within the *Basic WoE Framework* (Figure 1). As a whole, it is a process in which scientific evidence is assembled, evaluated, and integrated to make a technical inference (U.S. EPA, 2016)[3]. The WoE Framework presented and discussed in this report has three steps (Assemble Evidence, Weight Evidence, and Weigh the Body of Evidence). The process is conducted within the context of the core principles of transparency, documentation, and communication.

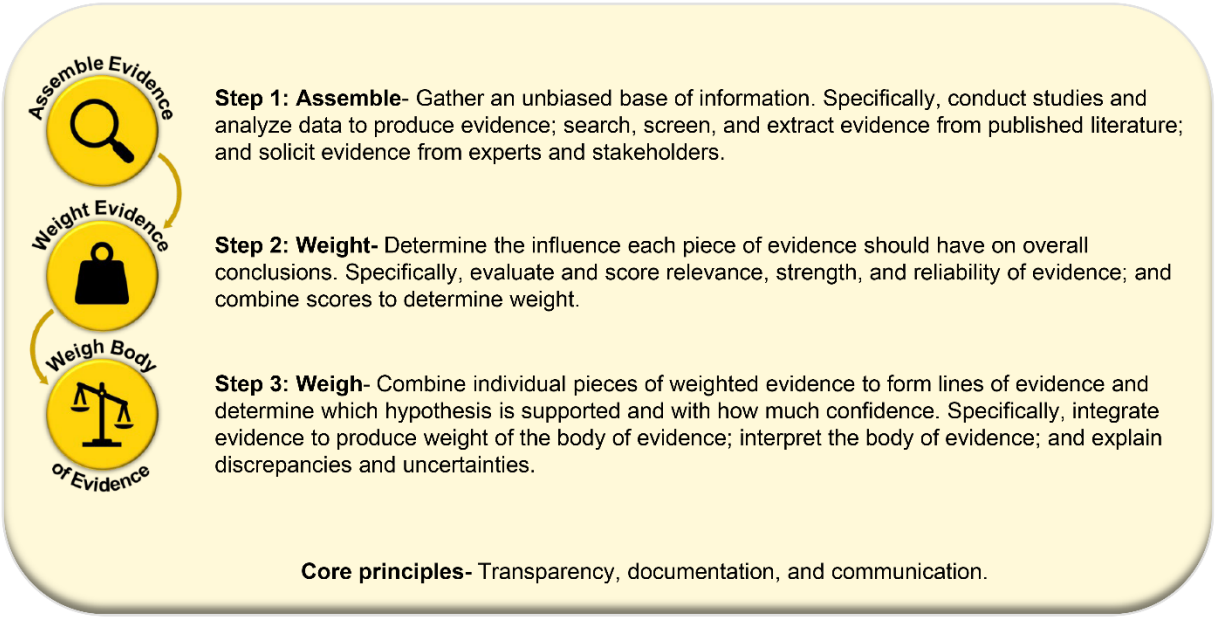


Figure 1. Basic Weight-of-Evidence Framework

The steps and core principles of the Basic Weight-of-Evidence Framework (U.S. EPA, 2016). Note that the similar sounding "Weight" and "Weigh" steps comprise distinct activities.

WoE as a concept is well-established and applicable to many programs under the CWA, including (but not limited to) criteria development (Table 1). WoE can be used to infer qualities, such as identifying likely causes of impairment, or to infer quantities, such as a numeric nutrient criterion (Suter et al., 2017a, b). NNC development is considered a quantitative, predictive assessment that informs a type and

[3] Weight-of-evidence has been defined by the World Health Organization as "a process in which all of the evidence considered relevant for a risk assessment is evaluated and weighted" (WHO & FAO, 2009). The European Food Safety Authority defines a weight-of-evidence assessment as "a process in which evidence is integrated to determine the relative support for possible answers to a scientific question" (EFSA, 2017).

level of effect not to be exceeded, and then relates it to a level of exposure that constitutes the criterion (Suter and Cormier, 2008).

Table 1. Using WoE in Environmental Assessment under the Clean Water Act

Four types of environmental assessment, the Clean Water Act program and questions related to water body condition that fall within each type, and examples of diverse evidence that might be combined to answer the questions.

Assessment Type	Clean Water Act Program	Question	Evidence that could be combined to answer the question using WoE
Condition assessment	33 U.S.C. § 1315(b)-State reports on water quality	Is a specific water body biologically impacted?	Multiple biological endpoint measurements (e.g., algal biomass, macroinvertebrate index, fish abundance)
Causal assessment	33 U.S.C. § 1313(d)-TMDLs	What is the likely cause of impairment?	Observational field evidence showing time order and biological sufficiency; experimental evidence showing dose-response relationships; literature evidence showing causal relationships under similar circumstances
Predictive assessment	33 U.S.C. § 1314(a)-Criteria development	What level of the cause will reduce impairment? What level of the cause will ensure attainment of designated uses?	Reference condition; stressor-response for multiple endpoints; literature evidence showing levels under similar circumstances
Outcome assessment	33 U.S.C. § 1315(b)-State reports on water quality	Does reaching the predicted level of the cause result in recovery?	Multiple biological endpoint measurements

For the purposes of this report, we use specific terminology to describe both the discrete and aggregate components of evidence. The terms and their conceptual relationship to one another are presented in Box 1. Understanding this terminology is important for understanding the methods described later in the report.

Box 1. Weight-of-Evidence Terminology

Modified from EFSA (2017); U.S. EPA (2016).

Basic WoE Framework: A process in which scientific evidence is assembled, evaluated, and integrated to make a technical inference.

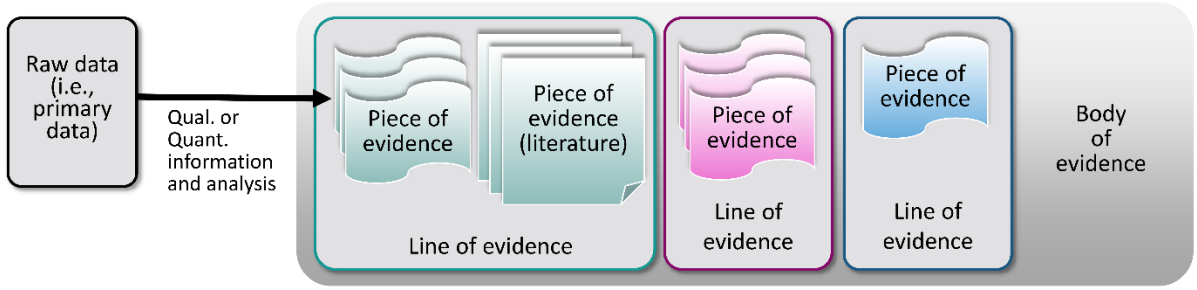
WoE method: The specific operational details embedded within the Basic WoE Framework.

Evidence: Information that enables inferences regarding a condition, cause, prediction, or outcome. Raw data (also called primary data) are generally not considered evidence until qualitatively or quantitatively analyzed.

Piece of evidence: The basic unit of evidence, e.g., a single study, expert judgment or experience, a model, or even a single observation. Pieces of evidence may be created de novo during the development of criteria or assembled from existing, published literature. They may be combined to form a line of evidence.

Line of evidence: A set of evidence that is similar in some way and establishes coherent reasoning. A line of evidence may be made up of a single piece of evidence or more than one piece of evidence and can include multiple causal or logical steps. This term is often used interchangeably with "type of evidence."

Body of evidence: All the applicable pieces and lines of evidence used to make inferences concerning a hypothesis.



The application of the *Basic WoE Framework* to NNC and other water quality benchmarks aligns with federal water quality standards regulations (40 CFR § 131.11(a)).[4] Specifically, WoE methods provide a transparent basis for selecting between or combining multiple lines of evidence. In addition, the strengths and weaknesses of diverse, individual pieces of evidence, lines of evidence, and collective bodies of evidence are made explicit, which aids both in derivation of criteria and in communication with stakeholders. Further, selection of WoE methods can be customized in anticipation of specific decision contexts (e.g., site-specific water quality criteria development, protection of highest quality waters designation).

1.2. Purpose of this report

The purpose of this report is to:

1. Describe the core principles and essential steps of the *Basic WoE Framework* and how the framework aligns with the phases of criteria development.

[4] "States must adopt those water quality criteria that protect the designated use. Such criteria must be based on sound scientific rationale and must contain sufficient parameters or constituents to protect the designated use" (40 CFR § 131.11(a)).

2. Provide a suite of state-of-the-art WoE methods to combine diverse evidence generated from various data types. Methods will be appropriate for different evidence and decision contexts that may be encountered by state and tribal nutrient managers.
3. Provide examples for communicating (especially visually) WoE as a process and its conclusions.

With this report, we anticipate that states will be able to:

1. Maximize the use of available evidence during the development of NNC.
2. Decide which WoE methods to use, given their own unique evidence, resource, and timeline constraints.
3. Further strengthen the transparency and defensibility of both the NNC development process and the derived criteria.

1.3. Approach of this report

We begin this report by briefly describing the EPA's *Basic WoE Framework* in the context of the criteria development phases (Section 1.4). Then, each criteria development phase is addressed in sequential chapters (Chapters 2-5). For each phase, we present options and details for a suite of WoE methods and provide additional resources and references. Throughout the text, we include examples of WoE methods that have been successfully applied by states in the past when developing NNC or examples that appear in the scientific literature. In several places, we develop purely hypothetical examples to demonstrate the application of a method or potential outcome of choosing one method over another. In Chapter 6, we describe two more forward-looking and complete examples based on a range of real situations and challenges that states and tribes may be facing in developing NNC. For these two examples, we suggest WoE methods that could be appropriate given different time, resource, evidence, and decision constraints. We end the main report with conclusions (Chapter 7). Two appendices provide additional background on the examples in Chapter 6 (Appendix A) and optional exercises for readers to practice weighting and weighing evidence themselves (Appendix B).

This report is complementary to EPA-published technical guidance and scientific support for criteria development. The report should not be construed as formal EPA regulatory guidance. The report is focused on freshwater NNC, although the *Basic WoE Framework* and the methods herein could also be applied to estuarine and wetland criteria. It is written in plain language to avoid being overly technical and is rich in visual examples.

Additionally, this report is not prescriptive, in that it does not require the use of any methods presented. It is also not proscriptive; it does not exclude the use of specific methods but may comment on the appropriateness of methods in the context of NNC. The report is not exhaustive, in that it does not attempt to cover all WoE methods in the extensive literature on the subject. If a method does not appear, that does not necessarily mean it is not useful or appropriate given the right circumstances. Finally, we do not attempt to derive criteria for examples involving real situations or states.

The work within this report was conducted under EPA-approved Quality Assurance Project Plans (QAPPs).[5] Independent QA audits were not deemed necessary because the QAPPs apply to a Category B project, under which audits are voluntary. The report was reviewed by QA management, three internal technical reviewers, and three external peer reviewers.

1.4. Crosswalk between the nutrient criteria development process and the *Basic WoE Framework*

As summarized in *N-STEPS Online*, the process of developing NNC follows a series of phases: **Planning**, **Problem Formulation**, **Analysis**, **Criteria Derivation**, and **Criteria Adoption** (U.S. EPA, 2022b). These phases are designed to be carried out in order, although in practice there may be iterations that involve one or more phases. The *Basic WoE Framework* introduced in Section 1.1 can be applied to the process of NNC, because they share common elements with EPA's risk assessment paradigm ((U.S. EPA, 1998); Figure 2).

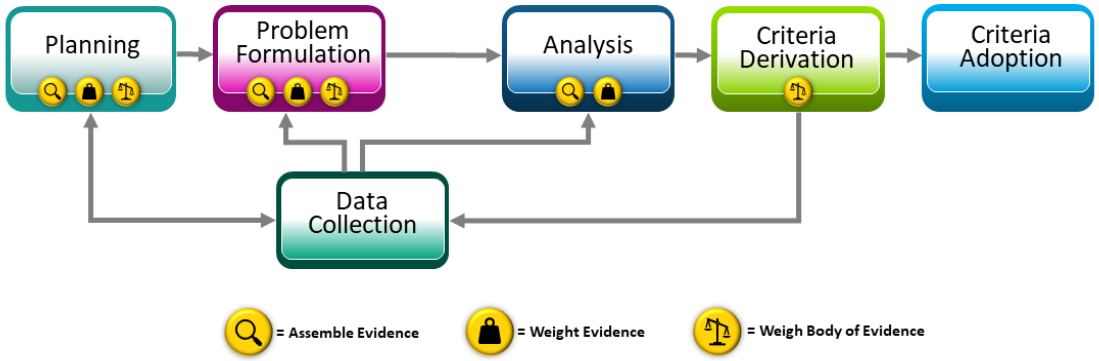


Figure 2. Phases of Criteria Development

A crosswalk between criteria development phases (boxes) and steps of the *Basic Weight-of-Evidence Framework* (circles). While "Data Collection" is shown outside of the criteria development phases, it is closely associated with and both feeds into and is fed by various activities that happen during multiple criteria development phases.

This report largely focuses on the application of the *Basic WoE Framework* across the phases of NNC development. Given that focus, the methods described in most detail are in the **Analysis Phase,** which encompasses Assemble Evidence and Weight Evidence, and for the **Criteria Derivation Phase**, which encompasses Weigh the Body of Evidence in the *Basic WoE Framework*. However, there are activities within the **Planning** and **Problem Formulation Phases** that are key for success in the following phases, so it is not recommended that the reader skip those chapters (Chapters 2 and 3). In addition, although not the focus of this report, there is limited, general information about how to apply WoE methods within the **Planning Phase** (e.g., for water body grouping decisions; Chapter 2) and the **Problem Formulation Phase** (e.g., for selection of assessment endpoints; Chapter 3).

We draw most heavily on the *Basic WoE Framework* in U.S. EPA (2016), which itself is based on decision theory and a robust body of work and experience in land and water-based condition, causal, and risk

[5] *Supporting Water Quality Goals through Literature and Weight of Evidence* (L-HEEAD-0032824-QP-1-1, approved 2-4-21) and *Development of Technical Resources for Managing Aquatic Ecosystem Stressors* (L-HEEAD-0033234-QP-1-2, approved 11-18-21).

assessments (e.g., Cormier et al. (2008); U.S. EPA (2006)). Other WoE frameworks may also be applicable to NNC development (for a review, see Martin et al. (2018)). Furthermore, straightforward conceptual analogies can enhance understanding of key components of WoE that apply to NNC development ((Salafsky et al., 2019); Figure 3).

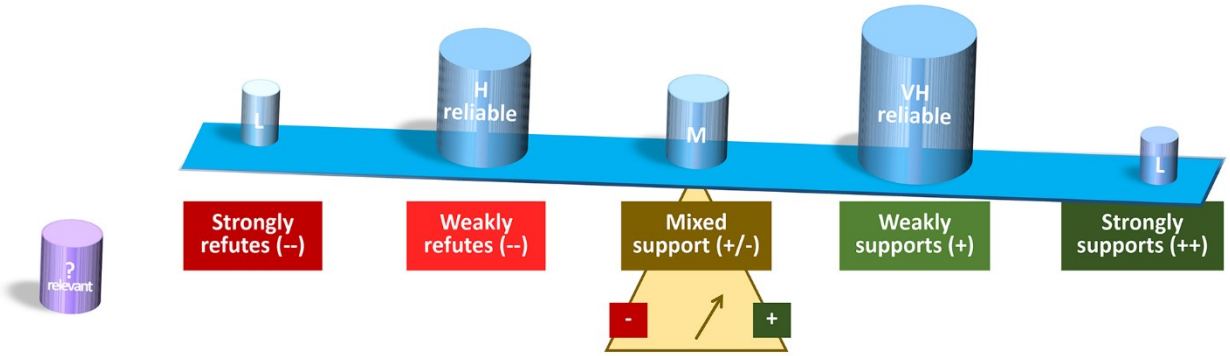
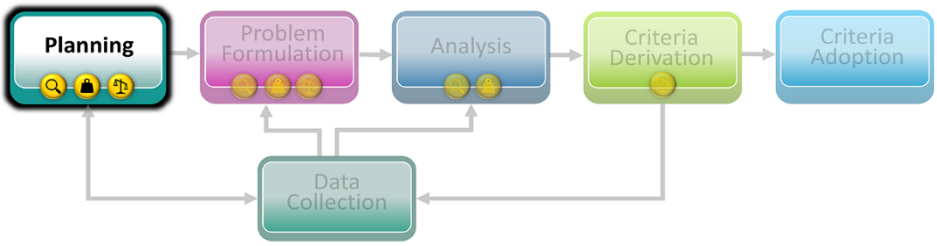


Figure 3. A Scale Analogy for Weight-of-Evidence

A simplified analogy that helps illustrate some of the important aspects of a weight-of-evidence process. Each container is a piece or line of evidence. Blue containers that are on this scale are **relevant** evidence. One container shaded purple has unclear (?) relevance. The size of the container reflects the **reliability** of the evidence. Reliability is judged to be very high (VH), high (H), medium (M), or low (L). Where it is placed on the scale reflects the **strength** of the evidence. When experts gather the containers, decide if they go on the scale, how big they are, where they are placed, and thus ultimately determine which way the scale tips, they are conducting a weight-of-evidence process. Modified from Salafsky et al. (2019).



Planning Phase

Take home: Activities undertaken during Planning provide a transparent foundation for developing NNC; transparency is a core principle of WoE. Grouping water bodies during Planning is a process to which WoE could be applied when diverse evidence needs to be combined.

Planning and scoping are important parts of any process in which technical information will be applied to decision-making (U.S. EPA, 1998, 1992). Effective planning and scoping activities ensure that stakeholder needs are considered from the outset and that the process for handling and communicating technical information will meet those needs.

During the **Planning Phase** of NNC development, state water quality standards staff determine an appropriate scope and establish management goals to direct the remaining steps in the criteria development process. During this phase, if the decision is made to develop state-specific NNC (as opposed to adoption of criteria previously developed by EPA) and more than one piece of evidence will likely be used, it is appropriate to consider WoE methods going forward.

An opportunity exists during planning to articulate the WoE methods that will be used in all other steps leading to NNC adoption and ensure that they will meet stakeholder needs. Proposed methods can be general and flexible enough to account for uncertainties or challenges that might arise. Trade-offs are inevitable when selecting WoE methods because of potentially limited resources (e.g., time, amount of evidence, expertise, and capacity to communicate with stakeholders and decision-makers (EFSA, 2017)). Therefore, a plan and potential set of WoE methods that are practical for each state will be unique.

Undertaking and documenting planning and scoping is key for transparency, which is a core principle of the *Basic WoE Framework*. For instance, a planning document can include text and figures addressing how water bodies will be classified, how candidate endpoints will be selected, and what evidence and analysis approaches will be considered. Figure 4 shows a hypothetical example of a workflow that includes WoE methods that will be used by the technical team.

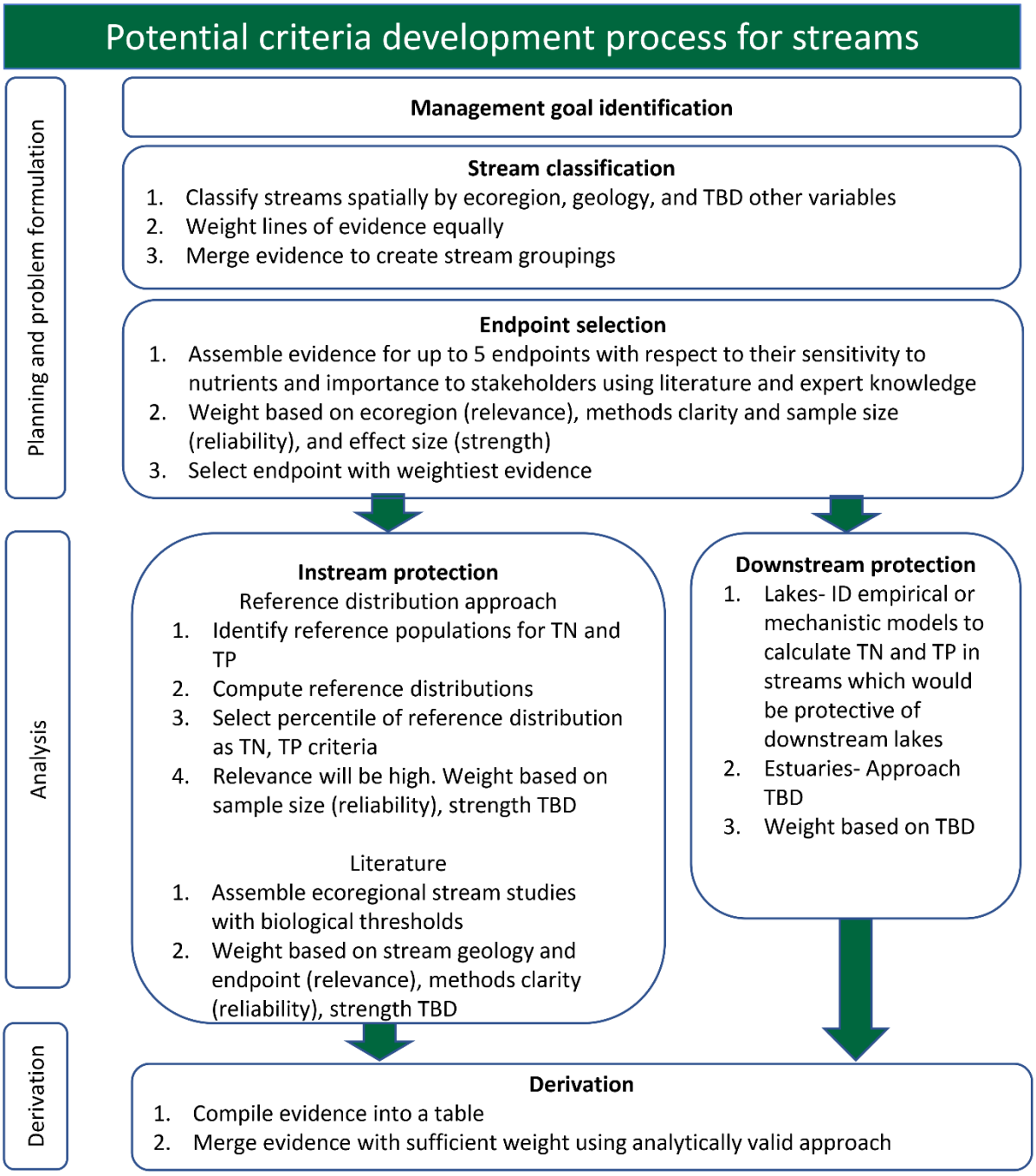


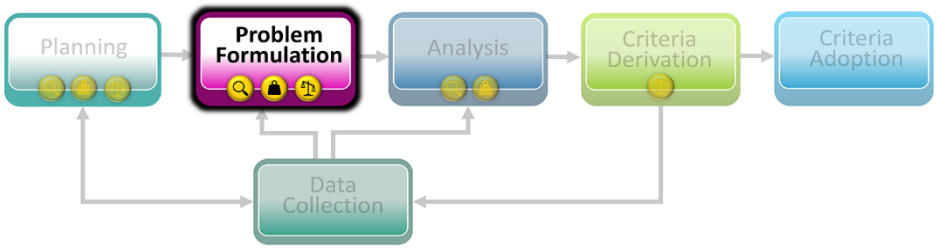
Figure 4. Hypothetical Example NNC Plan Incorporating WoE Methods

Hypothetical plan for developing numeric nutrient criteria for streams in a state that includes several references to WoE methods. Note that some methods are to be determined (TBD), because the technical team may need additional information or input before deciding how to proceed.

In addition to devising an overall plan that incorporates WoE methods, the **Planning Phase** is the time when water bodies are grouped so that management goals can be set and NNC can be prioritized and developed. Appropriately grouped water bodies can minimize the chance that candidate criteria are too

variable during the **Criteria Derivation Phase**. Delineating groups of waters may rely on multiple lines of evidence, depending on the number and variety of water bodies within a state. Indeed, diverse sources of evidence (physical, chemical, and/or biological attributes; expert knowledge) are already recommended by EPA and being used by states to group similar water bodies for which NNC can be developed (U.S. EPA, 2022b). For instance, Arkansas' *Nutrient Criteria Development Plan* proposes multiple classification variables, such as ecoregion, stream order, watershed size, gradient, and fluvial geomorphology, for streams and rivers (Arkansas DEQ, 2012). In Florida, unique underlying geology helped to define nutrient watershed regions for nitrogen and phosphorus numeric criteria along with information about biological communities and an understanding of how upstream regions affect downstream water quality (U.S. EPA, 2010a). These two examples demonstrate the existing opportunity for applying WoE methods to assemble, weight, and weigh evidence for grouping water bodies. Methods for doing so are not specifically described here, but material in Chapters 4 and 5 may be consulted for ideas.

Application of WoE methods during the **Planning Phase** likely will be most successful if collective expertise is leveraged. This may take the form of a workgroup composed of internal and external experts developing the plan, or through internal and/or external experts reviewing any plan that is developed. For instance, the Virginia Department of Environmental Quality has used collaborative workgroups of experts since 2011 as well as independent reviewers to develop strategy and refine approaches to establish numeric chlorophyll *a* criteria for segments of the James River (Virginia DEQ, 2019). Both of these options can identify capabilities and constraints of a given set of WoE methods and opportunities for refinement.



Problem Formulation Phase

Take home: Selecting endpoints during Problem Formulation is also a process to which WoE could be applied when diverse evidence needs to be combined. Conceptual models developed during problem formulation can help inform what evidence should be assembled for the Analysis Phase.

During NNC development, problem formulation is a process for generating and evaluating preliminary hypotheses about why ecological and/or human health effects have occurred or might occur as the result of nutrient pollution. As indicated in U.S. EPA (2022b), this process should result in (1) assessment endpoints that adequately reflect selected ecosystems and management goals, and (2) conceptual models that describe key relationships between nutrients and assessment endpoints.

Much like using evidence to determine water body groupings in the **Planning Phase** (see Chapter 2), selecting assessment endpoints within the **Problem Formulation Phase** is a process that could utilize the *Basic WoE Framework*. Selection of appropriate endpoints is critical to minimizing interpretation and inference challenges that could arise later in the **Criteria Derivation Phase** (see Chapter 5). Initially, water quality managers should identify candidate endpoints that are both intuitive and representative of the ecosystem and management goals expressed in state water quality standards (e.g., designated uses). *N-STEPS Online* lists many of these candidate endpoints and provides background information on their characteristics (U.S. EPA, 2022b). Evidence derived from primary data analyses, literature, existing syntheses, and expert knowledge (see Section 4.1) is then assembled to determine whether each candidate endpoint is:

1. Relevant to management goals,
2. Measurable,
3. Ecologically relevant,
4. Sensitive to nutrients, and
5. Important to stakeholders.

Recently, evidence was assembled to guide selection of a set of endpoints to protect designated uses of lakes from nutrient pollution (U.S. EPA, 2021a). Endpoints identified (e.g., zooplankton biomass and dissolved oxygen to protect aquatic life use) have national relevance and can be refined using additional state-level data (U.S. EPA, 2020a, b). The evidence used to support endpoint selection in this case included mechanism-based reasoning, literature, and existing stressor-response models. This example highlights the need to select both an endpoint (e.g., zooplankton) and attribute (e.g., biomass) during the **Problem Formulation Phase**. In another example of selecting endpoints, lake managers from

northeastern states have been working to develop diatom-based tools that support NNC derivation. Evidence being assembled in this case includes analyses of diatom assemblage data, the literature, and expert knowledge (Merrell and Lee, 2022; Potapova et al., 2022), which show diatoms are sensitive to nutrients and diatom assemblages are among the first biota in aquatic ecosystems to change in response to nutrient concentrations. Regional collaboration on endpoint selection also provides opportunities to ensure evidence and tools are not limited to ecologically arbitrary state political boundaries (i.e., they demonstrate ecological relevance). These two examples demonstrate the existing opportunity for applying WoE methods to assemble, weight, and weigh evidence for selecting endpoints. Further detail on methods for doing so are not provided here, but material in Chapters 4 and 5 may be consulted for ideas.

Problem Formulation should also result in conceptual models that visually summarize how nutrients affect selected endpoints. There are many examples to draw on ((U.S. EPA, 2022b); Figure 5). Conceptual models in the context of development of criteria serve many functions, including (1) helping to consider the kinds of evidence that might need to be assembled in the **Analysis Phase**, based on distinct linkages in the model (see Chapter 4); and (2) helping to visually communicate how the body of evidence results in the final conclusions.

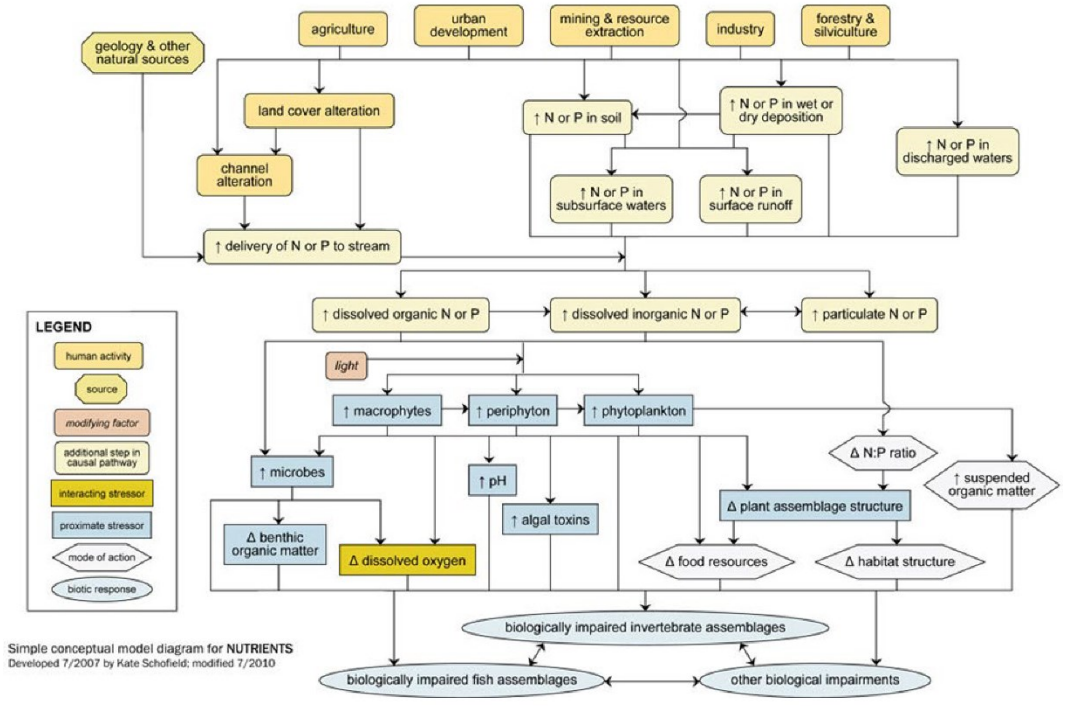
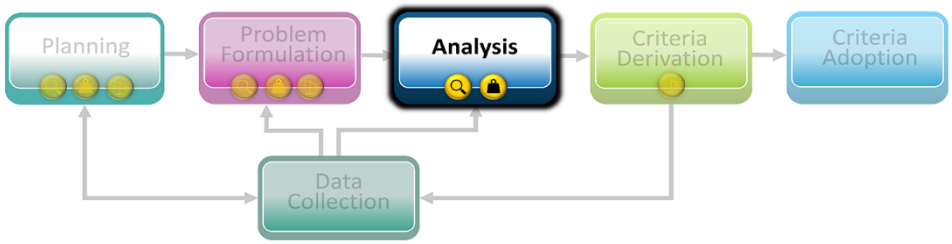


Figure 5. Conceptual Model Linking Nutrients to Aquatic Life

An example of a generic conceptual model from the CADDIS website (www.epa.gov/caddis) that links nutrients to impacts on aquatic life in streams. The boxes and arrows depicted in this model can help identify potentially relevant datasets and specific stressor-response relationships between nutrients, other nutrient-related stressors (e.g., dissolved oxygen), and endpoints that can be studied (via both *de novo* analysis of primary data and/or literature) to generate pieces and lines of evidence. The model can also be used to illustrate how evidence related to various boxes and arrows contribute to final conclusions. Taken from (U.S. EPA, 2017a).



Analysis Phase

4.1 Assemble evidence

Take home: Unbiased assembly of evidence is a best practice to ensure NNC are based on transparent data and information of sufficient amount and quality.

Evidence assembly involves gathering available pieces of information that can be weighted and integrated to support derivation of candidate criteria. Evidence itself is generated through data analysis and interpretation of results (Box 1). Raw data are input to generate evidence but are not evidence in and of themselves. Because criteria derivation often depends on generating new evidence from primary data analyses (e.g., establishing reference conditions or developing stressor-response curves using raw data from relevant sites), assembling evidence can be the most time-consuming and technical part of NNC development. However, evidence assembly using WoE methods is especially helpful for derivation of NNC, because nutrients are likely not the only stressor in aquatic ecosystems, and nutrients affect—and are affected by—many other potential stressors. The limitations of single lines of evidence and single stressor models can be overcome by gathering multiple lines of evidence from independent sources as recommended by (U.S. EPA, 2000c) and others such as Babitsch et al. (2021).

Unbiased assembly of evidence is the goal during the **Analysis Phase.** (Bias can also exist in data collection and conducting analyses; see Section 4.1.1 and other EPA guidance for how to minimize these types of biases (U.S. EPA, 2022b, 2010b)). Unbiased evidence assembly increases the defensibility and objectivity of the final conclusions. Plans and processes that were developed and reviewed during the **Planning Phase**, including which analysis approaches and sources are within scope, should be faithfully carried out; any changes should be documented and justified. While states nearly always include some primary data analysis, evidence needed to characterize and evaluate potential moderating, confounding, and interacting environmental factors can be determined from conceptual models and the literature. Thus, it is helpful to consider both primary data analysis and literature-based evidence. Furthermore, evidence from outside of geographic boundaries (e.g., from neighboring states) can be valuable, especially if the data were collected from sites with similar environmental conditions. For example, published results from field observations in streams of West Virginia and New York with conductivity and nutrient regimes representative of conditions in eastern Tennessee filled a data gap and enabled development of macroinvertebrate species sensitivity distributions (Coffey et al., 2014). States with limited primary data may need to assemble more evidence from literature, while states with plentiful primary data may not need as much evidence from literature.

Once evidence starts accumulating, observations of both positive or negative and strong or weak relationships, along with contextual information to improve understanding of moderating factors and spatial and temporal variation (e.g., across ecoregions), should be included as evidence when they are available. It is important to consider that bias could be introduced if evidence is down-weighted or omitted without justification, especially when dealing with results that may be surprising or not in agreement with preconceived expectations. Surprising results may be motivation to further explore existing data, gather additional data, and improve understanding of potential moderating and confounding factors.

While it is likely impossible to eliminate all bias during evidence assembly, documenting the steps taken to minimize bias and acknowledge remaining biases provides additional transparency to NNC derivation. Methods for minimizing bias within four different sources evidence are further discussed below.

4.1.1 Source 1: Primary data analyses

Evidence based on primary data refers to the results of data analyses conducted specifically for criteria development. These can be the results of field monitoring data analyses to assess reference conditions or quantify stressor-response relationships, experiments (including controlled laboratory, mesocosm, or field-scale exposure studies) to test stressor-response relationships, development of species sensitivity distributions, modeling outputs, change point analyses (e.g., Threshold Indicator Taxa Analysis; (Taylor et al., 2018)), or other *de novo* analyses using primary data (U.S. EPA, 2010b). All analyses have strengths and weaknesses (e.g., (U.S. EPA, 2015)), which should be considered during their selection. Section 4.2 may be consulted for ideas in addition to those below on how to maximize relevance, reliability, and strength of analyses.

The data underlying these analyses can have a big impact on the utility of the evidence (U.S. EPA, 2022b). For example, low statistical power due to small sample sizes or inadequate replication could limit the ability to detect biological responses to nutrient stressors (Francoeur, 2001). Biological responses to nutrient stressors also may be muted if ambient nutrient concentrations are already high, if other factors such as light or flow are moderating the stressor-response relationships, or if heterotrophic organisms (e.g., fungi) are driving the system's metabolism of nutrients but only autotrophic responses are measured. In addition, there may be a lag period before nutrients result in observed biological responses, which could affect the strength of measured stressor-response relationships. For example, diatoms may have a stronger relationship with nutrient concentrations from averaging periods >1 week prior to sampling than the nutrient concentration of one-time grab samples taken when diatoms were sampled (Smucker et al., 2022; Yuan et al., 2022).

For states in earlier phases of NNC development, it is helpful to consider statistical power, confounding and moderating factors, and the incorporation of lagged responses in both the field sampling design and the data analysis plan. If possible, coupling data from field observations and controlled mesocosm experiments is useful for understanding temporal and spatial dynamics, as well as confounding factors, that can influence stressor-response relationships (Taylor et al., 2018). Furthermore, states may have interest in assembling data for both N and P if pursuing dual nutrient criteria or using N:P data as part of primary data analyses for some endpoints such as harmful algal blooms (Paerl et al., 2016).

Biological assemblage composition data (e.g., diatoms and macroinvertebrates) are most likely to detect a response to nutrients if they are taxonomically consistent; assessing consistency is especially

important if the data span multiple years and/or taxonomists (Lee et al., 2019). For example, upfront management of taxonomy from counting methods, sample preparation, image-based documentation, and quality control methods will help to avoid loss of species-level information and improve the ability to generate scientifically robust evidence based on diatom assemblage composition (Alers-García et al., 2021; Tyree et al., 2020b; Tyree et al., 2020a). It is best practice to transparently document taxonomic inconsistencies that cannot be eliminated, even after extensive *post-hoc* harmonization (Potapova et al., 2022). If conducting *post-hoc* harmonization steps is not feasible, genus-level data can be used to generate multi-metric indices (MMIs) (Riato et al., 2022).

Evidence can be assembled from analyses, models, and tools derived from data from multiple spatial scales (i.e., state-specific, regional, and/or national datasets). For example, state data may be applied to regional diatom MMIs developed from national-scale datasets as an initial step to assess biological condition (Schulte et al., 2021). National-scale data from the National Lakes Assessment were used to develop models for deriving candidate criteria for TN and TP in lakes and reservoirs (U.S. EPA, 2020c). Examples of how to incorporate state data into these models are provided in U.S. EPA (2021a). Models of stressor-response relationships are useful for predicting how nutrient stressors may impact high quality waters or waters that are formally assessed as attaining the applicable nutrient water quality standard. Spatial data are necessary for these applications and geographic information system (GIS) layers are useful for visualizing the data to understand natural variation, checking if the results are sensible, and communicating with decision-makers and stakeholders.

For unbiased assembly of primary data, it is helpful to provide detailed methods for conducting the data analyses and transparent weighting of potentially different nutrient concentration values resulting from the analyses. For example, Smith and Tran (2010) provided detailed descriptions of how primary data were collected and analyzed to produce three principal types of evidence: (1) stressor-response analysis using non-parametric changepoint analysis (nCPA); (2) a multivariate assemblage change analysis using the median nutrient concentrations associated with reference, medium, and high nutrient concentrations using Bray-Curtis cluster analysis (BCA); and (3) reference analysis using the 25th percentile of all site nutrient concentrations and the 75th percentile of reference site concentrations. Empirical statistical analyses of primary data were used to create all three models. In addition to providing detailed methods for the statistical analyses, Smith and Tran (2010) also acknowledged the use of best professional judgement in weighting of the results from the three models and provided a comparison of the results to values from the published literature. Unbiased assembly of evidence from primary data analyses also requires transparency and justification of any analysis results that may be heavily down-weighted or omitted.

4.1.2 Source 2: Published literature

There are now decades of research on the effects of nutrient pollution on aquatic ecosystem structure and function (e.g., as reviewed and summarized by Carpenter et al. (1998), and Bennett et al. (2021)). Assembling this evidence from the literature entails searching, screening, and extracting evidence from publications in as transparent, rigorous, and standardized a way as possible given a state's goals and constraints. Literature-based evidence refers to the results of individual studies published in the peer-reviewed scientific literature, gray literature (e.g., government reports), and/or databases of evidence that compile results from individual studies. The most methodical and comprehensive approach to assembling literature-based evidence is a systematic review, but conducting a systematic review is not

always feasible or necessary for WoE (see Suter et al. (2020) for the essential features of systematic reviews and how to integrate with WoE, if desired). Literature-based evidence can be useful in deriving NNC, particularly when primary data are limited, even if it has lower relevance than evidence derived using data specifically matched to the sites of interest. For example, many streams in New Zealand and Montana have similar cold-water temperatures and low nutrient conditions that seem to support proliferations of the same diatom species, *Didymosphenia geminata* (Kumar et al., 2009).

In assembling literature-based evidence, bias can be minimized by conducting thorough searches of the literature with an objective screening process. Bias is more likely if evidence is only gathered from literature that is easily accessible or familiar. Specifying criteria that will be used to screen the studies that will be included as pieces of evidence reduces selection bias and provides transparency into the process of gathering literature (Suter and Cormier, 2016). The scope of a search can be accommodated based on specific needs. For example, if the objective of literature-based evidence is to support development of NNC with *comprehensive* evidence, a broad, extensive search including multiple search terms across several databases would be warranted. On the other hand, if the objective of literature-based evidence is to ensure that the *most relevant* literature supports development of NNC, a more targeted search with comparably fewer search terms and databases would be justified.

Regardless of the underlying objective, characterizing *a priori* search parameters and screening criteria and reporting *post-hoc* results of each step of the screening process are paramount to minimizing bias when assembling literature-based evidence. For instance, evidence assembly might begin with identifying the names of databases, search terms, date of when the search was conducted, and range of captured publication years. The next step is screening the literature search results using the pre-defined inclusion criteria. To detect potential bias in which studies are included or excluded, it is useful to have more than one person replicate screening for a portion of the search results.

The next step is reviewing the full text of individual studies, extracting key information (e.g., nutrient forms, biological endpoints, sample sizes, quantitative estimates of effect sizes and their associated uncertainty), and evaluating reliability (see Section 4.2.1.2). The extracted information may be saved into a spreadsheet, other database form, or annotated bibliography. A spreadsheet is useful for capturing quantitative data from the literature, such as effect sizes of stressor-response relationships and values of nutrient stressors, biological responses, and contextual variables, which could be used in a meta-analysis. To detect potential bias in what information from individual studies is extracted or potentially missed, it is useful to replicate data extraction for a portion of the literature. In addition, it is important to keep in mind that it is common for publications to omit results that are not statistically significant based on p-values, but this practice is a false binary test that misses the gradual notion of evidence supported by available data and could result in missing results with biological or ecological significance (Muff et al., 2022). Detecting bias related to statistical significance is one aspect of reliability and how results from the studies should be weighted (see Section 4.2). If data or reports are from potentially biased sources, it may be useful to conduct a sensitivity analysis to determine how much those sources contribute to the conclusions derived from literature-based evidence.

4.1.3 Source 3: Existing syntheses

Syntheses, literature reviews, and meta-analyses refer to published literature that analyzes and/or synthesizes results of a collection of individual studies. Syntheses are useful for identifying knowledge gaps and providing a scientific evidence base that bolsters or refines general, baseline knowledge of how

excess nutrients are expected to affect water bodies. Generally, existing literature syntheses are less numerous and more well-known (e.g., more cited) relative to the individual studies they summarize and analyze. Because of the relative rarity of syntheses, unbiased assembly of existing syntheses may be as simple as stating in planning documents whether they are in or out of scope of the NNC development effort. Analyzing data from multiple studies can be useful for increasing statistical power, which is often too low in individual studies to detect smaller but biologically important responses to nutrients, especially in field experiments prone to high variability (Francoeur, 2001). Examples of meta-analyses of studies examining nutrient effects on stream biota include Ardón et al. (2020), and Bennett et al. (2021). The evidence base associated with Bennett et al. (2021) along with additional biotic endpoints is available as an online database (www.epa.gov/ecodiver) that allows users to visualize and explore data from the literature while applying filters of interest (e.g., state, country, ecoregion).

4.1.4 Source 4: Expert knowledge

Expert knowledge can also be a source of evidence based on information that different stakeholders and partners may bring to the table. Subject matter experts may be selected based on their experience and contributions to the relevant scientific field, such as their publication record (e.g., U.S. EPA (2018)). Intentionally diverse panels or workgroups are also important for the unbiased assembly of expert knowledge, not only for obtaining information about the ecological system but to gain a more holistic understanding of all water body uses that need protection (Box 2). Unbiased assembly of expert knowledge also requires efforts to minimize conflicts of interest. Documenting experts' credentials is important for increasing transparency (e.g., subject matter expert biographies in U.S. EPA (2018)). Expert knowledge can be critical for understanding site-specific processes and land management histories that may contribute to unique conditions (e.g., fish species introductions or stocking, legacy nutrients, history of acidification, naturally colored waters attenuating light). Expert knowledge can be used to determine conceptual model pathways that are more likely to be important for a site or region and thus most crucial for assembling evidence. Methods for unbiased assembly of expert knowledge might include public calls for information, or independent peer review of the proposed process used for the project prior to its implementation, to strengthen confidence in the project's conclusions. For example, expert and stakeholder knowledge was assembled through extensive public comment opportunities during the development of inland nutrient criteria in Florida (U.S. EPA, 2010a).

Box 2. Indigenous Knowledge

Native American Tribes and Nations have been stewards of land and water resources since time immemorial. Natural resources and the environment play important roles in sustaining many aspects of traditional lifeways. Tribes can take on CWA authority for their Tribal lands, and a growing number of Tribes are working on setting standards, monitoring, assessing water quality, and developing goals to safeguard and restore water resources (see examples at https://mywaterway.epa.gov/state-and-tribal). Standards, including NNC, may be developed by Tribes in the same way as states to protect designated uses of water bodies, such as recreation, aquatic life, and drinking water. Tribes may also choose to protect waters designated for cultural uses. For example, the Minnesota Chippewa Tribe (Fond du Lac Band) has assessed lakes and reservoirs that support wild rice (Manoomin) areas and aesthetic waters, two categories of cultural use designations for water bodies that are significant to the preservation or exercise of the traditional value system of the Tribe (https://mywaterway.epa.gov/tribe/FONDULAC).

Indigenous knowledge has been federally recognized as an important source of information and a valid form of evidence to include and apply to research and decision-making, when it is appropriate and with the consent of the Tribe(s) involved (see Prabhakar and Mallory (2022) for an overview of understanding and applying indigenous knowledge). Integration of Indigenous Knowledge in environmental science and decision-making can enable a more holistic response to environmental impacts (U.S. EPA, 2011b). NNC development teams can consider collaboration with Tribal Nations and inclusion of indigenous knowledge in all phases and steps of the *Basic WoE Framework*. General examples and considerations for each phase include:

Planning: Tribes can be invited as collaborators or co-producers of knowledge. Early and prior consent from Tribal collaborators to participate in the process is valuable. It is important to plan how to have fair and meaningful engagement with Tribal collaborators. Even if the specific water bodies are not under Tribal jurisdiction, including Tribal collaborators in state NNC processes could result in mutual benefits (e.g., development of lessons learned that are applicable to additional water bodies and/or diverse water body types).

Problem formulation: Indigenous knowledge can provide holistic perspectives about the elements and connections among elements that should be included in conceptual models. This input may be critical for developing research questions, selecting and prioritizing endpoints, and informing the sampling design and data collection.

Analysis: Indigenous knowledge can include evidence acquired through direct contact with the environment and extensive observations passed down over generations. The use and dissemination of indigenous knowledge and data from Tribal lands and waters for any purpose should follow data sovereignty agreements with Tribal collaborators. Useful practices for indigenous data governance have been described as Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE) principles, which complement Findable, Accessible, Interoperable, and Reusable (FAIR) open science principles (Carroll et al., 2020). Metrics for judging relevance, strength, and reliability of evidence types derived from indigenous knowledge may need to be unique.

Criteria derivation: Opportunities for engagement, communication, review and/or input by tribes on how evidence will be integrated to derive criteria can lend credibility to the process.

With any project that requires a team effort, clear and well-reasoned plans for assembling evidence can enable consistency of methods, transparency, and accountability to achieve an unbiased body of evidence.

4.2 Weight evidence

Take home: Weighting evidence by establishing, objectively evaluating, and documenting qualities of that evidence shows how much influence individual evidence will have on overall NNC conclusions.

Once a body of evidence is assembled, individual pieces of evidence are weighted. If pieces of evidence differ in their weight, they exert different amounts of influence on NNC derivation. Three key qualities of a piece of evidence are relevance, reliability, and strength. Weighting involves evaluating evidence with respect to these qualities and assigning a qualitative or quantitative "score" that reflects the evaluation. In the following sections, we define these three qualities, give examples of how they might be judged, and discuss their application to primary data analyses, literature-based evidence, existing syntheses, and expert knowledge. We discuss methodological options for weighting, examples, and best practices for this step in the *Basic WoE Framework*.

4.2.1 Qualities of evidence

4.2.1.1 Relevance

Relevance is the degree to which a piece of evidence (e.g., an individual study, a particular stakeholder's knowledge) matches key conditions (such as type of water body, endpoints of interest, and environmental conditions at field sites), as well as the degree to which the evidence addresses other aspects of scope (e.g., management goal, designated use) laid out in the **Planning** and **Problem Formulation Phases**.

Questions to keep in mind when evaluating the relevance of evidence are:

- How closely does the analysis/study/knowledge coincide with abiotic conditions of waters for which NNC are being derived?
- How closely do the nutrient stressors and biological endpoints used in the analysis/study/knowledge align with those under consideration for NNC?

For primary data analyses, relevance will likely be high, especially for analyses/evidence generated from data specifically gathered for a state's NNC development, provided that the available data, associated analyses, and grouping of water bodies during the **Planning Phase** appropriately account for environmental variation. Literature-based evidence can also have high relevance even though it is less likely to be state-specific, especially in studies within an ecoregion of interest. Relevance of literature-based evidence can also be assured through clear inclusion and exclusion criteria during screening (see Section 4.1.2). Relevance of existing syntheses is determined by evaluating how well the original purpose and assembled evidence for the synthesis matches the NNC context.

Expert knowledge is likely to have broad relevance, while stakeholder knowledge might have high relevance about very specific places or aspects of designated use. Indigenous knowledge might be the only form of evidence relevant for understanding cultural uses (Box 2). User perception surveys, in which expert or non-expert users observe and report on factors they deem to be important for recreational use, are sometimes utilized and observations are correlated to measurements of nutrient

variables used in developing NNC (U.S. EPA, 2021b). The relevance of such surveys may depend on how well the characteristics of observed water bodies align with those under consideration for NNC development.

4.2.1.2 Reliability

Reliability is based on the inherent properties of evidence that make it convincing and is aligned with reproducibility. Reliability can depend on many aspects of experimental design, analysis, bias, and transparency (Frampton et al., 2022; Mupepele et al., 2016; Bilotta et al., 2014).

Questions to keep in mind when evaluating the reliability of evidence are:

- Are the data collection and analysis practices appropriate?
- Are confounding factors minimized?
- Are methods and results reported clearly and completely?

There are some general considerations when understanding the reliability of pieces of evidence (Figure 6). Pieces of evidence without quantitative data (e.g., individual expert opinion) are generally considered the least reliable, whereas the most reliable evidence is generally obtained from systematic reviews that combine data from multiple studies and are highly documented. It is not the case that the least reliable review is *always* more reliable than evidence generated from the single best reference/control or observational study. See below for additional information on judging specific reliability characteristics of reviews and primary data analyses. Furthermore, reliability can be increased by combining corroborating evidence, which is essentially what WoE is designed to do. For example, weighting and weighing multiple pieces of evidence from observational studies (i.e., studies in tier 3) within a line of evidence increases the reliability of that evidence (i.e., moves it to tier 2; Figure 6).

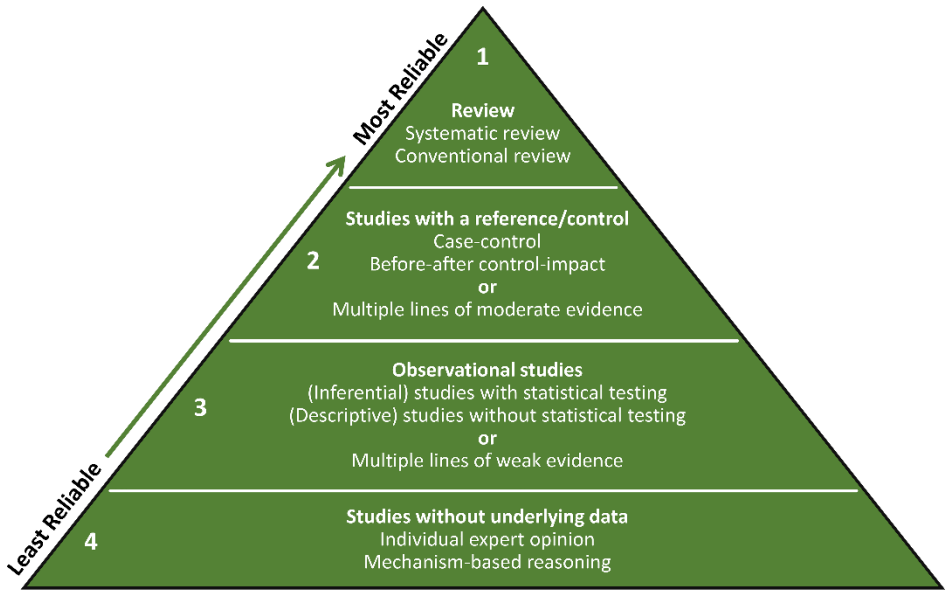


Figure 6. Evidence Reliability Pyramid

General considerations for judging the reliability of evidence. The bottom of the pyramid contains evidence that tends to be most plentiful but also the least reliable, while the top of the pyramid contains evidence that tends to be rare but most reliable. Modified from (Mupepele et al., 2016).

Reliability of evidence can also be evaluated by using additional, individual characteristics of an analysis, study, or stakeholder knowledge (Box 3). Many of these characteristics reflect aspects of experimental design (e.g., use of standard methods, minimization of confounding factors and other risks of bias), as well as the context of the evidence in the larger evidence base (e.g., corroboration, consistency, known modes of action). Not every characteristic will apply to every piece of evidence, but reliability of a piece of evidence is unlikely to be determined solely by one characteristic. Decisions about which characteristics are used to judge reliability should be transparent and justified.

Box 3. Characteristics That Contribute to Evidence Reliability

Taken from (U.S. EPA, 2016).

Design and execution: Evidence generated with a good study design that is well performed is more reliable.

Abundance: Evidence from more numerous data is more reliable, because it reflects greater replication or resolution.

Minimized confounding: Evidence is more reliable when the sampling design or analysis controls extraneous correlates.

Specificity: Evidence (e.g., a symptom or set of symptoms) specific to one cause or a few related causes is more reliable.

Potential for bias: Evidence from a study that is not funded by an interested party, is not produced for advocacy, and is not produced by an investigator with conflicts of interest is more reliable.

Standardization: A standard method decreases the likelihood that the evidence is biased or analyses are inaccurate.

Corroboration: Using models, indicators, or symptoms that have been verified by many studies and are accepted technical practice can greatly increase reliability.

Transparency: Complete descriptions of methods, inferential logic, and availability of data for reanalysis provide the means to check the results and are presumed to increase reliability by reducing the likelihood of hidden faults.

Peer review: Independent peer review of a study increases reliability of a source of information.

Consistency: The degree to which evidence does not vary in repeated instances within a study (e.g., across years, locations, sampling teams, or methods) is an indicator of reliability of a piece of evidence. When weighting types of evidence, consistency across studies of the same type is an indicator of reliability of the type.

Consilience: Evidence shown to be consistent with scientific knowledge and theory, particularly with respect to underlying mechanisms, is more reliable.

Additional considerations may be appropriate when evaluating the reliability of expert or stakeholder knowledge, including aspects of credibility and legitimacy like embeddedness in the scientific community (e.g., publication record), perception of bias, and the validity of past conclusions (Clark et al., 2002). Mechanisms to ensure reliability of expert opinion include establishing proper expert selection criteria, training of experts, discussing differences among experts, and presenting opinions in the context of other scientific data and observations.

While reviews tend to be the most reliable form of evidence (Figure 6), there is a lot of variation in review methodologies that can affect reliability. The Collaboration for Environmental Evidence (CEE) developed the CEE Synthesis Assessment Tool (CEESAT), which has multiple criteria for judging the reliability of reviews that that are self-identified as “systematic” (Box 4). These criteria are useful for evaluating the reliability of other types of existing syntheses (e.g., narrative literature reviews and meta-analyses), as well.

Box 4. Criteria for Judging the Reliability of Systematic Reviews

Taken from (CEE, 2022).

Review question:
Are the elements of the review question clear?

Protocol:
Is there an a-priori method/protocol document?

Searching:
Is the approach to searching clearly defined, systematic, and transparent?
Is the search comprehensive?

Including studies:
Are eligibility criteria clearly defined?
Are eligibility criteria consistently applied to all potentially relevant articles and studies found during the search?
Are eligibility decisions transparently reported?

Critical appraisal:
Does the review appraise each study?
During critical appraisal was an effort made to minimise subjectivity?

Data extraction:
Is the method of data extraction fully documented?
Are the extracted data reported for each study?
Were extracted data cross-checked by more than one reviewer?

Data Synthesis:
Is the choice of synthesis approach appropriate?
Is a statistical estimate of pooled effect (or similar) provided together with measures of variance and heterogeneity among studies?
Is variability in the study findings investigated and discussed?

Limitations:
Have the authors considered limitations of the synthesis?

4.2.1.3 Strength

Strength of evidence is the degree of differentiation between exposed or treated replicates from control conditions, reference conditions, or randomness. Strength is typically assessed through statistical parameters that communicate the magnitude, direction, association, or number of elements (Box 5). It is important to note that a piece of evidence can be strong and support a conclusion OR strong and refute a conclusion (the latter is sometimes referred to as “negative evidence”).

Questions to keep in mind when evaluating the strength of evidence are:

- What is the magnitude of the association?
- What is the direction of the association?

Box 5. Measurements of Evidence Strength

Modified from U.S. EPA (2016).

Magnitude: Commonly expressed as the effect size, difference between means, or a ratio of means.

Direction: Sign of an effect (i.e., positive or negative).

Association: Commonly expressed as a correlation coefficient or slope.

Number: The number of elements within a piece of evidence (e.g., of symptoms or overt effects in a response or of steps in a causal pathway) that are reported to be observed or the number of occurrences. This should not be confused with a candidate criterion value.

Evaluating the strength of primary data analyses and literature-based evidence is usually straightforward and based on the statistical results. The strength of syntheses and meta-analyses is usually expressed as an overall effect size if it is calculated (e.g., in a meta-analysis), which in turn depends on the strength of the individual studies included. Sometimes the number of studies within a synthesis that support or refute a hypothesis are tallied to reflect strength, but this type of "vote counting" is not universally accepted (CEE, 2022). Strength of expert and stakeholder knowledge may be difficult to determine, especially if it is not elicited with this type of evaluation in mind. If the opportunity exists to collect this source of evidence via surveys or focus groups, questions can be posed that elicit qualitative or quantitative expert and stakeholder opinions about, for example, the strength of association between a stressor and response. These data then can be analyzed using standard techniques such as in a probability distribution format or range (Burton et al., 2002).

4.2.2 Scoring and assigning weights

Weighting involves evaluating pieces of evidence with respect to relevance, reliability, and strength and assigning a score that reflects the evaluation. Scores for relevance, reliability, and strength are assessed independently—for example, a very relevant study could have a large effect size (high strength), but not have addressed important confounding effects (low reliability). Scoring is ideally as objective as possible with clear criteria for judgments determined *a priori*, but some subjectivity is likely unavoidable. Therefore, thorough documentation of the scoring process is advised.

Best practice for scoring evidence is to make scores symbolic, resulting in weights that are conceptual. There are examples of quantitative weighting of evidence in NNC development, but generally quantitative weighting implies a level of precision that is difficult to justify and so the decision to pursue this method should be carefully considered (Smith and Tran, 2010).

A variety of scoring schema and options for communication exist that are appropriate for the NNC development process. Scoring relevance can be done based on the questions and concepts presented in Section 4.2.1.1. Scores can span multiple categories, as long as the categories can be easily distinguished from one another. Scoring reliability can be done based on the general considerations in Figure 6 or a more thorough set of characteristics, as in Box 3. A hypothetical example of a reliability scoring table is provided in Figure 7, using reliability characteristics from Bennett et al. (2021).

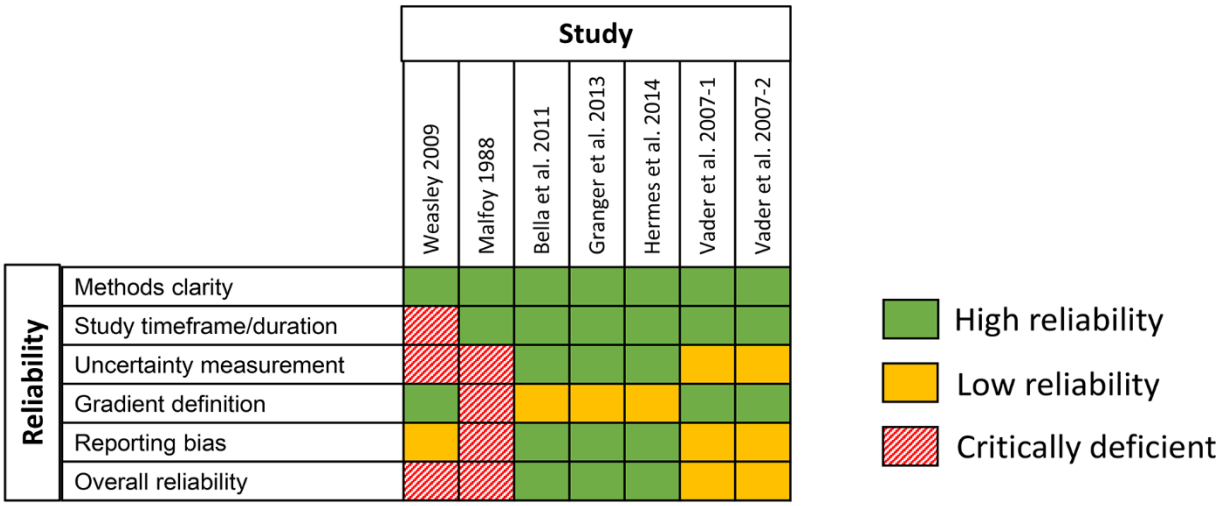


Figure 7. Communicating Evidence Reliability

An example of communicating the reliability of a set of literature-based evidence, where reliability is judged on five different factors. Each factor is judged as low or high reliability, with critically deficient assigned when evidence is severely flawed. Methods clarity = Clarity of the reported methods (not repeatable or repeatable); Study timeframe/duration = Study duration (single season or multiple seasons); Uncertainty measurement = Measurement of uncertainty (not reported or calculated and reported); Gradient definition = The gradient across which the stressor-response was measured (not planned or planned as part of the experimental design); Reporting bias = Completeness of reported results (incomplete reporting of results or all results reported regardless of statistical significance).

In Figure 7, individual pieces of literature-based evidence are scored as low reliability, high reliability, or critically deficient for each of the five factors and color-coded. Once these scores are assigned, an overall reliability score can be provided based on aggregation of these component scores. Note that in studies that score critically deficient for a minimum number of characteristics (it could be one or more than one), the whole study overall may be scored as critically deficient.

Strength of evidence can be evaluated and scored qualitatively and/or quantitatively, depending on the needs of the decision-maker and the amount and types of evidence available. In a meta-analysis context, individual studies are often weighted by the inverse of their variance. Table 2 illustrates a quantitative cutoff for correlation coefficients, calculated from regional field data examining the effect of major ions (and potential confounding factors) on invertebrate genera (U.S. EPA, 2011a). These cutoffs inform a qualitative, categorical weighting score based on the authors' expert judgement. In this example, a correlation coefficient (r) > 0.75 is considered relatively strong and studies that report high r values are given a ++ weighting score. Moderately strong evidence (+) are those studies with r values between 0.75 and 0.25. Weak r values receive a negative score (-), and those that refute the hypothesis (i.e., the correlation is in the opposite direction) are scored.

Table 2. Example of Weighting Evidence Strength

An example of weighting evidence strength using the absolute value of a correlation coefficient (*r*). Cutoff values of a correlation coefficient (or other statistical parameter) chosen to determine a score may vary across assessments. Taken from (U.S. EPA, 2016).

Assessment	Logical Implication and Strength	Score
The sign of the correlation coefficient depends on the relationship. For toxic relationships such as the correlation between conductivity and number of Ephemeroptera, the sign should be negative. Weak or positive correlations weaken the case for that candidate cause.	$\|r\| > 0.75$	+ +
	$0.75 \geq \|r\| \geq 0.25$	+
	$0.1 < \|r\| < 0.25$	0
	$\|r\| \leq 0.1$	-
	r has the wrong sign	- -

The values shown in Table 2 represent one weighting scheme for strength of evidence. The precise values chosen for these types of cutoffs may vary across assessments or be based on different statistical parameters (e.g., mean difference between impaired and reference sites). The scoring scale may also be specific to the situation. For example, instead of a scale ranging from ++ to --, it might range from +++ to 0. In all cases, scoring criteria should be determined and documented in advance to reduce bias within the weighting step and increase transparency and consistency.

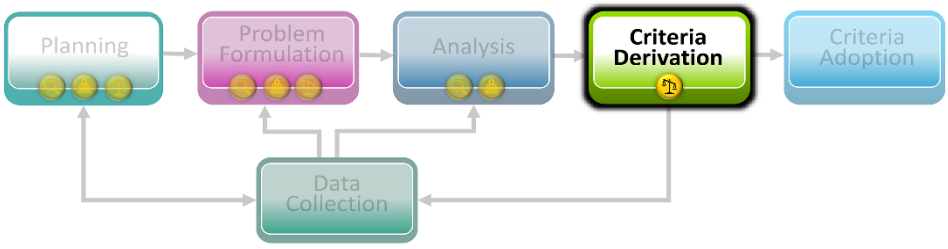
After being scored, relevance, reliability, and strength may be combined to produce an overall weight for each piece of evidence. An example of a complete evidence weighting table example is given in Table 3. Here, four pieces of evidence describe the relationship between diatoms and TP: (1) a diatom stressor-response analysis with field data collected within the state; (2) a diatom stressor-response analysis with field data collected outside the state; (3) a meta-analysis published in the literature; and (4) a mesocosm phosphorus dosing experiment published in the literature. Conceptual weights for relevance, reliability, and strength are assigned for each piece of evidence. Reasoning for these scores is clearly documented in the results section. For instance, in the diatom stressor-response analysis generated with data from outside the state, the environmental similarity to the state in question was close enough to pass the literature inclusion criteria and be scored as ++ for relevance; because the data were for a single season only, reliability was scored as 0. The meta-analysis, by contrast, was considered less relevant (being a synthesis over the entire US) but high study quality ("methods well documented and repeatable") resulted in a high score for reliability. Overall weights are derived from these independent scores of the three qualities, and the reasoning behind these weights are clear, consistent, and documented. Note that as in Figure 6, overall weights might not be as simple as adding up or "averaging" scores. For example, a piece of evidence with 0 reliability might result in an overall weight of 0, regardless of scores for relevance or strength.

Table 3. Summary Evidence Table Example

A hypothetical example of an evidence table communicating judgments of relevance, reliability, and strength for four pieces of evidence.

Piece of Evidence	Relevance (Rv)	Reliability (Rb)	Strength (St)	Overall Weight	Explanation
(1) TP-Diatom Index S-R curve: Analysis generated with state field data	+++	++	++	++	Field data from streams inside state (Rv) shows Index changepoint at TP=x mg/L with narrow CI (St); large sample size and wide nutrient gradient included (Rb).
(2) TP-Diatom Index S-R curve: Analysis generated with field data outside of state	++	0	+	0	Field data from streams outside state but with good environmental similarity (Rv) shows Index changepoint at TP=y mg/L with wide CI (St); single season data only (Rb).
(3) Meta-analysis of TP-Diatom richness relationship: Literature	+	+++	++	++	Meta-analysis of stream studies across the US (Rv) show a negative correlation between nutrient and biological endpoint for TP= >z mg/L (St); methods are well documented and repeatable (Rb).
(4) Mesocosm phosphorus dosing experiment: Literature	+	++	0	+	Experiment conducted in realistic stream mesocosm (Rv) shows no statistical change (St) in diatom richness with increasing doses of phosphorus; good sample size, reported experimental and analysis methods would be repeatable (Rb).
Integrated weight across all pieces of evidence	+/++	++	+/++	+/++	Evidence is largely consistent and weightiest for (1) and (3). The greatest uncertainty is the relation of mesocosm experiment to field exposures.

Once individual pieces of evidence are weighted, they are aggregated and integrated to arrive at a conclusion. Ultimately, the collective weight of an overall evidence base is a function of the weight of the individual sources and the way they were assembled, screened, and evaluated.



Criteria Derivation Phase

Take home: This phase includes weighing the body of evidence by integrating and interpreting evidence, as well as communicating conclusions. Methods for integrating evidence to derive criteria can range from simple to sophisticated; selected methods should be logical, informed by evidence availability and stakeholder needs, and communicated clearly.

5.1 Weigh body of evidence

The **Criteria Derivation Phase** aligns with weighing the body of evidence in the *Basic WoE Framework* (Figure 2). This part of the process begins by putting the pieces of evidence evaluated for relevance, reliability, and strength in the previous step into a form that facilitates integration. Evidence integration can take place in a single step where pieces of weighted evidence are integrated all at once. This generally works best when the evidence is of one type. However, evidence assembled for NNC tends to be diverse. Aggregating pieces of evidence into lines of evidence before integration may allow assessors to see patterns within and across evidence types, as well as facilitate communication with stakeholders.

5.1.1 Evidence aggregation

Pieces of evidence can be logically aggregated in more than one way (Figure 8). For instance, all of the new evidence generated from primary data analysis could be aggregated and integrated before integrating with literature-based evidence; stressor-response, reference condition, thresholds, and mechanistic modeling evidence could each be aggregated as separate lines of evidence before being integrated with each other. In a figure developed while deriving candidate criteria for TN and TP in headwater streams, Utah color-coded lines of evidence to show which were of a similar type (Utah DEQ, 2019). In addition, producing a flow chart can be helpful for showing aggregation steps (EFSA, 2017). For instance, Montana Department of Environmental Quality used a flow chart to communicate that evidence was aggregated into three lines before deriving TN and TP criteria protective of recreational use (Figure 9, (Suplee and Watson, 2013)).

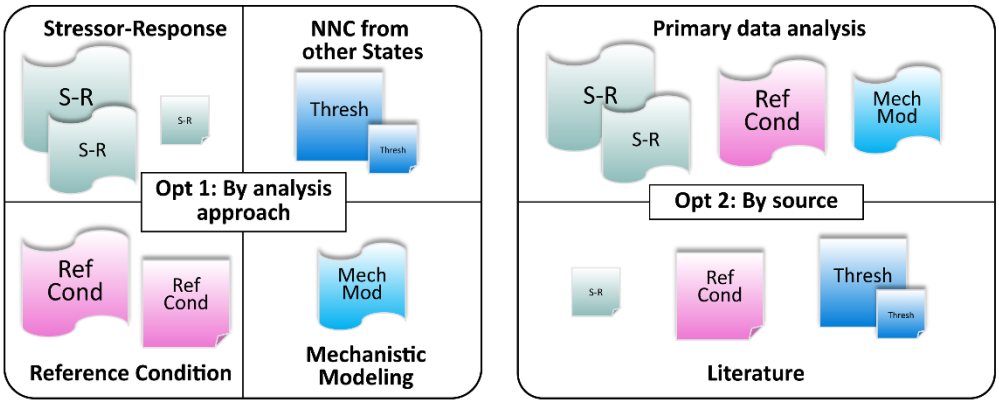


Figure 8. Options for Aggregating Pieces of Evidence into Lines of Evidence

Two options for aggregating pieces of evidence into lines of evidence. If grouping by analysis approach, the result is four lines of evidence (left). If grouping by evidence source, the result is two lines of evidence (right). Both options assume pieces of evidence have been assembled that represent multiple analysis approaches and sources, but not every combination needs to be present in real situations for aggregation to be useful. S-R=stressor-response, Ref Cond=reference condition, Thresh=Threshold NNC from other States, Mech Mod= Mechanistic Modeling.

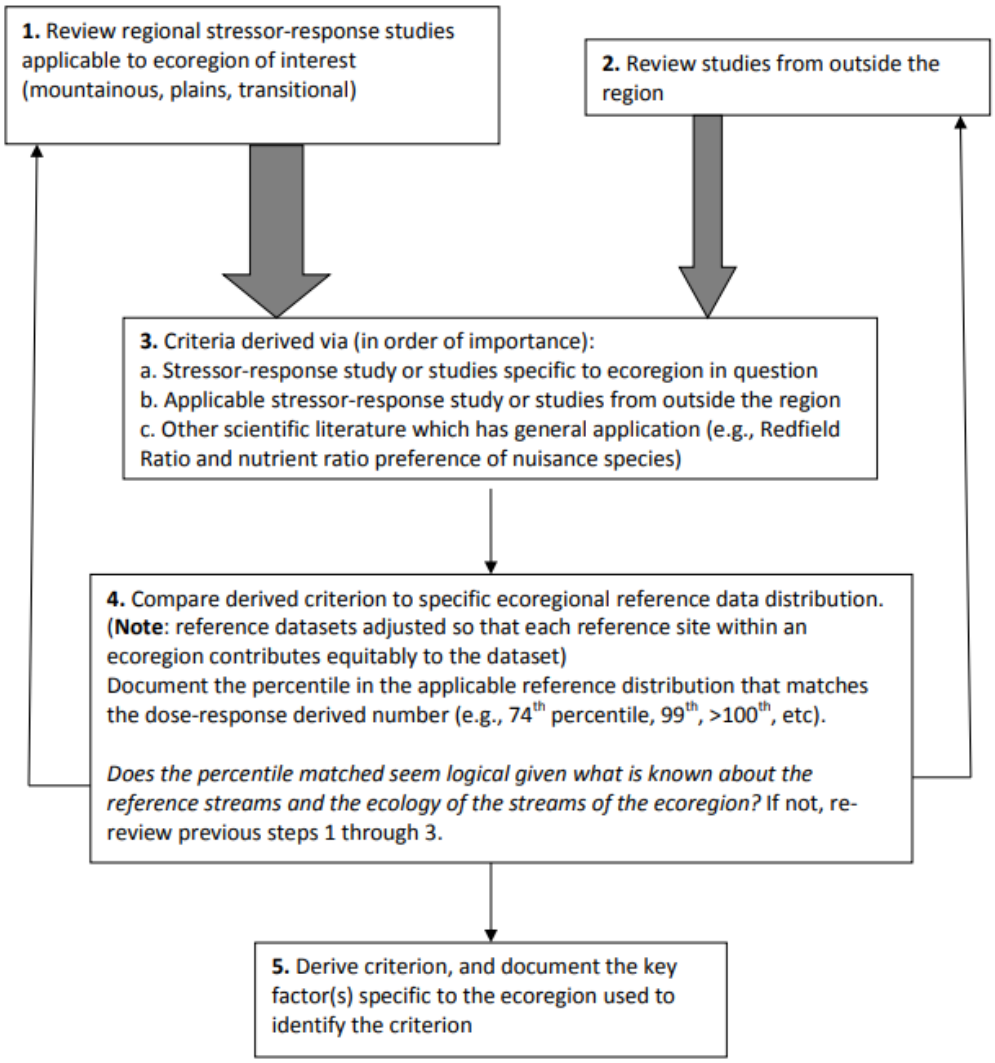


Figure 9. A Flow Chart Showing Evidence Aggregation

A flow chart showing how evidence from multiple sources was aggregated into three lines of evidence (see box in the chart labeled 3) before deriving TN and TP criteria for wadeable streams. Additionally, size of the arrows going into box 3 represent importance of the information sources. Taken from (Suplee and Watson, 2013).

5.1.2 Evidence integration

Evidence integration can take several forms. There may be only one line of evidence that has sufficient weight to inform the decision. This might occur if other lines of evidence under consideration are determined to be unacceptably weak in one or more areas of relevance, reliability, and/or strength. It is also possible that more than one line of evidence has sufficient weight to inform the decision, but ultimately the weightiest is chosen that best protects the designated use (see Table 4). U.S. EPA's guidance and models for the development of nutrient criteria in lakes creates a path for developing a single relevant, strong, and reliable line of evidence that informs numeric criteria (U.S. EPA, 2021a, 2020a, b). Even using a single line of evidence, challenges like uncertainty can arise, which can be overcome with appropriate strategies (see challenges section, below) (U.S. EPA, 2021a, 2020a, b).

Table 4. Example of Evidence Integration by Selecting Weightiest Evidence

A hypothetical example showing integration of evidence from Table 3 by selecting the weightiest.

Piece of Evidence	Overall Weight	Candidate criterion	Explanation
(1) TP-Diatom Index S-R curve: Analysis generated with State field data	++	TP=x mg/L with narrow CI	Primary data analysis that has resulted in weightiest evidence.
(2) TP-Diatom Index S-R curve: Analysis generated with field data outside of State	0	TP=y mg/L with wide CI	Similar primary data analysis as in (1), but underlying data represent only a single season, so evidence has unacceptably low reliability. Working with neighboring state to include multiple seasons in future analyses.
(3) Meta-analysis of TP-Diatom richness relationship: Literature	++	TP=α mg/L	Threshold identified, but endpoint is not sensitive to nutrient change at low TP concentrations, resulting in substantial uncertainty around candidate criterion.
(4) Mesocosm phosphorus dosing experiment: Literature	+	TP=β-γ mg/L	TP candidate criterion range identified from low and medium dosage, but not statistically different from high dosage concentration. β<x<γ.
Conclusion	++	TP=x mg/L	(1) selected to inform final criterion value. (4) is not strong, but corroborates value derived from (1).

Evidence integration can also take the form of merging multiple lines of evidence with sufficient weight. Merging evidence has led to successful **Criteria Adoption** in a number of instances, including for lakes and rivers in Minnesota (Heiskary et al., 2013; Heiskary and Wilson, 2005).

When multiple lines of evidence are merged, noting the characteristics of the evidence base as a whole can help convey the level of confidence in conclusions. For example, bodies of evidence with many pieces or lines of evidence from diverse sources that are free of bias and logically coherent lend confidence in the conclusions drawn from those bodies of evidence (U.S. EPA, 2016; Norton et al., 2015).

The methods for deriving a numeric criterion by merging multiple lines of evidence range from very simple to sophisticated. Regardless of which method is chosen, decisions about which to use should be logical and communicated clearly. On the simpler end, merging can be done using arithmetic and geometric means. Means are appropriate to use if all lines of evidence are weighted equally, the exposures are to the same stressor, and the endpoint entity and attribute are the same. This was done to establish a numeric chlorophyll *a* criterion for the Chesapeake Bay to prevent the human health effects of harmful algal blooms; the two lines of evidence that were averaged were an existing threshold established by the WHO and a stressor-response analysis using data from the Chesapeake Bay (U.S. EPA, 2007).

Weighted means have also been used where lines of evidence are unequal in weight. For example, based on biological changepoints and cluster analysis, a weighted mean was calculated to recommend a TP criterion for large rivers in New York (Smith and Tran, 2010). In this case, endpoint and attribute were

not the same for each line of evidence; careful consideration should be given to whether a designated use is adequately protected when the mean of multiple endpoints (some which may be more sensitive than others) is calculated.

A median could also be calculated to merge lines of evidence. For example, multiple lines of evidence were developed for setting nutrient criteria in Minnesota rivers. Reference condition evidence was weighted less heavily, and threshold values were weighted more heavily. The recommended criterion for TP was approximately the median across all lines of evidence for the Northern River Nutrient Region (Heiskary et al., 2013).

An optional approach outlined by U.S. EPA includes an additional element to develop NNC that integrate causal (nitrogen and phosphorus) and response parameters into one water quality standard (U.S. EPA, 2013). Criteria developed with this combined criteria approach allow for consideration of both nutrient level, duration, and frequency *and* an appropriate response level, duration, and frequency in determining when a designated use is met. Notably, if there is a sufficient understanding of the response parameter's relevance to management goals, measurability, ecological relevance, sensitivity to nutrients, and importance to stakeholders, then fully incorporating it into the criteria development process from the beginning would be possible.

More sophisticated methods for merging multiple lines of evidence include models used in meta-analysis, Bayesian network models (e.g., Carriger et al. (2016)), and multi-criteria decision analyses (e.g., Linkov et al. (2011)). To produce valid results, these methods require inputs of sufficient and uniform evidence, as well as quantitative expertise for conducting them and interpreting their results. Caution is advised in presuming more sophisticated methods will always lead to more precise or justified NNC, especially when assumptions and uncertainties surrounding the structure or other aspects of these merging methods are not clear or transparent.

If scoping and **Problem Formulation** are carefully done, there should be a low chance that *no* lines of evidence have sufficient weight to inform derivation of a criterion. However, if that situation does arise, reviewing the reason(s) can help strategically inform collection of additional evidence. For instance, if the environmental relevance of available evidence was insufficient, resources could be targeted at data collection and evidence development for more environmentally similar sites.

Other challenges can arise late in the process of criteria development. Many can be avoided by careful and deliberate scoping and **Problem Formulation**, and none need to derail criteria derivation. For example:

- Uncertainty- This is an incomplete understanding of a state or true value.[6] Uncertainty may be quantitative (e.g., the standard error = *x*) or qualitative (e.g., the amount of uncertainty is unacceptably high to stakeholders). Quantified, statistical uncertainty surrounding pieces of evidence generated from primary data analysis may be reduced by increasing sample sizes by searching for additional existing, reliable data or by collecting new, reliable data (see Section 4.1.1). Reducing qualitative uncertainty may involve developing new pieces or lines of evidence to address the specific concerns of experts and/or stakeholders.

[6] We do not attempt a full accounting of all the possible ways uncertainty may be defined and measured. For one collaborative effort of this type see https://dictionary.helmholtz-uq.de/content/landing_page.html#.

- Variability- This is inherent heterogeneity of data. To address variability in a piece or line of evidence, it might be necessary to include additional co-variates during analysis and/or more finely subdivide the unit of analysis (e.g., via more specific water body groupings in the **Planning Phase**).
- Ambiguity- This is when evidence has no clear meaning or more than one possible meaning. To address evidence ambiguity, an independent expert review of the evidence could be utilized. Additionally, one can acknowledge when evidence has more than one plausible interpretation and be transparent about which ultimately informs the decision.
- Discrepancy- This is an inconsistency in the evidence base in which evidence implies different answers. It is important to understand which discrepancies could have a logical basis (e.g., based on knowledge of biology or the particular analyses performed) and which do not. Discrepancies can lead to a critical examination of the underlying reliability of evidence and potential exclusion of evidence that is found to be too weak. Lacking an explanation for discrepancies, follow-up studies could be designed that target their resolution.

Documenting and clearly communicating the process of integrating evidence and how the conclusions (i.e., derived criteria) are supported by the evidence are core principles of the *Basic WoE Framework*. Quantitative derivation of a criterion should be accompanied by interpretation, explanation, and description of any outstanding ambiguities or uncertainties. When available, uncertainty may be expressed statistically as a range and/or probability of possible conclusions (i.e., criteria) (EFSA, 2017). Utilizing tables of evidence can help communicate conclusions. Tables that might be included at the **Criteria Derivation Phase** complement those presented earlier to communicate evidence weights (Table 5). The Minnesota Pollution Control Agency has used evidence tables to communicate both site-specific and water-body category NNCs (e.g., Heiskary et al. (2013); Heiskary and Wasley (2011)). In each example shown here, candidate criteria from multiple lines of evidence are listed separately alongside final proposed numbers. Evidence from Smith and Tran (2010) was compiled into a table format for this report and shows how each TP estimate was weighted before being combined using a weighted average to propose the criterion for large rivers in New York (Table 6).

Table 5. Examples of Compiling Evidence and Conclusions into Tables

Evidence utilized in proposing NNC for (a) Lake Pepin and (b) rivers in the Southern River Nutrient Region of Minnesota are compiled into tables modified from (Heiskary et al., 2013; Heiskary and Wasley, 2011). Tables such as these support the core principles of transparency, documentation, and communication in the *Basic WoE Framework*.

(a)

	2002 303(d) listing[1]	Recent 10-year mean[2]	2009 means	Criteria & goal ranges[3]	Diatom-inferred P from c.1900-1960[4]
TP µg/L	198	171	152	80-120	~110-140
Chl-a µg/L	25	30	32	28-120	--

[1] 1991-2000
[2] 2000-2009
[3] Represents draft values discussed or proposed at various points in overall process.
[4] Estimate #1 (Engstrom and Almendinger 2000)

(b)

Line of Evidence	TP (µg/L)	Chl-a (µg/L)	DO Flux (mg/L)	BOD_5 (mg/L)
25th %ile Threshold Concentrations (Table X)	145	21*	3.1*	3.1
IQR for Minimally impacted MN streams (Table X)	185-320		--	2.4-6.1
IQR for USEPA Ecoregion Summaries (Table X)	170-403	--	--	--
75th %ile for MN Reference Sites (Table X)	302	19	--	--
Predicted Concentration Using TP-Chla-BOD_5 Threshold Models (Figure X)	129-149	28-39	--	--
Predicted Concentration Using TP-BOD_5 Threshold Models (Figure X)	168-193	--	--	--
Predicted Concentration Using 75th %ile Water Quality Models (Table X)	--	36-39	4.8	2.5-2.7
Recommended Criterion (Table X)	150	35	4.5	3.0

*Indicates threshold is based on statewide data.
IQR= Interquartile Range
%ile= Percentile
TP= Total Phosphorus
Chl-a= Chlorophyll *a*
DO Flux= Diel Dissolved Oxygen Flux
BOD_5= Biochemical Oxygen Demand

Table 6. Example of Compiling Weighted Evidence and Conclusions into a Table

Evidence from Smith and Tran (2010) was compiled for this report into a table format to show how each TP estimate was weighted before being merged via a weighted average, resulting in a proposed TP criterion.

Line of Evidence	Indicator	Weight	TP (mg/L)	Weight x TP
Stressor-Response	NBI-P invertebrates	2	0.011	0.022
Stressor-Response	% mesotrophic diatoms	2	0.009	0.018
Stressor-Response	% eutrophic diatoms	2	0.020	0.040
Stressor-Response	BAP invertebrates	2	0.070	0.140
Cluster Analysis	Invertebrate Medium Group Median	1.5	0.037	0.056
Cluster Analysis	Diatom Medium Group Median	1.5	0.037	0.056
Reference	Median of Two Reference Estimates	1	0.023	0.023
	Sum	12		0.354
	Weighted Average (mg/L)			0.030

NBI-P= Nutrient Biotic Index for TP
BAP= Biological Assessment Profile
Medium Group= Cluster of biologically similar sites determined to have moderate nutrient concentrations

Figures can also be used to help communicate conclusions. There are many good general references about what makes an effective figure when displaying scientific information (e.g., Rougier et al. (2014)). Several recent examples demonstrate how figures can be used to display evidence and proposed NNC. For instance, the Utah Department of Environmental Quality shows the ranges of all individual lines of evidence and proposed TN and TP criteria for headwater streams in a single graphic (Figure 10).

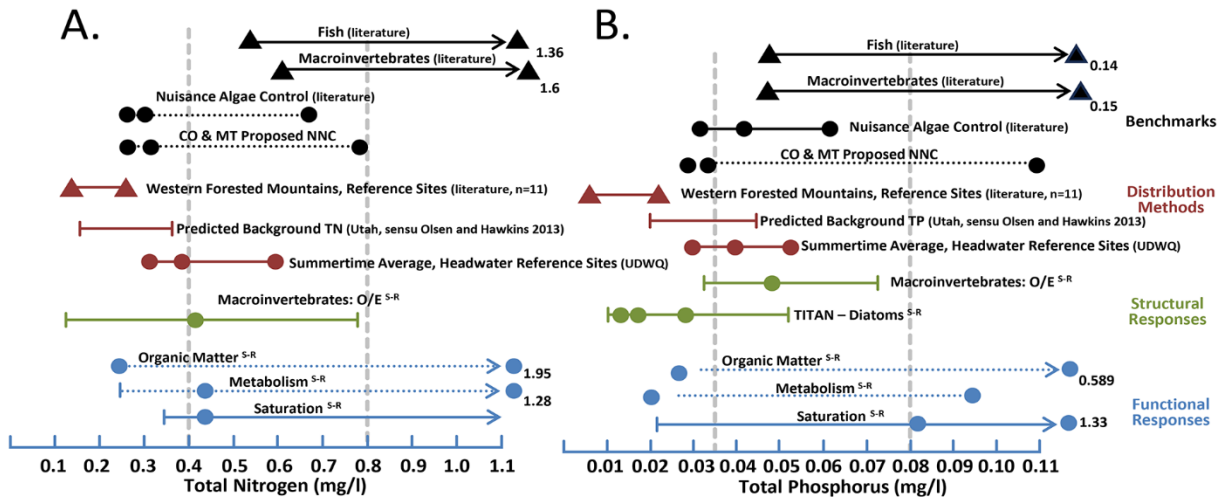


Figure 10. Example of Compiling Lines of Evidence and Conclusions into a Figure

A depiction of multiple lines of evidence assembled for the derivation of TN and TP numeric criteria for Utah headwater streams. Lines of evidence are labeled and shown as horizontal lines. Proposed numeric criteria are shown in relation to the lines of evidence as vertical grey dashed lines. A figure such as this supports the core principles of transparency, documentation, and communication in the *Basic WoE Framework*. Taken from (Utah DEQ, 2019).

The New Mexico Environment Department, with the support of Tetra Tech through the N-STEPS program, prepared evidence summaries for both stressor-response and reference distribution approaches to develop TN and TP numeric criteria (Tetra Tech (2015); Figure 11).

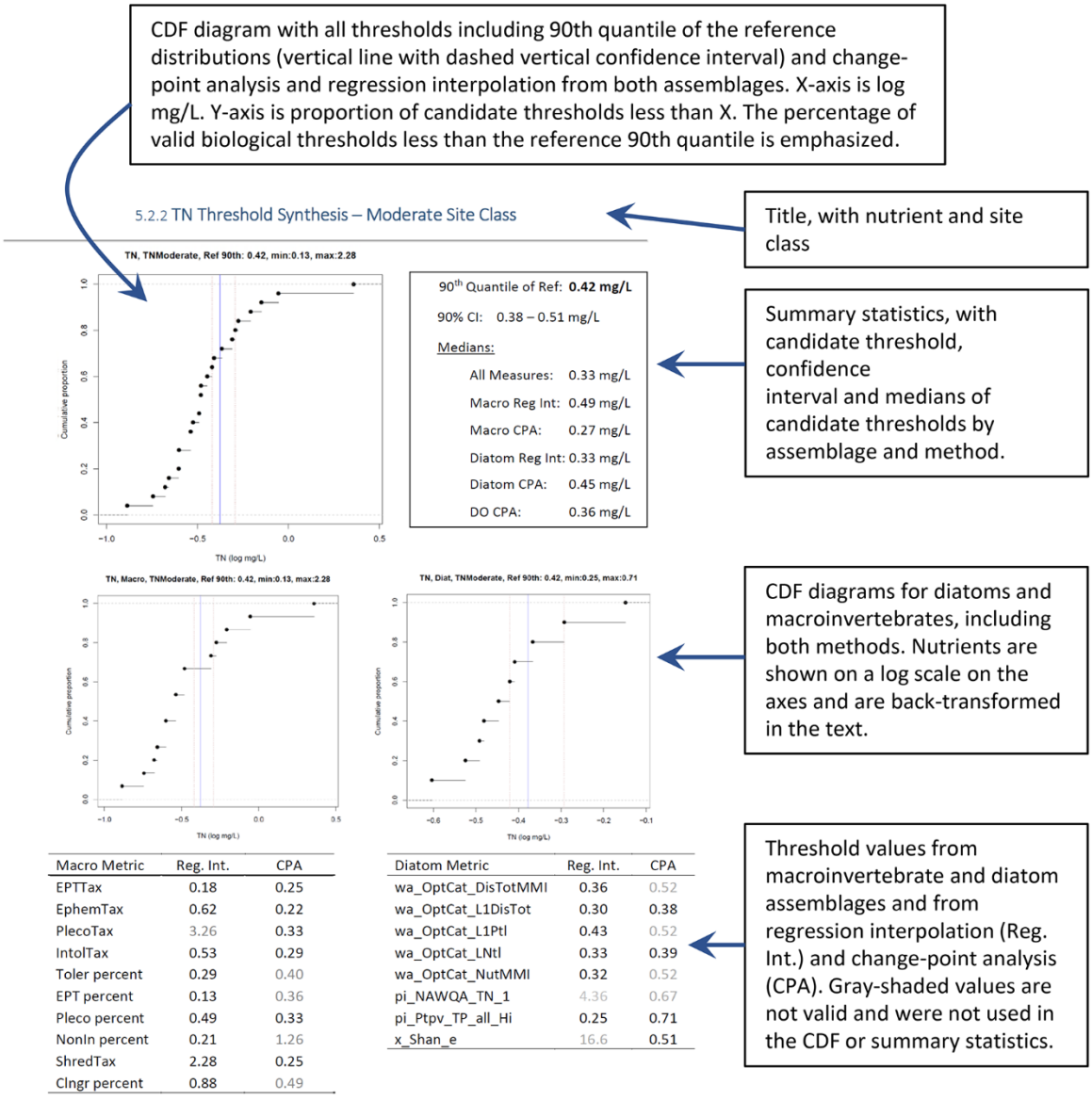


Figure 11. Example of Compiling Evidence with Interpretation into a Figure

Legend provided for a series of one-page evidence summaries used in developing TN and TP numeric criteria for New Mexico perennial wadeable streams. This legend aids in interpretation of the evidence summaries that follow in the document with the guiding text in the accompanying boxes. Each one-page summary shows multiple lines of evidence, including stressor-response based endpoints relevant to reference values and tabular values of the results for individual endpoint stressor-response analyses by method. The document also includes additional explanation and interpretation of each line of evidence with the pros and cons of each. Modified from Tetra Tech (2015).

Other innovative ways of visualizing pieces and lines of evidence that contribute to the development of criteria or benchmarks have also been proposed. Hall et al. (2017) suggest a plot to visually communicate relevance and reliability of evidence, grouped within levels of biological organization (Figure 12). While such a plot was envisioned for ecotoxicological evidence, some of its features are adaptable to show evidence and proposed criteria or benchmarks related to nutrients.

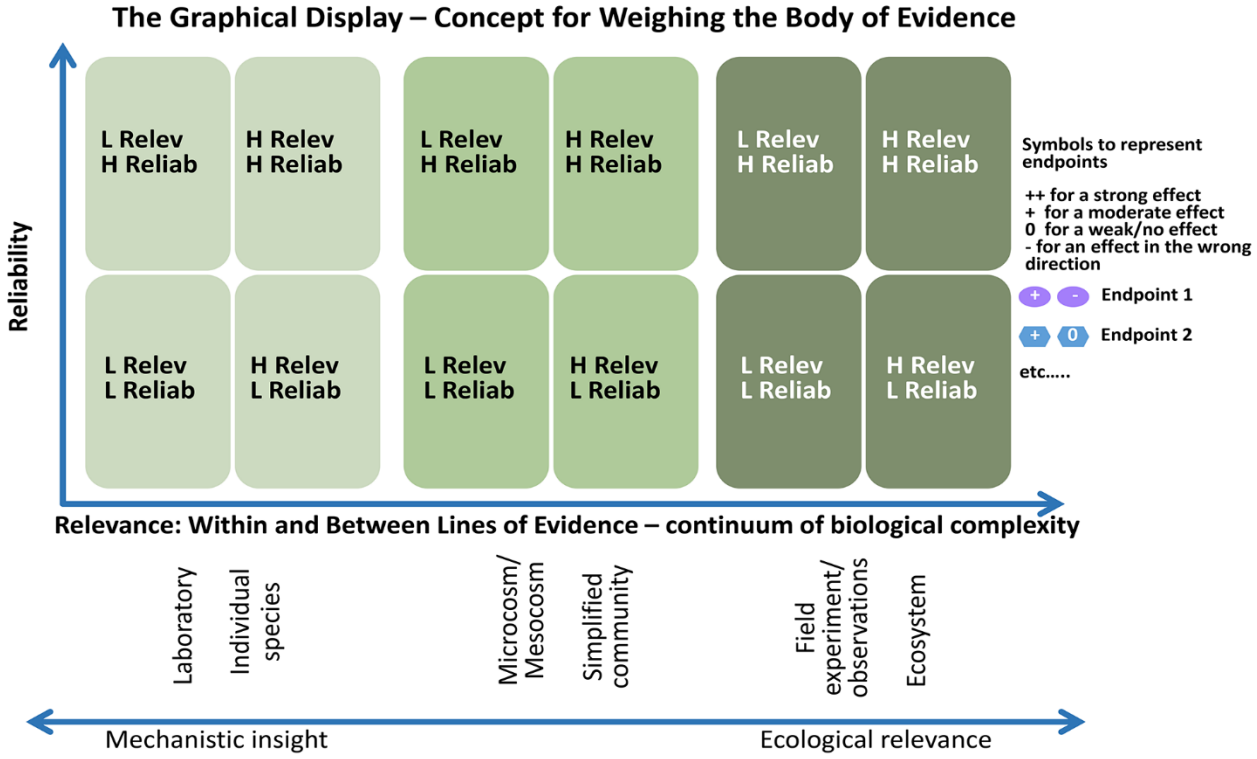


Figure 12. Innovative Example for Compiling Evidence into a Figure

A potential way to present multiple lines of evidence in a visual format. The figure provides a way to incorporate scale (small-scale evidence to the left, large-scale evidence to the right), relevance and reliability (by positioning evidence within a 2x2 table for each scale of evidence), strength (by using +, 0, and -), and different endpoints (shapes). Symbols to represent strength and examples of the continuum of biological complexity were modified from Hall et al. (2017) to show the figure's application to NNC evidence more clearly.

Derivation of criteria is usually the culmination of multi-year efforts involving technical aspects of derivation and weighing the body of evidence conducted by the same people who conducted the earlier phases of criteria development. As a predecessor to the next phase, **Criteria Adoption**, there may be wider interest and scrutiny on the technical team's methods and conclusions. This wider interest can be acknowledged and addressed through a technical review and/or public comment process. Additionally, at this phase it may be helpful to maintain a distinction between the technical team who have assembled, weighted, and weighed evidence and the decision-makers who use it to ultimately set an NNC. In a risk assessment paradigm, establishing responsibilities of risk assessors (i.e., the technical team) and risk managers (i.e., the decision-makers) lends additional transparency to the criteria development process.

Completion of the NNC development process is typically marked by detailed documentation and reporting of each phase of the process. The following are examples of best practices for documenting NNC development:

- Documentation begins at the **Planning Phase**. A technical plan can provide transparency and a benchmark for how the team intends the process to go. When things change, the team can specify how and why in relation to the plan.
- The best documentation enables reproducibility. Just like the "Materials and Methods" section of a journal article, good documentation enables someone not involved in the work to repeat the process and generate similar results, including the application of WoE methods.
- Documentation can be simplified by using checklists and templates. These tools can facilitate recording and communicating the complete details of the process.

Applications

The previous chapters of this report describe the reasons for using WoE methods and options that exist for each phase of the criteria development process. Methods are mentioned that have been successfully applied by states in the past when developing NNC or that appear in the scientific literature. Most of the examples are discrete; they demonstrate what an appropriate method is in isolation from other parts of the criteria development process. Thus, it can be challenging to fully appreciate the larger context and rationale for selecting WoE methods when that context is not apparent.

In this chapter, we complement examples in Chapters 2-5 with two more complete examples based on a range of real situations and challenges that states may be facing in developing NNC. We assembled "profiles" of each state that include details of factors affecting criteria development (Appendix A). A summary of these factors is presented in Table 7. State "A" and State "C" differ in the categories of water bodies for which they are developing criteria; how close they are to **Criteria Adoption**; their access to relevant primary data and capacity to analyze those data; the availability of literature-based evidence; their capacity to conduct new studies to fill evidence gaps; the lines of evidence they are or are interested in developing; and their proposed evidence integration methods. The contrasting situations of these states result in different WoE methods that would be appropriate. In the following sections, we suggest WoE methods that could be appropriate given the contrasting time, resource, evidence, and decision constraints experienced by these two states.

Table 7. Summary of State A and State C

A summary of factors relevant to the phases of the NNC development process for two real but anonymous example states, State A and State C. The table was current at the time of interviews and information collection but may not reflect the current status of NNC development within these states. S-R= Stressor-Response

Phase	**Planning and Problem Formulation**	**Analysis**				**Criteria Derivation**	
Factors	Water Body Type and Time to Criteria Adoption	Availability of Relevant Primary Data	Capacity to Analyze Primary Data	Availability of Evidence from Published Literature	Capacity to Conduct New Studies	Lines of Evidence Used	Evidence Integration Method
State A	Inland waters: <1 year	High	Not Limited	Low/Medium	High	Literature S-R Reference Condition	Mean
State C	Lakes and streams: 3-5 years	Medium	Limited	Medium/High	Medium	TBD, but considering Literature S-R Reference Condition	TBD

6.1 How might State A and State C conduct planning and problem formulation using WoE methods?

A state like State C that is relatively early in the process for developing NNC for lakes and streams has an opportunity to plan and shape their work using WoE methods nearly from the beginning. Because there are resource constraints arising from having a small staff, external expertise could assist a state like State C in reviewing and strengthening plans developed during the **Planning** and **Problem Formulation Phases**. External expertise could also be leveraged for helping to group water bodies and select endpoints using WoE methods (as well as leveraged later in the **Analysis** and **Criteria Derivation Phases**). On the other hand, a state like State A that is much further along in developing NNC for its inland waters could ensure its original planning process, water body grouping approach, and endpoint selection process are documented in written form with as much detail as possible to enhance transparency. One key detail to bring out in that documentation would be that State A decided to undertake development of criteria for multiple water body types at the same time, yet each set of assembled evidence and the process to use the evidence for criteria development was distinct.

6.2 How might State A and State C assemble evidence using WoE methods?

For both State A and State C, it is important to make sure there is complete documentation of data collection, compilation, and clean up (e.g., for taxonomic consistency), as well as statistical methods and other decisions made about including or excluding certain data points or datasets. Potential sources of unavoidable bias or information gaps should be acknowledged. A state like State A with abundant primary data and capacity to analyze primary data can explore multiple endpoints and statistical approaches or models. For State C, it could be worthwhile to incorporate data from publicly available national datasets or the literature to augment evidence from the state's own datasets.

In the case of State C, with 3-5 years until **Criteria Adoption** and the potential for a relatively large amount of relevant evidence in the literature, it could be worthwhile to take a rigorous and systematic approach to build a strong literature-based evidence base and conduct a quantitative meta-analysis. For states with situations more similar to State A, with less than 1 year remaining until **Criteria Adoption** and not a large amount of relevant published literature, it may make sense to describe and justify a more qualitative approach to narratively review key studies. When there is limited time to find key studies, it could be helpful to focus on existing literature syntheses (e.g., by reviewing their reference lists and summarized findings) or rely on expert recommendations.

6.3 How might State A and State C weight evidence using WoE methods?

A state like State A that relies largely on abundant primary data and the capacity to analyze it to create evidence should focus on documenting the qualities of that evidence. It would be appropriate to clarify aspects of relevance, reliability, and strength that led the technical team to equally weight the lines of evidence produced through the reference condition and stressor-response approaches that ultimately informed criteria derivation. If a state like State C decides to emphasize literature-based evidence in their criteria development process, creating and applying clear relevance, reliability, and strength metrics for that source of evidence is recommended. See Sections 4.2.1.1, 4.2.1.2, and 4.2.1.3 for

questions that could be utilized in determining relevance, reliability, and strength of literature-based evidence. In addition, utilizing a visual communication tool (see Figure 7) could enhance transparency of weighting judgments.

6.4 How might State A and State C aggregate evidence using WoE methods?

Again, a state like State A that is further into the process of criteria development likely already has a sense of which pieces of evidence are suitable for aggregation. In this situation, it is important to clearly document how pieces of evidence are grouped into lines of evidence (e.g., stressor-response, reference condition). In addition, it is appropriate to describe within each line of evidence how many pieces of evidence there are, and whether they tend to be coherent. On the other hand, a state like State C that is earlier in the process of criteria development may not have determined whether or how to aggregate evidence. Because of the interest in and access to multiple sources of evidence (e.g., primary data analyses and literature-based evidence) and multiple analysis approaches (e.g., stressor-response, reference condition), it is likely that some sort of aggregation will be appropriate.

6.5 How might State A and State C integrate evidence using WoE methods?

States like State A that arrive at the **Criteria Derivation Phase** with more than one line of evidence with sufficient weight to influence their conclusions will have options for how to merge those lines of evidence. State A has several alternatives and is exploring merging evidence using a mean. This method appears justified and easy to communicate with stakeholders. States like State C will need to decide how to integrate evidence once it is all assembled and weighted. If only one line of evidence has sufficient weight, it will determine the criterion. If more than one line of evidence has sufficient weight, the State may choose to select the best line of evidence or merge multiple lines of evidence. In both a State A and State C situation, characteristics of the body of evidence as a whole (number of lines, diversity of evidence, bias, and coherence) can be described and inform the level of confidence in conclusions.

Conclusions

Weight-of-evidence (WoE) is a process in which evidence is assembled, evaluated, and integrated to make a technical inference in an assessment. The following are take-home messages for the role WoE can play in strengthening each phase of NNC development.

Planning Phase- Activities undertaken during **Planning** provide a transparent foundation for developing NNC; transparency is a core principle of WoE. Grouping water bodies during **Planning** is a process to which WoE could be applied when diverse evidence needs to be combined.

Problem Formulation Phase- Selecting endpoints during **Problem Formulation** is also a process to which WoE could be applied when diverse evidence needs to be combined. Conceptual models developed during **Problem Formulation** can help inform what evidence should be assembled in the **Analysis Phase**.

Analysis Phase- This phase includes assembling evidence and weighting evidence. Unbiased assembly of evidence is best practice and can ensure NNC are based on transparent data and information of sufficient amount and quality. Weighting evidence by establishing, objectively evaluating, and documenting qualities of that evidence shows how much influence individual evidence will have on overall NNC conclusions.

Criteria Derivation Phase- This phase includes weighing the body of evidence by integrating and interpreting evidence, as well as communicating conclusions. Methods for integrating evidence to derive criteria can range from simple to sophisticated; selected methods should be logical, informed by evidence availability and stakeholder needs, and communicated clearly.

Overall, there is a set of intended outcomes when WoE methods are applied to NNC development (Table 8). Those outcomes occur during the process of criteria development but are ultimately achieved through improved water quality and the protection of designated uses.

Table 8. Summary of Suggested Practices and Intended Outcomes

This table summarizes suggestions for how to carry out WoE at different phases of criteria development and what can be achieved.

Criteria Development Phase	Basic WoE Framework Element	Key Suggested Practices	Intended Outcomes
Planning	Core principle	Planning is transparent, documented, and leverages collective expertise.	Decision-makers and stakeholders understand and trust the criteria development process. Planning minimizes bias, is realistic, and meets stakeholder needs.
	Assemble, weight, weigh	WoE methods are used to group water bodies.	When **Criteria Derivation Phase** is reached, candidate criteria for each water body grouping have acceptably low amounts of variation.
Problem Formulation	Assemble, weight, weigh	WoE methods are used to select endpoints.	Endpoints are relevant to management goals, measurable, ecologically relevant, sensitive to nutrients, and important to stakeholders.
Analysis	Assemble	Evidence is assembled in an unbiased way.	Conclusions reached in the **Criteria Derivation Phase** are objective and defensible because they are based on evidence that accurately and fairly represents what is known about nutrients and their effects in water bodies.
	Weight	Weighting criteria are established ahead of time; relevance, strength, and reliability of evidence are assessed and documented.	Each piece of evidence has influence on the conclusions in the **Criteria Derivation Phase** that appropriately corresponds to its objectively evaluated relevance, strength, and reliability.
	Core principle	Processes for assembling and weighting evidence are documented and communicated clearly.	Decision-makers and stakeholders understand the pieces of evidence that make up the body of evidence and how they influence conclusions in the **Criteria Derivation Phase**.
Criteria Derivation	Weigh	If necessary, evidence is logically aggregated. Integration method is appropriate for the evidence.	Derived criteria are sound and defensible, because the method to either (a) select the weightiest evidence or (b) merge multiple lines of sufficiently weighty evidence is technically appropriate and justified to protect the designated use.
	Core principle	Conclusions are clearly communicated.	Decision-makers and stakeholders understand and trust the derived criteria.

References

Alers-García, J; Lee, SS; Spaulding, SA. (2021). Resources and Practices to Improve Diatom Data Quality. Limnol Oceanogr Bull 30: 48-53. http://dx.doi.org/10.1002/lob.10433

Ardón, M; Zeglin, LH; Utz, RM; Cooper, SD; Dodds, WK; Bixby, RJ; Burdett, AS; Follstad Shah, J; Griffiths, NA; Harms, TK; Johnson, SL; Jones, JB; Kominoski, JS; Mcdowell, WH; Rosemond, AD; Trentman, MT; Van Horn, D; Ward, A. (2020). Experimental nitrogen and phosphorus enrichment stimulates multiple trophic levels of algal and detrital-based food webs: A global meta-analysis from streams and rivers. Biol Rev Camb Philos Soc 96: 692-715. http://dx.doi.org/10.1111/brv.12673

Arkansas DEQ. (2012). State of Arkansas nutrient criteria development plan. North Little Rock, AR: Arkansas Department of Environmental Quality, Water Division – Planning Branch. https://www.adeq.state.ar.us/water/planning/pdfs/ar_nutrient_plan_update.pdf

Babitsch, D; Berger, E; Sundermann, A. (2021). Linking environmental with biological data: Low sampling frequencies of chemical pollutants and nutrients in rivers reduce the reliability of model results. Sci Total Environ 772: 145498. http://dx.doi.org/10.1016/j.scitotenv.2021.145498

Bennett, MG; Lee, SS; Schofield, KA; Ridley, CE; Washington, BJ; Gibbs, DA. (2021). Response of chlorophyll a to total nitrogen and total phosphorus concentrations in lotic ecosystems: A systematic review [Review]. Environ Evid 10: 23. http://dx.doi.org/10.1186/s13750-021-00238-8

Bilotta, GS; Milner, AM; Boyd, IL. (2014). Quality assessment tools for evidence from environmental science. Environ Evid 3. http://dx.doi.org/10.1186/2047-2382-3-14

Burton, GA; Chapman, PM; Smith, EP. (2002). Weight-of-evidence approaches for assessing ecosystem impairment. Hum Ecol Risk Assess 8: 1657-1673. http://dx.doi.org/10.1080/20028091057547

Carpenter, SR; Caraco, NF; Correll, DL; Howarth, RW; Sharpley, AN; Smith, VH. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol Appl 8: 559-568. http://dx.doi.org/10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2

Carriger, JF; Barron, MG; Newman, MC. (2016). Bayesian Networks Improve Causal Environmental Assessments for Evidence-Based Policy. Environ Sci Technol 50: 13195-13205. http://dx.doi.org/10.1021/acs.est.6b03220

Carroll, SR; Garba, I; Figueroa-Rodríguez, OL; Holbrook, J; Lovett, R; Materechera, S; Parsons, M; Raseroka, K; Rodriguez-Lonebear, D; Rowe, R; Sara, R; Walker, JD; Anderson, J; Hudson, M. (2020). The CARE principles for Indigenous data governance. Data Sci J 19: 43. http://dx.doi.org/10.5334/dsj-2020-043

CEE. (2022). Collaboration for Environmental Evidence Synthesis Appraisal Tool (CEESAT) (Version 5.1).

Clark, JA; Hoekstra, JM; Boersma, PD; Kareiva, P. (2002). Improving U.S. Endangered Species Act recovery plans: Key findings and recommendations of the SCB recovery plan project. Conserv Biol 16: 1510-1519. http://dx.doi.org/10.1046/j.1523-1739.2002.01376.x

Coffey, DB; Cormier, SM; Harwood, J. (2014). Using field-based species sensitivity distributions to infer multiple causes. Hum Ecol Risk Assess 20: 402-432. http://dx.doi.org/10.1080/10807039.2013.767071

Cook, SC; Housley, L; Back, JA; King, RS. (2018). Freshwater eutrophication drives sharp reductions in temporal beta diversity. Ecology 99: 47-56. http://dx.doi.org/10.1002/ecy.2069

Cormier, SM; Paul, JF; Spehar, RL; Shaw-Allen, P; Berry, WJ; Suter, GW, 2nd. (2008). Using field data and weight of evidence to develop water quality criteria. Integr Environ Assess Manag 4: 490-504. http://dx.doi.org/10.1897/IEAM_2008-018.1

EFSA. (2017). Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J 15: 1-69. http://dx.doi.org/10.2903/j.efsa.2017.4971

Frampton, G; Whaley, P; Bennett, M; Bilotta, G; Dorne, JCM; Eales, J; James, K; Kohl, C; Land, M; Livoreil, B; Makowski, D; Muchiri, E; Petrokofsky, G; Randall, N; Schofield, K. (2022). Principles and framework for assessing the risk of bias for studies included in comparative quantitative environmental systematic reviews. Environ Evid 11. http://dx.doi.org/10.1186/s13750-022-00264-0

Francoeur, SN. (2001). Meta-analysis of lotic nutrient amendment experiments: Detecting and quantifying subtle responses. J North Am Benthol Soc 20: 358-368. http://dx.doi.org/10.2307/1468034

Hall, AT; Belanger, SE; Guiney, PD; Galay-Burgos, M; Maack, G; Stubblefield, W; Martin, O. (2017). New approach to weight-of-evidence assessment of ecotoxicological effects in regulatory decision-making. Integr Environ Assess Manag 13: 573-579. http://dx.doi.org/10.1002/ieam.1936

Heiskary, S; Bouchard, RW, Jr; Markus, H. (2013). Minnesota nutrient criteria development for rivers (draft). St. Paul, MN: Minnesota Pollution Control Agency. https://www.pca.state.mn.us/sites/default/files/wq-s6-08.pdf

Heiskary, SA; Wasley, D. (2011). Lake Pepin site specific eutrophication criteria. (WQ-S6-10). St. Paul, MN: Minnesota Pollution Control Agency. https://www.pca.state.mn.us/sites/default/files/wq-s6-10.pdf

Heiskary, SA; Wilson, CB. (2005). Minnesota lake water quality assessment report: Developing nutrient criteria (3rd ed.). (WQ-LAR3-01). St. Paul, MN: Minnesota Pollution Control Agency. https://www.pca.state.mn.us/sites/default/files/wq-lar3-01.pdf

Hilton, J; O'Hare, M; Bowes, MJ; Jones, JI. (2006). How green is my river? A new paradigm of eutrophication in rivers [Review]. Sci Total Environ 365: 66-83. http://dx.doi.org/10.1016/j.scitotenv.2006.02.055

Kumar, S; Spaulding, SA; Stohlgren, TJ; Hermann, KA; Schmidt, TS; Bahls, LL. (2009). Potential habitat distribution for the freshwater diatom Didymosphenia geminata in the continental US. Front Ecol Environ 7: 415-420. http://dx.doi.org/10.1890/080054

Lee, SS; Bishop, IW; Spaulding, SA; Mitchell, RM; Yuan, LL. (2019). Taxonomic harmonization may reveal a stronger association between diatom assemblages and total phosphorus in large datasets. Ecol Indicat 102: 166-174. http://dx.doi.org/10.1016/j.ecolind.2019.01.061

Linkov, I; Welle, P; Loney, D; Tkachuk, A; Canis, L; Kim, JB; Bridges, T. (2011). Use of multicriteria decision analysis to support weight of evidence evaluation. Risk Anal 31: 1211-1225. http://dx.doi.org/10.1111/j.1539-6924.2011.01585.x

Martin, P; Bladier, C; Meek, B; Bruyere, O; Feinblatt, E; Touvier, M; Watier, L; Makowski, D. (2018). Weight of Evidence for Hazard Identification: A Critical Review of the Literature [Review]. Environ Health Perspect 126: 076001. http://dx.doi.org/10.1289/EHP3067

Merrell, K; Lee, SS. (2022). NE Lakes Sediment Diatom Collaboration: Regional Collaboration and Perseverance. LakeLine. Newsletter of the North American Lake Management Society. Summer 2022: 10-14.

Muff, S; Nilsen, EB; O'Hara, RB; Nater, CR. (2022). Rewriting results sections in the language of evidence. Trends Ecol Evol 37: 203-210. http://dx.doi.org/10.1016/j.tree.2021.10.009

Munn, MD; Frey, JW; Tesoriero, AJ; Black, RW; Duff, JH; Lee, K; Maret, TR; Mebane, CA; Waite, IR; Zelt, RB. (2018). Understanding the influence of nutrients on stream ecosystems in agricultural landscapes. (U.S. Geological Survey Circular 1437). Reston, VA: U.S. Geological Survey. http://dx.doi.org/10.3133/cir1437

Mupepele, AC; Walsh, JC; Sutherland, WJ; Dormann, CF. (2016). An evidence assessment tool for ecosystem services and conservation studies. 26: 1295–1301. http://dx.doi.org/10.1890/15-0595

Norton, SB; Cormier, SM; Suter, GW, II. (2015). Ecological causal assessment. Boca Raton, FL: CRC Press.
Paerl, HW; Scott, JT; Mccarthy, MJ; Newell, SE; Gardner, WS; Havens, KE; Hoffman, DK; Wilhelm, SW; Wurtsbaugh, WA. (2016). It takes two to tango: When and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. Environ Sci Technol 50: 10805-10813. http://dx.doi.org/10.1021/acs.est.6b02575
Potapova, MG; Lee, SS; Spaulding, SA; Schulte, NO. (2022). A harmonized dataset of sediment diatoms from hundreds of lakes in the northeastern United States. Scientific Data 9: 540. http://dx.doi.org/10.1038/s41597-022-01661-3
Prabhakar, A; Mallory, B. (2022). Guidance for federal departments and agencies on Indigenous knowledge. Washington, DC: Executive Office of the President, Office of Science and Technology Policy, Council on Environmental Quality. https://www.whitehouse.gov/wp-content/uploads/2022/12/OSTP-CEQ-IK-Guidance.pdf
Riato, L; Hill, RA; Herlihy, AT; Peck, DV; Kaufmann, PR; Stoddard, JL; Paulsen, SG. (2022). Genus-level, trait-based multimetric diatom indices for assessing the ecological condition of rivers and streams across the conterminous United States. Ecol Indicat 141: 1-13. http://dx.doi.org/10.1016/j.ecolind.2022.109131
Rougier, NP; Droettboom, M; Bourne, PE. (2014). Ten simple rules for better figures [Editorial]. PLoS Comput Biol 10: e1003833. http://dx.doi.org/10.1371/journal.pcbi.1003833
Ryan, DF. (2021). Biological responses to stream nutrients: A synthesis of science from experimental forests and ranges. (PNW-GTR-981). Portland, OR: U.S. Department of Agriculture, Forest Service, Pacific Northwest Research Station. https://www.fs.usda.gov/research/treesearch/63549
Salafsky, N; Boshoven, J; Burivalova, Z; Dubois, NS; Gomez, A; Johnson, A; Lee, A; Margoluis, R; Morrison, J; Muir, M; Pratt, SC; Pullin, AS; Salzer, D; Stewart, A; Sutherland, WJ; Wordley, CFR. (2019). Defining and using evidence in conservation practice. 1. http://dx.doi.org/10.1111/csp2.27
Schulte, N; Lee, S; Spaulding, S; Stevenson, J; Edlund, M; Heinlein, J; Burge, D; Vaccarino, M. (2021). Application of diatom multi-metric indices to state monitoring data. https://diatoms.org/news/webinar-multi-metrics-diatoms
Smith, AJ; Tran, CP. (2010). A weight-of-evidence approach to define nutrient criteria protective of aquatic life in large rivers. J North Am Benthol Soc 29: 875-891. http://dx.doi.org/10.1899/09-076.1
Smucker, NJ; Pilgrim, EM; Wu, H; Nietch, CT; Darling, JA; Molina, M; Johnson, BR; Yuan, LL. (2022). Characterizing temporal variability in streams supports nutrient indicator development using diatom and bacterial DNA metabarcoding. Sci Total Environ 835: 154960. http://dx.doi.org/10.1016/j.scitotenv.2022.154960
Suplee, MW; Watson, V. (2013). Scientific and technical basis of the numeric nutrient criteria for Montana's wadeable streams and rivers: Update 1. (WQPBWQSTR-002). Helena, MT: Montana Department of Environmental Quality. https://deq.mt.gov/files/water/wqpb/standards/pdf/sciencetech2013fnlcom.pdf
Suter, G; Cormier, S; Barron, M. (2017a). A weight of evidence framework for environmental assessments: Inferring qualities. Integr Environ Assess Manag 13: 1038-1044. http://dx.doi.org/10.1002/ieam.1954
Suter, G; Cormier, S; Barron, M. (2017b). A weight of evidence framework for environmental assessments: Inferring quantities. Integr Environ Assess Manag 13: 1045-1051. http://dx.doi.org/10.1002/ieam.1953

Suter, G; Nichols, J; Lavoie, E; Cormier, S. (2020). Systematic review and weight of evidence are integral to ecological and human health assessments: They need an integrated framework. Integr Environ Assess Manag 16: 718-728. http://dx.doi.org/10.1002/ieam.4271

Suter, GW, 2nd; Cormier, SM. (2008). What is meant by risk-based environmental quality criteria? Integr Environ Assess Manag 4: 486-489. http://dx.doi.org/10.1897/IEAM_2008-017.1

Suter, GW; Cormier, SM. (2016). Bias in the development of health and ecological assessments and potential solutions. Hum Ecol Risk Assess 22: 99-115. http://dx.doi.org/10.1080/10807039.2015.1056062

Taylor, JM; Back, JA; Brooks, BW; King, RS. (2018). Spatial, temporal and experimental: Three study design cornerstones for establishing defensible numeric criteria in freshwater ecosystems. J Appl Ecol 55: 2114-2123. http://dx.doi.org/10.1111/1365-2664.13150

Tetra Tech. (2015). New Mexico nutrient thresholds for perennial wadeable streams: Final report. https://www-archive.env.nm.gov/wp-content/uploads/sites/25/2019/04/NSTEPS-NM_Nut_Crit_ReportFINAL_20150821.pdf

Tyree, MA; Bishop, IW; Hawkins, CP; Mitchell, R; Spaulding, SA. (2020a). Reduction of taxonomic bias in diatom species data. Limnology and Oceanography: Methods 18: 271-279. http://dx.doi.org/10.1002/lom3.10350

Tyree, MA; Carlisle, DM; Spaulding, SA. (2020b). Diatom enumeration method influences biological assessments of southeastern USA streams. Freshw Sci 39: 183-195. http://dx.doi.org/10.1086/707725

U.S. EPA. (1992). Framework for ecological risk assessment [EPA Report]. (EPA/630/R-92/001). Washington, DC. http://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=30004UKJ.txt

U.S. EPA. (1998). Guidelines for ecological risk assessment [EPA Report]. (EPA/630/R-95/002F). Washington, DC: U.S. Environmental Protection Agency, Risk Assessment Forum. https://www.epa.gov/risk/guidelines-ecological-risk-assessment

U.S. EPA. (2000a). Ambient water quality criteria recommendations. Information supporting the development of state and tribal nutrient criteria: Rivers and streams in nutrient Ecoregion IX [EPA Report] (pp. 150). (822-B-00-019). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=20003EHB.txt

U.S. EPA. (2000b). Ambient water quality criteria recommendations. Information supporting the development of state and tribal nutrient criteria: Rivers and streams in nutrient Ecoregion XI [EPA Report]. (EPA-822-B-00-020). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=20003ELH.txt

U.S. EPA. (2000c). Nutrient criteria technical guidance manual: rivers and streams [EPA Report]. (EPA/822/B-00/002). Washington, DC: U.S. Environmental Protection Agency, Office of Water. http://water.epa.gov/scitech/swguidance/standards/criteria/nutrients/rivers/index.cfm

U.S. EPA. (2001a). Ambient water quality criteria recommendations. Information supporting the development of state and tribal nutrient criteria: Rivers and streams in nutrient Ecoregion VIII [EPA Report]. (EPA-822-B-01-015). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=20003G83.txt

U.S. EPA. (2001b). Ambient water quality criteria recommendations. Information supporting the development of state and tribal nutrient criteria: Rivers and streams in nutrient Ecoregion X [EPA Report]. (EPA-822-B-01-016). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=20009MZI.txt

U.S. EPA. (2006). Framework for developing suspended and bedded sediment (SABS) water quality criteria [EPA Report]. (EPA-822-R-06-001). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=164423

U.S. EPA. (2007). Ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll a for the Chesapeake Bay and its tidal tributaries: 2007 chlorophyll criteria addendum [EPA Report]. (EPA/903/R-07/005). U.S. Environmental Protection Agency :: U.S. EPA. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100BTPK.txt

U.S. EPA. (2010a). Technical support document for U.S. EPA's final rule for numeric criteria for nitrogen/phosphorus pollution in Florida's inland surface fresh waters. https://web.archive.org/web/20110107120426/http://water.epa.gov/lawsregs/rulesregs/upload/floridatsd1.pdf

U.S. EPA. (2010b). Using stressor-response relationships to derive numeric nutrient criteria [EPA Report]. (EPA-820-S-10-001). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100IK1N.txt

U.S. EPA. (2011a). A field-based aquatic life benchmark for conductivity in Central Appalachian streams (Final Report). (EPA/600/R-10/023F). Cincinnati, OH: U.S. Environmental Protection Agency, National Center for Environmental Assessment. http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=233809

U.S. EPA. (2011b). Integration of traditional ecological knowledge (TEK) in environmental science, policy and decision-making. https://www.epa.gov/sites/default/files/2017-03/documents/tsc_tribal-ecological-knowledge-env-sci-policy-dm.pdf

U.S. EPA. (2013). Guiding principles on an optional approach for developing and implementing a numeric nutrient criterion that integrates causal and response parameters [EPA Report]. (EPA-820-F-13-039). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100KO37.txt

U.S. EPA. (2015). Preamble to the Integrated Science Assessments [EPA Report]. (EPA/600/R-15/067). Research Triangle Park, NC: U.S. Environmental Protection Agency, Office of Research and Development, National Center for Environmental Assessment, RTP Division. https://cfpub.epa.gov/ncea/isa/recordisplay.cfm?deid=310244

U.S. EPA. (2016). Weight of evidence in ecological assessment [EPA Report]. (EPA/100/R-16/001). Washington, DC: Office of the Science Advisor. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100SFXR.txt

U.S. EPA. (2017a). Causal Analysis/Diagnosis Decision Information System (CADDIS): Nutrients. https://www.epa.gov/caddis-vol2/nutrients

U.S. EPA. (2017b). National water quality inventory: Report to Congress. (EPA/841/R-16/011). U.S. Environmental Protection Agency :: U.S. EPA. https://www.epa.gov/sites/default/files/2017-12/documents/305brtc_finalowow_08302017.pdf

U.S. EPA. (2018). Mapping the vulnerability of human health to extreme heat in the United States (final report). (EPA/600/R-18/212F).

U.S. EPA. (2019). National Aquatic Resource Surveys: Rivers and streams 2013-2014 (data and metadata files). Retrieved from https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

U.S. EPA. (2020a). Nutrient Scientific Technical Exchange Partnership & Support (N-STEPS) Online: Chlorophyll criteria based on zooplankton. Available online at https://nsteps.epa.gov/apps/chl-zooplankton/ (accessed October 19, 2022).
U.S. EPA. (2020b). Nutrient Scientific Technical Exchange Partnership & Support (N-STEPS) Online: Hypoxia model. Available online at https://nsteps.epa.gov/apps/chl-hypoxia/ (accessed October 19, 2022).
U.S. EPA. (2020c). Nutrient Scientific Technical Exchange Partnership & Support (N-STEPS) Online: Nutrient - Chlorophyll Models. Available online at https://nsteps.epa.gov/apps/tp-tn-chl/
U.S. EPA. (2021a). Ambient water quality criteria to address nutrient pollution in lakes and reservoirs [EPA Report]. (EPA-822-R-21-005). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P1012YNU.txt
U.S. EPA. (2021b). Development of user perception surveys to protect water quality from nutrient pollution: A primer on common practices and insights [EPA Report]. (EPA 823-R-21-001). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P1012A1R.txt
U.S. EPA. (2022a). National lakes assessment: The third collaborative survey of lakes in the United States [EPA Report]. (EPA 841-R-22-002). Washington, DC: U.S. Environmental Protection Agency, Office of Water and Office of Research and Development. https://nationallakesassessment.epa.gov/webreport
U.S. EPA. (2022b). Nutrient Scientific Technical Exchange Partnership & Support (N-STEPS) Online. Available online at https://nsteps.epa.gov/
U.S. EPA. (2022c). State progress toward developing numeric nutrient water quality criteria for nitrogen and phosphorus. Available online at https://www.epa.gov/nutrient-policy-data/state-progress-toward-developing-numeric-nutrient-water-quality-criteria (accessed August 26, 2022).
Utah DEQ. (2019). Technical support document: Utah's nutrient strategy.
Virginia DEQ. (2019). Recommended numeric chlorophyll-a criteria for the James River Estuary.
WHO & FAO. (2009). Principles and methods for the risk assessment of chemicals in food [WHO EHC]. (EHC 240). Geneva, Switzerland: World Health Organization. https://www.who.int/publications/i/item/9789241572408
Yuan, LL; Smucker, NJ; Nietch, CT; Pilgrim, EM. (2022). Quantifying spatial and temporal relationships between diatoms and nutrients in streams strengthens evidence of nutrient effects from monitoring data. Freshw Sci 41: 100-112. http://dx.doi.org/10.1086/718631

Appendix A

The purpose of this appendix is to provide detailed profiles of two states involved in developing nutrient criteria. The profiles of these states were created to summarize the data and information available to them in making decisions related to numeric nutrient criteria (NNC) and their decision-making context. They differed in the timing of criteria development, extent of data and analytical ability, and familiarity with weight-of-evidence (WoE) concepts. In this way, they provide content important to informing the development of research tools and translational science to support future nutrient criteria development efforts.

Two profiles were constructed using a standard format to aid in comparability. Each profile is organized by factors coming into play during the 1) **Planning** and **Problem Formulation Phases** (decision and timetable); 2) **Analysis Phase** (types of data, analysis capacity, evidence from literature, capacity to commission new studies, and weighting evidence); and 3) **Criteria Derivation Phase** (criteria development method). Understanding the amount and types of available evidence, the resources and capacities of states, and overall factors like decision timelines are important for understanding which WoE methods are appropriate and feasible. The following information was gleaned through informational calls with each state, researching each state's nutrient water quality websites, and personal knowledge. Each state's name and information are anonymized.

The details described in these profiles represent a snapshot in time. It is expected that as the criteria development process advances, these details will change. However, as other states across the country undertake NNC development, they may find themselves in similar circumstances to the states profiled here. Therefore, an opportunity exists to learn how WoE methods could be applied in these common circumstances.

A.1 State A profile

A.1.1 State A – Streams and rivers

State A is close to completing the development of numeric criteria for a subset of inland waters and in early development of criteria for coastal waters.

A.1.2 Factors affecting the Planning and Problem Formulation Phases

A.1.2.1 Decision and timetable

State A is in the process of completing nutrient criteria development for multiple types of inland waters (streams, rivers, small impoundments, and some wetlands). The state is developing criteria to protect designated uses in three classes of inland waters that include inland flowing, some wetland and small impounded fresh surface waters; these designated uses include protection of aquatic life, recreation, and drinking water quality. While criteria development involves multiple water body types, we focus most of the details of the profile on streams and rivers.

The state has focused on a range of assessment endpoints to protect the aforementioned management goals, including dissolved oxygen, pH, aquatic life condition indicators, adverse microbial growths, transparency, chlorophyll *a*, and nuisance algal cover. At the time of discussion, they were at the point

of public engagement and eliciting feedback. The state has been in the process of planning, collecting data, developing analysis tools and analyzing data, synthesizing the data, and developing recommended criteria for more than 10 years.

This state did not produce a planning document per se. They formed a steering committee (composed of water quality standards (WQS) staff, biologists, engineers, and water quality managers) that worked collaboratively to develop a plan, although it was not formally written upfront. However, this process resulted in the general approach detailed in the Technical Support Document that underlies their criteria. This process included **Problem Formulation** and conceptual modeling, identification and addressing data gaps, selection of endpoints, analysis approaches, approaches for weighting and weighing evidence to derive values, and formulation of the decision framework. The ideas developed during planning were also vetted with the Regional EPA nutrient coordinator.

Clearly, consideration of available information and data was an important part of their planning and **Problem Formulation**, and the final information and data used are discussed below. They are now approximately a year away from adopting inland criteria which would then be sent to USEPA for approval. The state has begun preliminary work on coastal criteria for a single water body along their coast. They are proposing potential adoption of these essentially water body-specific criteria in 2022/2023 and using that as a demonstration to continue criteria development for the rest of their coastal waters.

A.1.3 Factors affecting the Analysis Phase

The types of relevant data, capacity to analyze data, evidence from published literature, capacity to conduct or commission new studies, and how evidence is weighted are all important elements of the **Analysis Phase**. These affect how evidence will be weighed for criteria derivation.

A.1.3.1 Types of relevant primary data

State A had a high amount of relevant primary data. They have a well-established, long-term monitoring program that collected abundant nutrient and response (assessment endpoint) data. In addition, they conducted several special studies focused on, for example, specific endpoints and filling gaps in their reference dataset. State A also invested in the development of unique indicators (e.g., algal indices) that informed their decision-making. They relied heavily on their state monitoring data for analysis.

The state prefers to use its own data primarily but is not against also utilizing data from nearby states or others that have similar streams and land-use types to provide a robust sample population. For context, the USEPA National Aquatic Resource Surveys program (U.S. EPA, 2019) has collected around 200 samples for rivers and streams for State A and there are an additional 169 samples in adjacent states for the range of endpoints collected by that program.

They did rely heavily on outside information from other states including user perception work by one other state and one other country with environmentally similar streams and rivers. In addition, they mentioned the utility of criteria development discussions with adjacent states, facilitated through an interstate organization that supports such meetings and interactions. Lastly, they stated their criteria development process also benefitted from national nutrient meetings including USEPA OST/HECD Nutrient Scientific Exchange and Partnership Support program (U.S. EPA, 2022) meetings, interactions with regional coordinators and the N-STEPS program; and the online N-STEPS Q&A content, which provided expert answers to many of their and others' questions.

A.1.3.2 Capacity to analyze primary data

State A was not limited in their capacity to analyze primary data in support of their analysis for nutrient criteria development. They have a small, but well-trained population of scientists and engineers who can conduct a wide range of statistical modeling and advanced technical tool development. They did, however, benefit from code (e.g., R packages) developed, demonstrated, and provided by USEPA and N-STEPS scientists and made available through that program as well as the USEPA nutrient criteria guidance document including the Stressor-Response guidance (U.S. EPA, 2010).

A.1.3.3 Evidence from published literature

A recently completed systematic review of literature (1970-2017) reporting stressor-response relationships between nutrient concentrations and biological communities in streams and rivers provides a rough estimate of available literature-based evidence. Only one published reference was found that measured macroinvertebrate response to nutrients in part based on streams and rivers sampled in State A. Another study with potential relevance was based on a national-scale dataset derived from the NAWQA program showing relationships between nutrients and diatom metrics. State A also shares level III ecoregions with 8 other states, and approximately 17 additional references were found that measured the response of chlorophyll, diatoms, and macroinvertebrates to nutrients in those states. Stressor-response relationships based on data collected in other states are not necessarily relevant to State A but could be looked at more closely to determine this. Overall, literature-based stressor-response evidence is low to medium and may be an underestimate given that the review did not cover all endpoints being considered by the state (e.g., the review did not include DO, pH, or transparency).

As noted above, the state relied on published information from one other state and one other country to inform user perception endpoints and did not conduct their own user perception studies for this effort. The state also relied on equations relating nutrients to sestonic chlorophyll *a* from globally and regionally comparable streams as a line of evidence for TP criteria in one of their use classes. Their scientists are also widely read and several continue to publish peer-reviewed literature; so they have also drawn from that experience.

A.1.3.4 Capacity to conduct or commission new studies

With a program funded by federal partners and capable staff, State A has a high capacity to conduct or commission new studies. The state collected new data for this effort, including experimental work that was mixed in terms of applicability. For example, they explored using nutrient diffusing substrates but did not do many experiments or rely on the output. They explored diatom and soft algal composition indicators, which they use for making aquatic life use assessment decisions independently of macroinvertebrates. They also developed stressor diagnostic tools from the diatom data (e.g., nutrients, conductivity, etc.). In addition, they played with developing a nutrient inference model (sometimes called transfer function models) using diatom nutrient optima similar to approaches applied by other states, but it did not work out. At the time of the discussion, the state was working on a Hilsenhoff biotic index (HBI) type-indicator with algae using nutrient tolerance values (TV) they developed.

A.1.3.5 Weighting evidence

State A considered quality of evidence at all points along the process. This, for example, led to identifying gaps in reference data, identifying recreational targets (using user perception study from a different state), and selecting endpoints.

The state did not judge the quality (e.g., relevance, strength, reliability) of analysis results (evidence) based on specific, previously established criteria. This was mostly done by best professional judgment of the Steering Committee, which went through the steps of evaluating decisions. For example, the Steering Committee seemed to view least disturbed reference population derived distributional values as more protective of uses than values derived from that population of sites meeting macroinvertebrate based biological condition targets. In the end, the state used both endpoints for criteria derivation but with no weighting applied.

For scientific literature, the state explained they similarly relied on the professional judgment of the Steering Committee and staff. They emphasized the criteria for one use class of streams whose values were particularly tricky; there was concern for their protectiveness and the literature helped resolve that.

The state communicated the qualities of the results and evidence in their technical support document, which details the process, logic, and decision-making.

A.1.4 Factors affecting the Criteria Derivation Phase

The methods for developing criteria (including how the state analyzes water quality data or aquatic life relevant to nutrient criteria) all affect how the conclusion of a WoE process (e.g., weighing the body of evidence) would take place.

A.1.4.1 Method for developing criteria and analyzing data relevant to management goals

In general, State A used what they described as a WoE process that adhered somewhat to the weight-of-evidence elements without a formal basis but following the approach organically. They developed several lines of evidence as a result of discussions and feedback from USEPA OW, their EPA region, NSTEPS and an internal Steering Committee. Out of this, they were able to derive and evaluate (weight) several lines of evidence (reference, stressor-response, and literature) and derive values (weigh) to protect different designated uses (aquatic life and recreation).

A.1.4.2 Weighing the body of evidence

The state process for integrating multiple lines of evidence considered three options: the mean, the minimum and an approach applying weights to different lines of evidence. There was a strong interest in weighting lines, but in the end the mean was the easiest to do and to communicate to stakeholders. The EPA Regional coordinator was interested in incorporating percentiles of sites attaining algal indicators; but the state stuck with just the invertebrate index for the line of evidence of attaining populations because the algal model is not yet adopted as a numeric criterion for aquatic life.

To address uncertainty, the state relied on statistical measures. For the logistic models, there was model uncertainty inherent in the analysis and the state chose to use a lower percentile of error around the 60 % probability value because of the risk of low invertebrate sensitivity to nutrients. They also used a conservative confidence interval around sestonic chlorophyll and TP relationships in streams. Lastly, they selected the 90th percentile of reference and 75th percentile of attaining as measures of uncertainty around the condition associated with those populations.

A.2 State C profile

A.2.1 State C – Lakes and streams

State C is an inland state which has developed numeric criteria for one lake and is in the process of developing nutrient thresholds for other lakes and streams.

A.2.2 Factors affecting the Planning and Problem Formulation Phases

The decisions being made and timeline affect how a WoE process would take place.

A.2.2.1 Decision and timetable

State C is earlier in the process of developing nutrient criteria for lakes and streams. The state has two nutrient translator thresholds for one lake which were derived using a multiple lines of evidence approach including published literature, ecoregional values, distributions of values in the lake, nutrient loading information, and lake water quality modeling. It is now in the process of developing an approach and analyses to generate nutrient thresholds to protect other lakes. Concurrently, the state is in the process of exploring numeric nutrient thresholds for streams or certain classes of streams. The estimated timeline is a minimum of 3-5y.

The state is developing criteria to protect designated uses in lakes and streams; these designated uses include protection of aquatic life, recreation, and drinking water quality. The state has focused on a range of assessment endpoints to protect the management goals including for lakes (chlorophyll *a*, cyanobacterial growth/phytoplankton composition, and dissolved oxygen) and for streams (periphyton biomass, macroinvertebrate indices, fish indices, and dissolved oxygen). The state has been in the process of planning, collecting data, developing analysis tools and analyzing data, synthesizing the data, and developing recommended criteria for more than 10 years. This state has a nutrient criteria development plan that is several years old, presented to the USEPA Region, which lays out the process they propose to use for developing criteria, parameters and rationale, approach, and application. It is not detailed regarding lines of evidence, weighting, or weighing. This document does not include the elements of the risk assessment-based approach (problem formulation, conceptual modeling, etc.).

A.2.3 Factors affecting the Analysis Phase

The types of relevant data, capacity to analyze data, evidence from published literature, capacity to conduct or commission new studies, and how evidence is weighted are all important elements of the **Analysis Phase**. These affect how evidence will be weighed for criteria derivation.

A.2.3.1 Types of relevant primary data

State C has a high amount of nutrient data in streams and lakes, but a relatively low amount of relevant primary response data (chlorophyll, algal composition, dissolved oxygen (DO) profiles) compared to the other states in lakes and medium amount for streams. For example, chlorophyll collection began in 2016. They have an established monitoring program that collects nutrient and response (assessment endpoint) data. In addition, they have conducted special studies focused on filling gaps in their streams data, including ecoregionally targeted studies. They rely on their state monitoring data for analysis.

The state prefers to use data collected from waters within state boundaries. This can be collected by a variety of agencies (e.g., DEQ, USEPA, USGS, universities, etc.). For context, the USEPA National Aquatic Resource Survey program (NARS) has collected around 280 samples for streams in State C and there are an additional 2000 samples in adjacent states for the range of endpoints collected by that program (U.S. EPA, 2019). For lakes, there are 245 NARS samples in State C and more than 1500 in adjacent states. The

NARS samples collected during 2007 and 2012 within State C are part of current N-STEPS projects.

A.2.3.2 Capacity to analyze primary data

State C has been limited in their capacity to analyze primary data in support of their analysis for nutrient criteria development, due to resource and time constraints. They have a small group of dedicated scientists collecting and managing data, but with insufficient time to conduct advanced analyses to support the ongoing work. They have relied on external consultants, academics, and N-STEPS to help with site-specific analysis for one lake and with exploratory analyses of existing data for regional lake and stream criteria development work.

A.2.3.3 Evidence from published literature

As noted above, the state relied on scientific literature, including USEPA ecoregional values, to develop site-specific nutrient translator thresholds.

A recently completed systematic review of literature (1970-2017) reporting stressor-response relationships between nutrient concentrations and biological communities in streams and rivers provides a rough estimate of available literature-based evidence. Twelve published references were found that measured biological responses to nutrients in whole or part based on streams and rivers sampled in State C. Another study with potential relevance was based on a national-scale dataset derived from the NAWQA program showing relationships between nutrients and diatom metrics. State C also shares level III ecoregions with 7 other states (excluding a very small overlap in 1 state), and >20 additional references were found that measured the response of chlorophyll, diatoms, and macroinvertebrates to nutrients in those states. Stressor-response relationships based on data collected in other states are not necessarily relevant to State C but could be looked at more closely to determine this. Overall, literature-based stressor-response evidence is relatively high and may be an underestimate because the review did not include all endpoints the state is considering (e.g., the review did not include DO).

No companion systematic review of the literature on stressor-response relationships between nutrient and biological responses in lakes has been completed. However, there is a long history of papers synthesizing data (e.g., phosphorus and chlorophyll stressor-response relationships) from lakes. As a small example, Dillon and Rigler (1974) assembled data from more than 95 lakes in North America, Canfield and Bachmann (1981) similarly assembled data for more than 709 lakes and reservoirs in the US, Smith (1982) compiled data for more than 127 temperature zone northern latitude lakes, OECD (1982) has data from 128 lakes from around the world including 40 in the US, and more recently Soranno et al. (2017) published the LAGOS-NE dataset composed of data from more than 12,000 lakes in the US, including 100,000s of TP and chlorophyll observations. State C shares ecoregions with many of the data from these studies. In addition, in an abbreviated search, we identified more than 56 peer reviewed studies exploring phosphorus and chlorophyll data for lakes in State C by one research group alone. Without extracting data from these studies, it is hard to rank lake specific stressor-response literature evidence, but it would appear to be, at a minimum, medium.

A.2.3.4 Capacity to conduct or commission new studies

State C has generally had a medium capacity to conduct or commission new studies, relying somewhat on help from other agencies. The site-specific lake nutrient translator criteria study, that included water

quality modeling for an important reservoir, was funded independently. The state has collected targeted stream data to fill geographic gaps in representative samples. They have also funded USGS to conduct studies in support of nutrient criteria work, including a filamentous algae study and a pilot study on one river basin that was shown to have insufficient nutrient gradient to generate response curves. In addition, they have received funding from USEPA to support academic consultants to analyze lake and stream data and are receiving support from NSTEPS to conduct lake and stream analyses to help facilitate progress in nutrient threshold development. They have not pursued experimental studies.

A.2.3.5 Weighting evidence

While not yet occurring, State C would appreciate guidance on how to weight evidence. The one site-specific study funded by a third party used multiple lines of evidence, which were all discussed, but there did not appear to be an analysis of relevance, strength, or reliability.

A.2.4 Factors affecting the Criteria Derivation Phase

The methods for developing criteria (including how the state analyzes water quality data or aquatic life relevant to nutrient criteria) all affect how the conclusion of a weight of evidence process (e.g., weighing evidence) would take place.

A.2.4.1 Method for developing criteria and analyzing data relevant to management goals

In general, State C used what they described as a WoE process using reference, stressor-response and literature evidence (and water quality modeling) in developing site-specific nutrient translators for one lake and they are certainly interested in multiple lines of evidence for future work. Their concern, however, is to only use evidence that they can deem reliable and defensible, but there was no specific definition of what that is and they would appreciate guidance to help define that defensibility or reliability threshold; even what elements to consider.

A.2.4.2 Weighing the body of evidence

Given the early stage in criteria development, State C has not yet considered how to weigh or combine the body of evidence to derive criteria. Multiple lines of evidence were used in the site-specific lake study, but it was not clear how those lines of evidence were weighted and integrated to derive the final numeric values.

A.3 Synthesis

This appendix described details relevant to nutrient criteria development in two states that differed in many regards but were selected as a representation of the variety of conditions that exist nationwide. This synthesis compares and contrasts their characteristics.

States varied from those within a year of promulgating rules to those early in the criteria development process. For State A which is close to promulgation, advice on WoE methods (see Chapter 6) will be less useful now but may assist in criteria review in the future. However, many states are early in the process for at least some water body types, if not all (e.g., State C) and, thus, advice on WoE methods will be very welcome.

States generally have access to relevant data, a result of substantial investment in both routine and targeted monitoring work. This includes specific projects supporting nutrient criteria development efforts. It is unlikely that data will be a limitation, except for specific unique response data (e.g., zooplankton in lakes); but for the core variables, most states will have adequate and relevant primary

data for a variety of analyses. Additional large federal agency monitoring effort data (e.g., EPA EMAP, EPA NARS) are available through the water quality portal (https://www.waterqualitydata.us/), and other agency data (e.g., USGS NAWQA) can be accessed through agency specific web portals.

The capacity to analyze data within states varied, ranging from not limited by in-house staff expertise to relying much more on external support from academic partners or federal agencies. States could always use support to assure continuity of skills with staff turnover, continued staff training, and emergent technology. States noted the benefit of technical support from USEPA in helping their ongoing efforts.

The availability of and reliance on evidence from published literature relevant to nutrient criteria development is, on the whole, moderate. The breadth and depth of nutrient criteria research is regional and where that research has been conducted, states generally use it. But where there are regional gaps in relevant research, this reliance is lower and limited to national scale of general studies that may lack state specificity.

States have at least medium capacity for new studies and tend to rely on opportunistic support from federal agency partners to fund and conduct new research to develop tools or analyses to support nutrient criteria efforts. Some states that are better funded have been able to fund targeted, internal studies (e.g., State A).

States have generally not formally evaluated their evidence in terms of relevance, strength, or reliability (weighting evidence). Where evidence was evaluated, it was mostly done ad-hoc with best professional judgment, sometimes by a team, but not based on a method or any specific rules.

In terms of criteria development methods, each state remains interested in multiple lines of evidence including literature, stressor-response (including, generally, multiple stressor-response results), and reference-based analyses (again, often including a few reference approaches). No method of objectively weighing was favored, although one state (State A) did consider using a straight mean. Where possible, states used statistical uncertainty in interpreting the results, but not specifically in weighing evidence. States evaluated the degree of convergence among lines to the degree possible, but that convergence varied and where there was more variability, states tended to rely on a transparent best professional evaluation of the evidence with regards to the linkage to management goals (e.g., lines tied closely to use protection), the need to be protective (e.g., least disturbed reference values were seen as more protective by State A), and what appeared like an evaluation of the reliability of the evidence in terms of sample size and statistical certainty. All the states expressed great interest in much more help and advice on how to weigh a body of evidence.

Appendix B

The purpose of this appendix is to provide an opportunity for you to think through the criteria development process while applying the weight-of-evidence (WoE) methods described in this report. The exercise presented here was first introduced at the **2018 USEPA Nutrient Criteria Workshop**; it has been updated to highlight where and how the *Basic WoE Framework* could be applied. Text boxes throughout the appendix are reminders of critical questions and decision-points that should be considered while working through the criteria development process.

Scenario: Congratulations! You are part of a team assigned the responsibility for developing and carrying out a process to derive numeric nutrient criteria (NNC) for your Agency. You will be recommending NNC to protect recreational uses (fishable/swimmable) in natural lakes. You are familiar with EPA guidance. You know that stressor-response models are available based on national lakes data. You also know that there are other scientifically defensible approaches described in the nutrient criteria guidance documents (e.g., other stressor-response relationship and mechanistic modeling, use of published literature) and that guidance suggests considering multiple lines of evidence is appropriate, as well. Your Agency has decided that using multiple, scientifically defensible lines of evidence for deriving numeric values is the preferred approach. Your task is to develop and document a process for deriving the magnitude component of NNC and to apply your expertise in WoE methods to enhance transparency and defensibility of your conclusions.

B.1 Planning Phase

You recall that creating plans that are transparent, documented, and that leverage collective expertise supports the core principles of the *Basic WoE Framework.* Think about how you would map out your plan for the criteria development process. Would a figure be appropriate (Figure B-1)? How could leveraging collective expertise to review the plan or to review specific aspects of analysis or derivation strengthen conclusions?

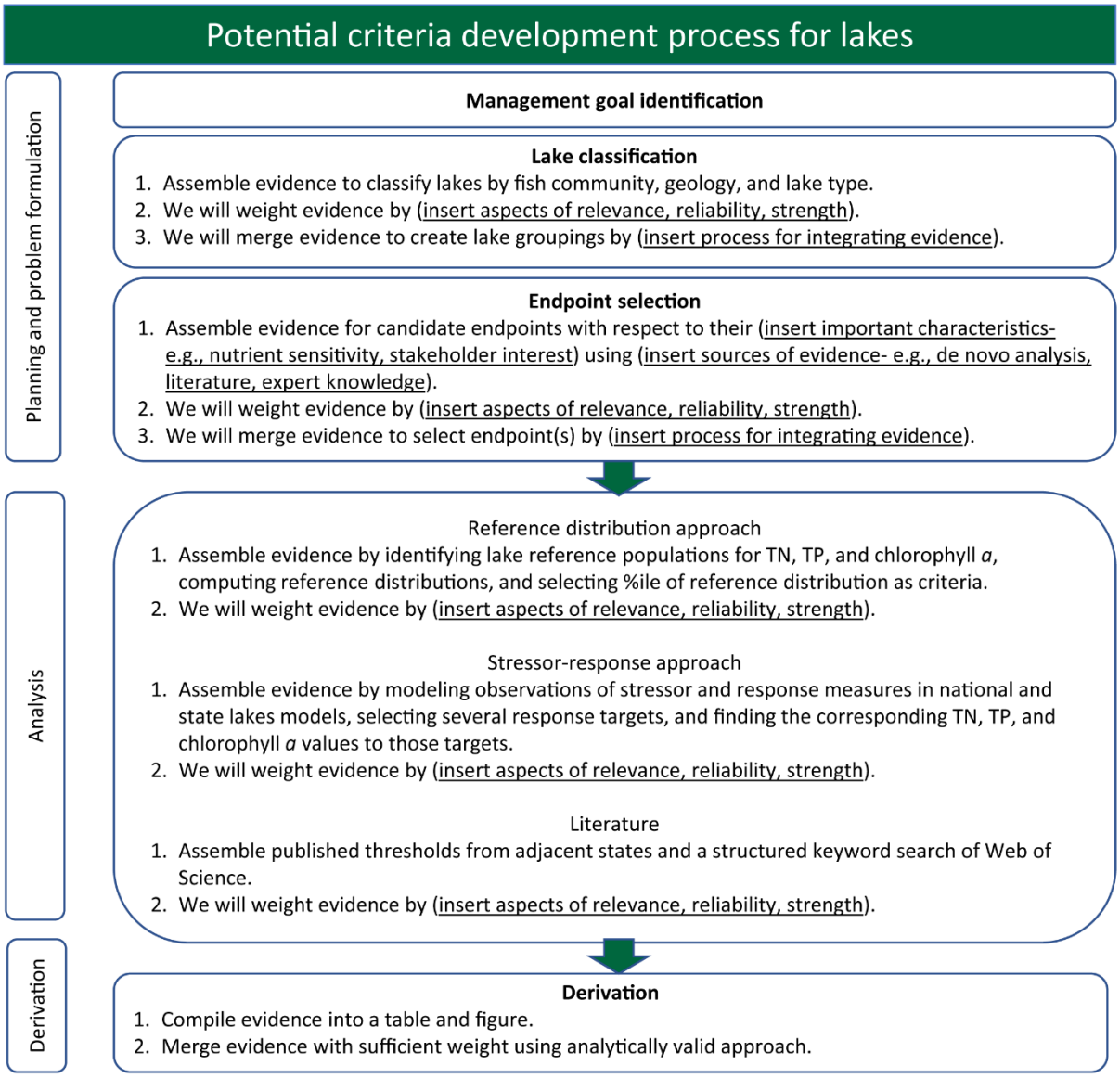


Figure B-1. Potential planning template for NNC to protect recreational use of lakes

This flow diagram is one way to document a planned process for NNC development. Think about ways you might communicate with stakeholders about your methods (including WoE methods) for conducting each phase of criteria development.

You also recall that the **Planning Phase** involves grouping water bodies and that WoE methods can help integrate evidence as you decide how to group lakes. If you had evidence on the following lake traits, think about and document how you would weight each (what makes evidence on lake traits relevant, reliable, and strong?). Also think about and document how you would integrate across evidence to create your groupings (will you use evidence about one or multiple lake traits?).

1. Fish community (e.g., Cold water, Cool water, Warm water)
2. Predominant Bottom Substrate (e.g., Rocky, Sand/silt, Mud)
3. Size/Depth (e.g., Small/Shallow, Medium/Medium, Large/Deep)
4. Lake Type (e.g., Natural, Reservoirs)

Grouping Lakes

• What are the lake traits YOUR state has to work with?

B.2 Problem Formulation Phase

You know that problem formulation includes selecting assessment endpoints, which is a process amenable to WoE methods. Assessment endpoints should be relevant to management goals, measurable, ecologically relevant, sensitive to nutrients, and important to stakeholders. If you had evidence on the following endpoints, think about and document how you would weight each (what makes evidence on endpoints relevant, reliable, and strong?). Also think about and document how you would integrate across evidence to select endpoints (will you move forward with all endpoints that have sufficiently weighty evidence?).

1. Water clarity
2. Phytoplankton
3. Harmful Algal Blooms (HABs)
4. Diatoms
5. Benthic fauna
6. Submerged Aquatic Vegetation (SAV)
7. Epiphytes
8. Dissolved Oxygen (DO)
9. Invasive species
10. Algal toxins

Conceptual models developed during problem formulation can help you visualize how sources, stressors, and endpoints are related (Figure B-2). In this case, it allows a user or manager to see how stressors, stressor sources, secondary factors that influence interactions, and ways in which the assessment endpoints are affected by and influence the ultimate management goal (restoration of fishable/swimmable recreational uses). Think about how your conceptual model might inform what evidence you assemble in the **Analysis Phase**.

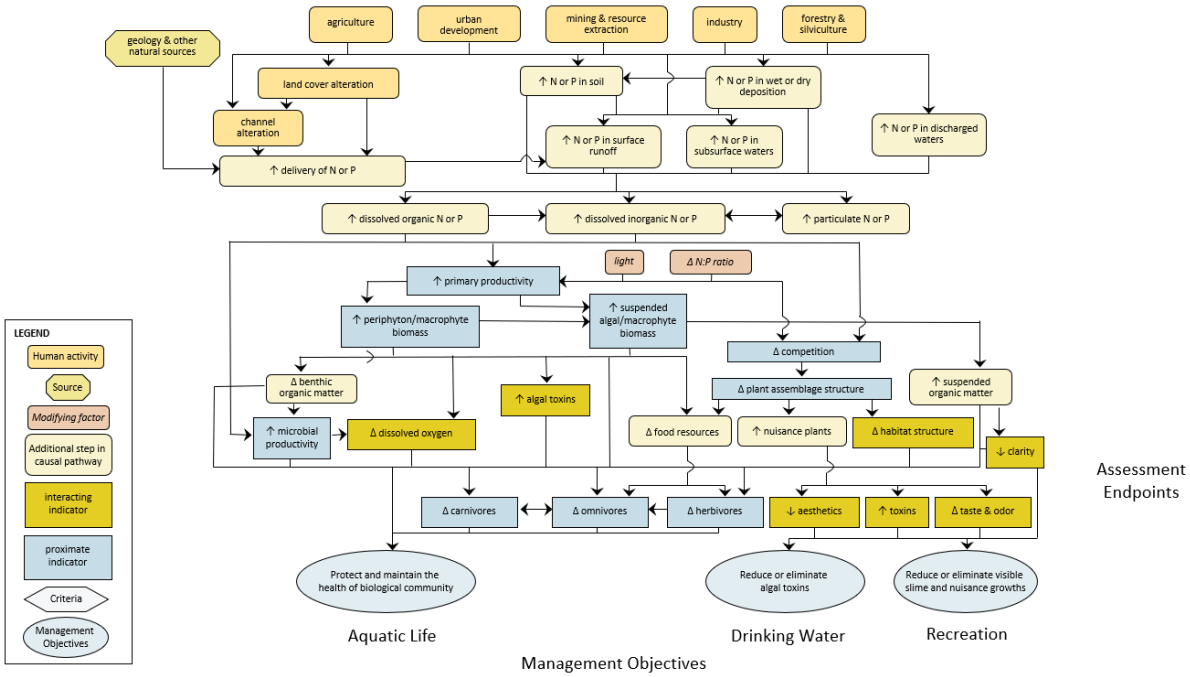


Figure B-2. Example lake conceptual model

This conceptual model includes sources of probable stressors and ways they would interact with various lakes found in any state, as well as potential assessment endpoints and management objectives. Think about how your state could use this model or a model like it to help communicate this process to managers and stakeholders.

B.3 Analysis Phase

B.3.1 Assemble Evidence

You are very lucky that your team has assembled the following evidence. Think about and document what practices your team would have used to assemble an unbiased set of evidence.

Table B-1. Reference Based Values

The following values are distributions of numeric values for TP, TN, and chlorophyll in the water column from different populations of sites within the state.

		Growing Seasonal Values								
Population		**TP (mg/L)**			**TN (mg/L)**			**Chlorophyll (ug/L)**		
	N	25th	50th	75th	25th	50th	75th	25th	50th	75th
Reference Lakes	20	0.002	0.008	0.012	0.200	0.400	0.650	0.5	1.7	3.5
All lakes	210	0.003	0.016	0.030	0.300	0.800	1.200	0.8	3.4	6.5
Assessed Lakes Known to be Meeting Uses	24	0.004	0.011	0.020	0.400	0.550	0.800	1.0	2.3	4.7
Impaired Lakes	7	0.012	0.030	0.054	0.600	1.500	2.160	3.0	5.8	9.4

Table B-2. Stressor-Response Values

The following are values for TP, TN, and chlorophyll derived from stressor-response relationship modeling from a national survey of lakes.

				Stressor Concentration to Meet Target		
Response	**Response Target**	**Allowable exceedance probability**	**Certainty level (%)**	**TP (mg/L)**	**TN (mg/L)**	**Chl a (µg/L)**
Microcystin concentration	6	0.02	90			12.3
Microcystin concentration	8	0.02	90			15.9
Chlorophyll	12		90	0.019	0.46	
Chlorophyll	16		90	0.024	0.51	

The following are values for TP, TN, and chlorophyll derived from stressor-response relationship modeling from lakes in the state. These models were developed before models were available from the national survey of lakes.

Response	**Response Target**	**Stressor Concentration to Meet Target**		
		TP (mg/L)	**TN (mg/L)**	**Chlorophyll (µg/L)**
Chlorophyll	2	0.008	0.46	
Chlorophyll	5	0.028	0.72	
Chlorophyll	15	0.072	2.1	
R^2 (p-value)		0.6 (<0.05)	0.54 (<0.05)	
Cyano Density	20,000	0.03		10
Cyano Density	50,000	0.04		15
Cyano Density	100,000	0.045		20
R^2 (p-value)		0.42 (<0.05)		0.60 (<0.05)
Microcystis Density	20,000	0.061	0.92	12
Microcystis Density	50,000	0.048	1.10	17
Microcystis Density	100,000	0.021	0.71	24
R^2 (p-value)		0.41 (<0.05)		0.62 (<0.05)
Hypolimnetic DO	0	0.03	1	5
Hypolimnetic DO	2	0.02	0.7	4
Hypolimnetic DO	4	0.015	0.63	3.2
Hypolimnetic DO	6	0.005	0.23	1
R^2 (p-value)		0.50 (<0.05)	0.48 (<0.05)	0.53 (<0.05)

Table B-3. Published Literature Values

The following are values for TP and chlorophyll derived from the scientific literature.

Citation	Assessment Endpoint	TP (mg/L)	Chl (µg/L)
Schupp and Wilson (1993)	Peak coldwater fish abundance	0.006	1
Johnston et al. (1999)	Coldwater fish growth peak	0.009	6
Elliott et al. (1996)	Coldwater fish growth increase (England)	0.011	14

Adjacent State A 2010	TP (mg/L)	TN	Chl a (µg/L)
Coldwater	<0.012		<3
Coolwater	<0.020		<6
Recreation	<0.030		<9

Adjacent State B 2016	TP (mg/L)	TN	Chl a (µg/L)
High altitude	0.012		2.6
Low altitude, excellent aesthetics	0.017		3.8
Low altitude, good aesthetics	0.018		7.0

Study	Location	Surveyed Group	Respondent Ranking	Chl-a Level (µg/L)
Hoyer et al. (2004)	FL	Citizen lake monitors	Excellent for swimming (rank=1,2)	7 to 12 (mean) 2.5 – 10.5 (range+)
			Slightly impaired for swimming (rank=3)	14 (mean) 5 – 11 (range+)
			Undesirable (rank=4,5)	5 to 80 (mean) 2.5 – 110 (range+)
Heiskary and Walker (1988)	MN	Agency staff	Excellent for swimming (rank=1,2)	5 to 10 ppb (mean) 2 – 17 ppb (range+)
			Slightly impaired for swimming (rank=3)	45 (mean) 15 – 60 ppb (range+)
			Undesirable (rank=4,5)	55 ppb (mean) 40 – 75 ppb (range+)

B.3.2 Weight evidence

In order to decide how much influence each piece of evidence should have on your conclusions, you know you need to determine its relevance, reliability, and strength. Use the blank table below to think about and document how you would judge these three qualities of the evidence you have assembled.

Weight Evidence

- What information do you have and what additional information would you want to know about the evidence above to accurately weight it?
- How will you score and assign weights?
- How will you be transparent about weighting decisions?
- Will you use a figure or evidence weighting table (e.g., Table B-4)?

Table B-4. Table to Weight Evidence

This table is an example that might be completed with "++, +, -, 0" based on the weight of particular evidence available. It serves as a visual representation of the evidence so each line can be compared.

Piece of Evidence	Relevance	Reliability	Strength	Overall	Explanation
Reference Condition Evidence 1					
Stressor-Response Evidence 1					
Stressor-Response Evidence 2					
Scientific Literature Evidence 1					
Stakeholder Surveys					

B.4 Criteria Derivation Phase

B.4.1 Evidence aggregation and integration

You are in the home stretch! In this final part of criteria development, you are ready to use your weighted evidence to derive criteria. First, you need to decide whether evidence aggregation is a step you want to take. If you only have a few pieces of evidence to integrate, it may be unnecessary. If you have so much evidence that it will be difficult to communicate how you are combining it to draw a conclusion, aggregation can be valuable.

Evidence Aggregation & Integration

- Do you have weighty enough evidence to make a decision?
- Will you pick the weightiest evidence and let that determine your decision?
- Will you merge multiple pieces or lines of evidence? How will you merge evidence?

Next, think about and document how you will integrate your pieces or lines of evidence. Don't forget that criteria derivation should be accompanied by interpretation, explanation, and description of any outstanding ambiguities or uncertainties. When available, uncertainty may be expressed statistically as a range and/or probability of possible conclusions.

Finally, think about how you will communicate your conclusions. Will you use a figure or table (e.g., Table B-5)?

Table B-5. Table for Weighing the Body of Evidence

This table is an example that would be completed with data values and weight ranges from Table B-4 (above). Specific lines of evidence can be selected, or multiple lines of evidence can be combined, depending on the weight of evidence and the needs of your particular state.

Line of Evidence	TP µg/L	TN mg/L	Chl-a µg/L	Notes
Reference Conditions				
Stressor-Response				
Scientific Literature				
Stakeholder Surveys				

You are DONE! With your hard work you are well on your way to **Criteria Adoption**.

B.5 Examples

The following section demonstrates how three different teams could have gone through the criteria development process to arrive at numeric criteria for various lake types. The hypothetical teams went through the following process, but focused their practice on WoE methods in Steps 4-6:

Step 1: Identify relevant waterbody type (**Planning**)
Step 2: Identify possible sources of stressors (**Problem Formulation**)
Step 3: Identify assessment endpoints (**Problem Formulation**)
*Before moving into Step 4, it is helpful to create a conceptual model (see Figure B-2)
Step 4: Assemble evidence – primary data analyses, published literature, expert knowledge (**Analysis**)
Step 5: Weight evidence (**Analysis**)
Step 6: Weigh the body of evidence (**Criteria Derivation**)

This appendix provides three examples that allow users to walk through the steps detailed above to see how the process might be followed in their state, using their data. Some of the choices will not always match every possible circumstance perfectly. This is by design. The examples are built in such a way that users can see where and why their state path may vary from what was followed here.

B.5.1 Example 1

B.5.1.1 Step 1 – Identify Waterbody Type

Fish community

- Cold water, **Cool water**, Warm water

Predominant Bottom Substrate

- Rocky, **Sand/silt**, Mud

Size/Depth

- Small/Shallow, **Medium/Medium**, Large/Deep

Natural Lake Type

- Continental Glacial, **Alpine Glacial**, Coastal Plain, Playas, Potholes, and Sandfill Lakes

Reservoirs

- Tributary storage, Run-of-the-river, Main stem storage

For this example, we identified *a cool water, sandy, medium-sized, natural glacial lake.*

B.5.1.2 Step 2 – Identify Possible Sources of Stressors

Point source

- **Wastewater Treatment Plant**
- Manufacturing by-products
- Underground storage tanks
- **Septic tanks**

Non-point source

- **Fertilizers**
- Land disturbance
- **Urban/Suburban Runoff**
- Wildfires
- Invasive species
- **Sedimentation**
- Pesticides
- Emerging Contaminants

We identified *a wastewater treatment plant, septic tanks, fertilizers, runoff, and sedimentation as possible stressor sources.*

B.5.1.3 Step 3 – Identify Assessment Endpoints

- Water clarity
- Phytoplankton
- HABs
- Diatoms
- **Benthic fauna**
- **SAV**
- Epiphytes
- DO
- Invasive species
- **Algal toxins**

In this example, we decided that the important stressors relevant to restoring and maintaining fishable/swimmable water quality included *water clarity* and *HABs*. Important assessment endpoints are *benthic fauna, submerged aquatic vegetation*, and *algal toxins*.

B.5.1.4 Step 4 – Assemble Evidence

As discussed in the document, there are multiple types of evidence that may or may not be available when developing NNC. For this example, we chose to use reference conditions, stressor-response values, scientific literature, and stakeholder surveys. (Recall that the data used in this example were part of an existing group exercise and have no actual connection to a specific state or region.) In the following steps, many tables are used to show exactly what pieces of evidence are chosen in order to be transparent about the selection process, document exactly what is decided on, and clearly communicate those decisions. This builds confidence in the evidence as well as the conclusions.

The first line of evidence we analyzed was reference condition (Table B-6). Data were prepared and we chose to use the 75th percentile of the reference distribution for TN, TP, and chlorophyll-a as protective of recreational use.

Table B-6. Example 1- Reference Conditions

The 75th percentile of the reference distribution for TP, TN, and chlorophyll-a are highlighted.

		Growing Seasonal Values								
Population		**TP (mg/L)**			**TN (mg/L)**			**Chlorophyll (µg/L)**		
	N	25th	50th	75th	25th	50th	75th	25th	50th	75th
Reference Lakes	20	0.002	0.008	**0.012**	0.200	0.400	**0.650**	0.5	1.7	**3.5**
All lakes	210	0.003	0.016	0.030	0.300	0.800	1.200	0.8	3.4	6.5
Assessed Lakes Known to be Meeting Uses	24	0.004	0.011	0.020	0.400	0.550	0.800	1.0	2.3	4.7
Impaired Lakes	7	0.012	0.030	0.054	0.600	1.500	2.160	3.0	5.8	9.4

RESULT	**TP: 12µg/L**	**TN: 0.65mg/L**	**Chl-a: 3.5µg/L**

The second line of evidence considered was stressor-response relationships (Table B-7). This evidence was prepared from models developed with national lakes data and from previously existing models developed with state data. We highlighted a range of conservative targets for recreational (fishable/swimmable) designated use.

Table B-7. Example 1- Stressor-Response Relationships

A range of conservative targets are highlighted.

				Stressor Concentration to Meet Target		
Response	**Response Target**	**Allowable exceedance probability**	**Certainty level (%)**	**TP (mg/L)**	**TN (mg/L)**	**Chl a (µg/L)**
Microcystin concentration	6	0.02	90			12.3
Microcystin concentration	8	0.02	90			15.9
Chlorophyll	12		90	**0.019**	**0.46**	
Chlorophyll	16		90	0.024	0.51	

Response	Response Target	Stressor Concentration to Meet Target		
		TP (mg/L)	TN (mg/L)	Chlorophyll (µg/L)
Chlorophyll	2	0.008	0.46	
Chlorophyll	5	**0.028**	**0.72**	
Chlorophyll	15	0.072	2.1	
R^2 (p-value)		0.6 (<0.05)	0.54 (<0.05)	
Cyano Density	20,000	**0.03**		**10**
Cyano Density	50,000	0.04		15
Cyano Density	100,000	0.045		20
R^2 (p-value)		0.42 (<0.05)		0.60 (<0.05)
Microcystis Density	20,000	**0.061**	**0.92**	**12**
Microcystis Density	50,000	0.048	1.10	17
Microcystis Density	100,000	0.021	0.71	24
R^2 (p-value)		0.41 (<0.05)		0.62 (<0.05)
Hypolimnetic DO	0	0.03	1	5
Hypolimnetic DO	2	0.02	0.7	4
Hypolimnetic DO	4	**0.015**	**0.63**	**3.2**
Hypolimnetic DO	6	0.005	0.23	1
R^2 (p-value)		0.50 (<0.05)	0.48 (<0.05)	0.53 (<0.05)

RESULT	TP: 15-61µg/L	TN:0.46-0.92mg/L	Chl-a: 3.2-12.3µg/L

The third line of evidence reviewed was scientific literature (Table B-8). Applicable criteria ranges were highlighted to show protective literature values for recreational uses.

Table B-8. Example 1- Scientific Literature Selections

Applicable ranges for TP and Chl-a are highlighted.

Citation	Assessment Endpoint	TP	Chl
Schupp and Wilson (1993)	Peak coldwater fish abundance	0.006	1
Johnston et al. (1999)	Coldwater fish growth peak	0.009	6
Elliott et al. (1996)	Coldwater fish growth increase (England)	0.011	14

Adjacent State A 2010	TP	TN	Chl a
Coldwater	<0.012		<3
Coolwater	<0.020		<6
Recreation	**<0.030**		**<9**

Adjacent State B 2016	TP	TN	Chl a
High altitude	**0.012**		**2.6**
Low altitude, excellent aesthetics	0.017		3.8
Low altitude, good aesthetics	0.018		7.0

Result	TP: 12-30µg/L	TN: n/a	Chl-a: 2.6-9µg/L

Finally, we used stakeholder surveys as the last line of evidence (Table B-9). There were two studies available that showed a range of chlorophyll-a values for excellent swimming conditions. Please note that only chlorophyll-a data was available, therefore there are no results in this section for TP or TN.

Table B-9. Example 1- Stakeholder Survey Results

Chl-a levels associated with excellent swimming conditions are highlighted.

Study	Location	Surveyed Group	Respondent Ranking	Chl-a Level (µg/L)
Hoyer et al. (2004)	FL	Citizen lake monitors	Excellent for swimming (rank=1,2)	**7 to 12 (mean)** 2.5 – 10.5 (range+)
			Slightly impaired for swimming (rank=3)	14 (mean) 5 – 11 (range+)
			Undesirable (rank=4,5)	5 to 80 (mean) 2.5 – 110 (range+)
Heiskary and Walker (1988)	MN	Agency staff	Excellent for swimming (rank=1,2)	**5 to 10 ppb (mean)** 2 – 17 ppb (range+)
			Slightly impaired for swimming (rank=3)	45 (mean) 15 – 60 ppb (range+)
			Undesirable (rank=4,5)	55 ppb (mean) 40 – 75 ppb (range+)

RESULT	TP: n/a	TN: n/a	Chl-a: 5-12µg/L

B.5.1.5 Step 5 –Weight Evidence

Following the instructions laid out in the report, we weighted the evidence (Table B-10). Note that all evidence provided addresses magnitude. None of it addresses duration or frequency. See the notes in Table B-10 for brief explanations of weighting decisions.

Table B-10. Example 1- Weight Evidence

Line of Evidence	Relevance	Reliability	Strength	Overall	Notes
Reference Conditions	++	++	++	++	Reference sites are well-defined and sensitive to natural variability. 75th percentile has solid precedent.
Stressor Response	+	+	+	+	For state models, estimated response is to the mid-point of the stressor, which might not be conservative. Target selection is undocumented.
Scientific Literature	+	++	++	++	Literature values are well-vetted and acceptable in other settings. Settings are not always specific to focal lake type.
Stakeholder Surveys	++	0	+	+	Very relevant for aesthetics. Highly subjective and variable results.

B.5.1.6 Step 6 – Weigh the body of evidence

The document describes different approaches for evidence integration (see Section 5.1.2). For Example 1, we decided to base our numeric nutrient criteria on the reference condition and scientific literature lines of evidence. Based on the values in Table B-10, they were the "weightiest" evidence. This results in candidate criteria with the ranges shown in Table B-11. See Figure B-3 for the same information shown in a graphical format.

Table B-11. Example 1- Evidence Summary

Line of Evidence	TP µg/L	TN mg/L	Chl-a µg/L	Notes
Reference Conditions	**12**	**0.65**	**3.5**	++
Stressor Response	15-61	0.46-0.92	3.2-12.3	+
Scientific Literature	**12-30**	n/a	**2.6-9**	++
Stakeholder Surveys	n/a	n/a	5-12	+
Candidate criteria	12-30	0.65	2.6-9	

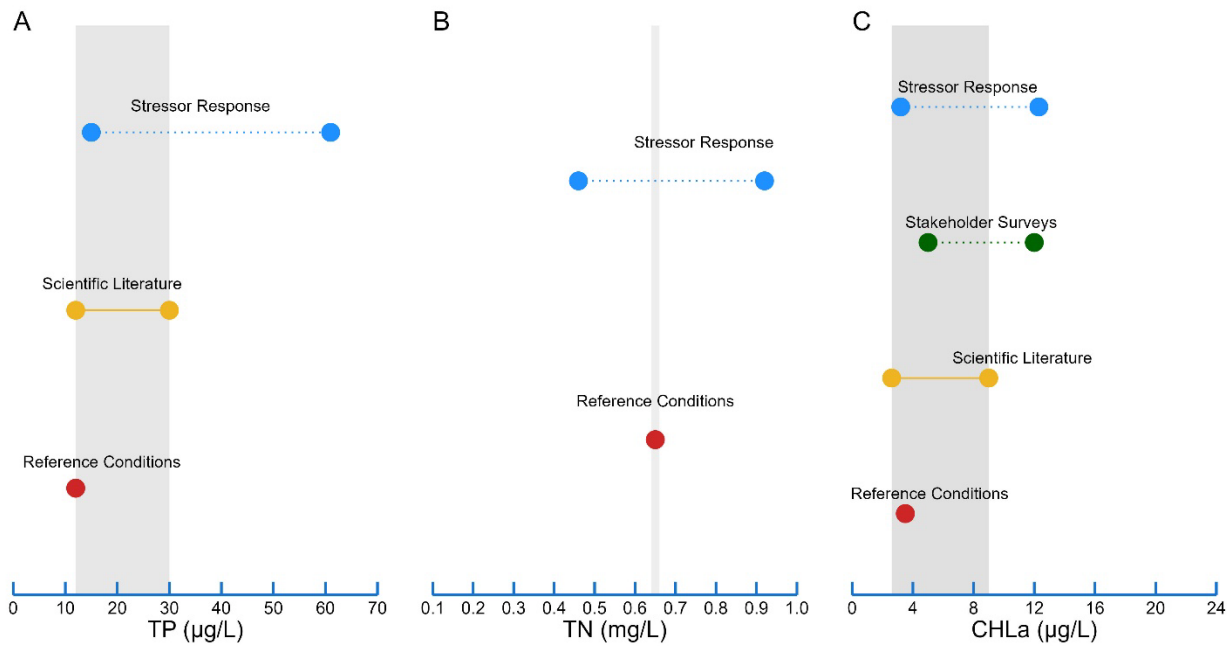


Figure B-3. Compiled Lines of Evidence and Conclusions for TP (A), TN (B), and Chl-a (C) in Example 1

Lines of evidence are labeled and shown as horizontal lines (solid = ++ weight; dashed = + weight). Proposed numeric criteria are shown in relation to the lines of evidence as vertical grey shading.

B.5.2 Example 2

B.5.2.1 Step 1 – Identify Waterbody Type

Fish community

- **Cold water**, Cool water, Warm water

Predominant Bottom Substrate

- **Rocky**, Sand/silt, Mud

Size/Depth

- Small/Shallow, **Medium/Medium**, Large/Deep

Natural Lake Type

- **Continental Glacial**, Alpine Glacial, Coastal Plain, Playas, Potholes, and Sandfill Lakes

Reservoirs

- Tributary storage, Run-of-the-river, Main stem storage

For this example, we identified a *cold water, rocky, medium-sized, continental lake*, as highlighted above.

B.5.2.2 Step 2 – Identify Possible Sources of Stressors

Point source

- Wastewater Treatment Plant
- **Manufacturing by-products**
- **Underground storage tanks**
- Septic tanks

Non-point source

- **Fertilizers**
- **Land disturbance**

- Urban/Suburban Runoff
- Wildfires
- **Invasive species**
- **Sedimentation**
- **Pesticides**
- Emerging Contaminants

We identified *manufacturing by-products, underground storage tanks, fertilizers, land disturbance, invasive species, sedimentation, and pesticides* as possible stressor sources.

B.5.2.3 Step 3 – Identify Assessment Endpoints

- Water clarity
- Phytoplankton
- HABs
- **Diatoms**
- Benthic fauna
- **SAV**
- Epiphytes
- DO
- Invasive species
- **Algal toxins**

In this example, we determined that the important stressors relevant to restoring and maintaining fishable/swimmable water quality included *water clarity* and *dissolved oxygen*. Important assessment endpoints included *diatoms, submerged aquatic vegetation,* and *algal toxins*.

B.5.2.4 Step 4 – Assemble Evidence

As discussed in the document, there are multiple types of evidence that may or may not be available when developing NNC. For this example, we chose to use reference conditions, stressor-response values, scientific literature, and stakeholder surveys. (The evidence used in this example was part of an existing group exercise and had no actual connection to a specific state or region.) In the following steps, many tables are used to show the pieces of evidence that were chosen. This allows us to be transparent about the process for selecting evidence, to document the selections, and to clearly communicate those decisions. This builds confidence in the evidence as well as the conclusions.

The first line of evidence we analyzed used distribution statistics from assessed lakes known to be meeting uses (Table B-12). Data was prepared and we chose to use the 75th percentile of the assessed lakes known to be meeting uses for TN, TP, and chlorophyll-a as protective of recreational use. In this example, it is not known if these data are from inside or outside of our state or region.

Table B-12. Example 2- Reference Conditions

The 75th percentile of the assessed lakes known to be meeting uses for TP, TN, and chlorophyll-a are highlighted.

		Growing Seasonal Values								
Population		**TP (mg/L)**			**TN (mg/L)**			**Chlorophyll (µg/L)**		
	N	25th	50th	75th	25th	50th	75th	25th	50th	75th
Reference Lakes	20	0.002	0.008	0.012	0.200	0.400	0.650	0.5	1.7	3.5
All lakes	210	0.003	0.016	0.030	0.300	0.800	1.200	0.8	3.4	6.5
Assessed Lakes Known to be Meeting Uses	24	0.004	0.011	**0.020**	0.400	0.550	**0.800**	1.0	2.3	**4.7**
Impaired Lakes	7	0.012	0.030	0.054	0.600	1.500	2.160	3.0	5.8	9.4

RESULT	TP: 20µg/L	TN: 0.80mg/L	Chl-a: 4.7µg/L

The second line of evidence was stressor-response relationships (Table B-13). This evidence was prepared from models developed with national lakes data and from previously existing models developed with state data. We highlighted a range of conservative targets for recreational (fishable/swimmable) designated use.

Table B-13. Example 2- Stressor-Response Relationships

A range of conservative targets are highlighted.

				Stressor Concentration to Meet Target		
Response	**Response Target**	**Allowable exceedance probability**	**Certainty level (%)**	**TP (mg/L)**	**TN (mg/L)**	**Chl a (µg/L)**
Microcystin concentration	6	0.02	90			**12.3**
Microcystin concentration	8	0.02	90			15.9
Chlorophyll	12		90	**0.019**	**0.46**	
Chlorophyll	16		90	0.024	0.51	

Response	**Response Target**	**Stressor Concentration to Meet Target**		
		TP (mg/L)	**TN (mg/L)**	**Chlorophyll (µg/L)**
Chlorophyll	2	0.008	0.46	
Chlorophyll	5	**0.028**	**0.72**	
Chlorophyll	15	0.072	2.1	
R^2 (p-value)		0.6 (<0.05)	0.54 (<0.05)	
Cyano Density	20,000	**0.03**		**10**
Cyano Density	50,000	0.04		15
Cyano Density	100,000	0.045		20
R^2 (p-value)		0.42 (<0.05)		0.60 (<0.05)
Microcystis Density	20,000	0.061	0.92	12
Microcystis Density	50,000	0.048	1.10	17
Microcystis Density	100,000	0.021	0.71	24
R^2 (p-value)		0.41 (<0.05)		0.62 (<0.05)
Hypolimnetic DO	0	0.03	1	5

Response	Response Target	Stressor Concentration to Meet Target		
		TP (mg/L)	TN (mg/L)	Chlorophyll (µg/L)
Hypolimnetic DO	2	0.02	0.7	4
Hypolimnetic DO	4	0.015	0.63	3.2
Hypolimnetic DO	6	**0.005**	**0.23**	**1**
R^2 (p-value)		0.50 (<0.05)	0.48 (<0.05)	0.53 (<0.05)

RESULT	TP: 5-48µg/L	TN: 0.23-1.1mg/L	Chl-a: 1-17µg/L

The third line of evidence used was peer-reviewed scientific literature to associate nutrient and chlorophyll values with recreational uses (Table B-14). Applicable ranges were highlighted for recreational uses.

Table B-14. Example 2- Published Literature Selections

Applicable ranges for TP and Chl-a are highlighted.

Citation	Assessment Endpoint	TP	Chl
Schupp and Wilson (1993)	Peak coldwater fish abundance	0.006	1
Johnston et al. (1999)	Coldwater fish growth peak	0.009	6
Elliott et al. (1996)	Coldwater fish growth increase (England)	0.011	14

Adjacent State A 2010	TP	TN	Chl a
Coldwater	<0.012		<3
Coolwater	<0.020		<6
Recreation	**<0.030**		**<9**

Adjacent State B 2016	TP	TN	Chl a
High altitude	0.012		2.6
Low altitude, excellent aesthetics	**0.017**		**3.8**
Low altitude, good aesthetics	0.018		7.0

RESULT	TP: 17-30µg/L	TN: n/a	Chl-a: 3.8-9µg/L

Finally, we used stakeholder surveys as the last line of evidence (Table B-15). There were two studies available that showed a range of chlorophyll a values for excellent swimming conditions. Please note that only chlorophyll a data was available, therefore there are no results in this section for TP or TN.

Table B-15. Example 2- Stakeholder Survey Results

Chl-a levels associated with excellent swimming conditions are highlighted.

Study	Location	Surveyed Group	Respondent Ranking	Chl-a Level (µg/L)
Hoyer et al. (2004)	FL	Citizen lake monitors	Excellent for swimming (rank=1,2)	**7 to 12 (mean)** 2.5 – 10.5 (range+)
			Slightly impaired for swimming (rank=3)	14 (mean) 5 – 11 (range+)
			Undesirable (rank=4,5)	5 to 80 (mean) 2.5 – 110 (range+)
Heiskary and Walker (1988)	MN	Agency staff	Excellent for swimming (rank=1,2)	**5 to 10 ppb (mean)** 2 – 17 ppb (range+)
			Slightly impaired for swimming (rank=3)	45 (mean) 15 – 60 ppb (range+)
			Undesirable (rank=4,5)	55 ppb (mean) 40 – 75 ppb (range+)

RESULT	TP: n/a	TN: n/a	Chl-a: 5-12µg/L

B.5.2.5 Step 5 – Weight Evidence

Following the instructions laid out in this document, we are now ready to weight evidence (Table B-16). Note that all evidence provided addresses magnitude. None of it addresses duration or frequency. See the notes in Table B-16 for brief explanations of weighting decisions.

Table B-16. Example 2- Weight Evidence

Line of Evidence	Relevance	Reliability	Strength	Overall	Notes
Reference Conditions	+	++	+	+	The lakes are not true reference sites but are known to be meeting existing uses. 75th percentile has solid precedent.
Stressor Response	+	+	+	+	For state models, estimated response is to the mid-point of the stressor, which might not be conservative. Target selection is undocumented.
Scientific Literature	+	++	++	++	Literature values are well-vetted and acceptable in other settings. Settings are not always specific to specific lake type.
Stakeholder Surveys	++	0	+	+	Very relevant for aesthetics. Highly subjective and variable results.

B.5.2.6 Step 6 – Weigh the body of evidence

The document describes different approaches for evidence integration (see Section 5.1.2). For Example 2, we decided to base our numeric nutrient criteria on the scientific literature line of evidence. Based on the judgements in Table B-16, it was evaluated as the strongest line of evidence to inform the conclusion. This results in candidate criteria with the ranges shown in Table B-17. Since no literature evidence was available for TN, additional evidence will be collected in future efforts.

Table B-17. Example 2- Evidence summary

Line of Evidence	TP µg/L	TN mg/L	Chl-a µg/L	Notes
Reference Conditions	20	0.80	4.7	+
Stressor Response	5-48	0.23-1.10	1.0-17	+
Scientific Literature	**17-30**	n/a	**3.8-9**	++
Stakeholder Surveys	n/a	n/a	5-12	+
Candidate criteria	17-30	n/a	3.8-9	

B.5.3 Example 3

B.5.3.1 Step 1 – Identify Waterbody Type

Fish community

- Cold water, Cool water, **Warm water**

Predominant Bottom Substrate

- Rocky, Sand/silt, **Mud**

Size/Depth

- **Small/Shallow**, Medium/Medium, Large/Deep

Natural Lake Type

- Continental Glacial, Alpine Glacial, **Coastal Plain**, Playas, Potholes, and Sandfill Lakes

Reservoirs

- Tributary storage, Run-of-the-river, Main stem storage

For this example, we identified a *warm water, muddy bottom, small/shallow coastal plain lake*. We tried to highlight options for a more urban lake setting, therefore some of the following selections may show more "liberal" targets as we do not expect to be able to return to pristine water quality.

B.5.3.2 Step 2 – Identify Possible Sources of Stressors

Point source

- **Wastewater Treatment Plant**
- Manufacturing by-products
- Underground storage tanks
- **Septic tanks**

Non-point source

- Fertilizers
- **Land disturbance**
- **Urban/Suburban Runoff**
- Wildfires
- Invasive species
- **Sedimentation**
- Pesticides
- **Emerging Contaminants**

We identified a wastewater *treatment plant, septic tanks, land disturbance, urban/suburban runoff, sedimentation, and emerging contaminants* as possible stressor sources.

B.5.3.3 Step 3 – Identify Assessment Endpoints

- Water clarity
- **Phytoplankton**
- **HABs**
- Diatoms
- Benthic fauna
- SAV
- Epiphytes
- DO
- Invasive species
- **Algal toxins**

In this example, we determined that the important stressors relevant to restoring and maintaining fishable/swimmable water quality included *water clarity* and *dissolved oxygen*. Important assessment endpoints included *phytoplankton, HABs,* and *algal toxins*.

B.5.3.4 Step 4 – Gather Evidence

As discussed in the document, there are multiple types of evidence that may or may not be available when developing NNC. For this example, we chose to use reference conditions, stressor-response values, scientific literature, and stakeholder surveys. (The evidence used in this example was part of an existing group exercise and had no actual connection to a specific state or region.) In the following steps, many tables are used to show the pieces of evidence that were chosen. This allows us to be transparent about the process for selecting evidence, to document the selections, and to clearly communicate those decisions. This builds confidence in the evidence as well as the conclusions.

The first line of evidence we analyzed used distribution statistics from urban lakes (Table B-18). Data was prepared and we chose to use the 25th percentile of impaired lakes for TN, TP, and chlorophyll-a as potentially protective of recreational use. In this example, it is not known if these data are from inside or outside of our state or region.

Table B-18. Example 3- Reference Conditions

The 25th percentile of impaired lakes for TP, TN, and chlorophyll-a are highlighted.

		Growing Seasonal Values								
Population		**TP (mg/L)**			**TN (mg/L)**			**Chlorophyll (µg/L)**		
	N	25th	50th	75th	25th	50th	75th	25th	50th	75th
Reference Lakes	20	0.002	0.008	0.012	0.200	0.400	0.650	0.5	1.7	3.5
All lakes	210	0.003	0.016	0.030	0.300	0.800	1.200	0.8	3.4	6.5
Assessed Lakes Known to be Meeting Uses	24	0.004	0.011	0.020	0.400	0.550	0.800	1.0	2.3	4.7
Impaired Lakes	7	**0.012**	0.030	0.054	**0.600**	1.500	2.160	**3.0**	5.8	9.4

RESULT	**TP: 12µg/L**	**TN: 0.60mg/L**	**Chl-a: 3.0µg/L**

The second line of evidence used was stressor-response relationships (Table B-19). This evidence was prepared from models developed with national lakes data and from previously existing models developed with state data. We highlighted a range of liberal targets for recreational (fishable/swimmable) designated use.

Table B-19. Example 3- Stressor-Response Relationships

A range of liberal targets are highlighted.

				Stressor Concentration to Meet Target		
Response	**Response Target**	**Allowable exceedance probability**	**Certainty level (%)**	**TP (mg/L)**	**TN (mg/L)**	**Chl a (µg/L)**
Microcystin concentration	6	0.02	90			12.3
Microcystin concentration	8	0.02	90			**15.9**
Chlorophyll	12		90	0.019	0.46	
Chlorophyll	16		90	**0.024**	**0.51**	

Response	**Response Target**	Stressor Concentration to Meet Target		
		TP (mg/L)	**TN (mg/L)**	**Chlorophyll (µg/L)**
Chlorophyll	2	0.008	0.46	
Chlorophyll	5	0.028	0.72	
Chlorophyll	15	**0.072**	**2.1**	
R^2 (p-value)		0.6 (<0.05)	0.54 (<0.05)	
Cyano Density	20,000	0.03		10
Cyano Density	50,000	0.04		15
Cyano Density	100,000	**0.045**		**20**
R^2 (p-value)		0.42 (<0.05)		0.60 (<0.05)
Microcystis Density	20,000	0.061	0.92	12
Microcystis Density	50,000	0.048	1.10	17
Microcystis Density	100,000	**0.021**	**0.71**	**24**
R^2 (p-value)		0.41 (<0.05)		0.62 (<0.05)

Hypolimnetic DO	0	0.03	1	5
Hypolimnetic DO	2	**0.02**	**0.7**	**4**
Hypolimnetic DO	4	0.015	0.63	3.2
Hypolimnetic DO	6	0.005	0.23	1
R^2 (p-value)		0.50 (<0.05)	0.48 (<0.05)	0.53 (<0.05)

RESULT	**TP: 20-72µg/L**	**TN: 0.51-2.1mg/L**	**Chl-a: 4-24µg/L**

The third line of evidence used peer-reviewed scientific literature to associate nutrient and chlorophyll values with recreational uses (Table B-20). Applicable ranges were highlighted for recreational uses.

Table B-20. Example 3- Published Literature Selections

Applicable ranges for TP and Chl-a are highlighted.

Citation	Assessment Endpoint	TP	Chl
Schupp and Wilson (1993)	Peak coldwater fish abundance	0.006	1
Johnston et al. (1999)	Coldwater fish growth peak	0.009	6
Elliott et al. (1996)	Coldwater fish growth increase (England)	0.011	14

Adjacent State A 2010	TP	TN	Chl a (µg/L)
Coldwater	<0.012		<3
Coolwater	<0.020		<6
Recreation	**<0.030**		**<9**

Adjacent State B 2016	TP	TN	Chl a (µg/L)
High altitude	0.012		2.6
Low altitude, excellent aesthetics	0.017		3.8
Low altitude, good aesthetics	**0.018**		**7.0**

RESULT	TP: 18-30µg/L	TN: n/a	Chl-a: 7-9µg/L

Finally, we used stakeholder surveys as the last line of evidence (Table B-21). There were two studies available that showed a range of chlorophyll-a values for slightly impaired swimming conditions. Please note that only chlorophyll-a data was available, therefore there are no results in this section for TP or TN.

Table B-21. Example 3- Stakeholder Survey Results

Chl-a levels associated with slightly impaired swimming conditions are highlighted.

Study	Location	Surveyed Group	Respondent Ranking	Chl-a Level (µg/L)
Hoyer et al. (2004)	FL	Citizen lake monitors	Excellent for swimming (rank=1,2)	7 to 12 (mean) 2.5 – 10.5 (range+)
			Slightly impaired for swimming (rank=3)	**14 (mean)** 5 – 11 (range+)
			Undesirable (rank=4,5)	5 to 80 (mean) 2.5 – 110 (range+)

Study	Location	Surveyed Group	Respondent Ranking	Chl-a Level (µg/L)
Heiskary and Walker (1988)	MN	Agency staff	Excellent for swimming (rank=1,2)	5 to 10 ppb (mean) 2 – 17 ppb (range+)
			Slightly impaired for swimming (rank=3)	**45 (mean)** 15 – 60 ppb (range+)
			Undesirable (rank=4,5)	55 ppb (mean) 40 – 75 ppb (range+)

RESULT	TP: n/a	TN: n/a	Chl-a: 14-45µg/L

B.5.3.5 Step 5 – Weight evidence

Following the instructions laid out in the document, we are now ready to weight evidence (Table B-22). Note that all evidence provided addresses magnitude. None of it addresses duration or frequency. See the notes in Table B-22 for brief explanations of weighting decisions.

Table B-22. Example 3- Weight Evidence

Line of Evidence	Relevance	Reliability	Strength	Overall	Notes
Reference Conditions	+	0	0	0	Reference sites are difficult to find in urban settings. 25th percentile usually used at impaired sites.
Stressor Response	+	+	+	+	For state models, estimated response is to the mid-point of the stressor, and the more liberal targets were selected. Target selection is lenient.
Scientific Literature	++	++	++	++	Literature values are well-vetted and acceptable in other settings. Settings are not always specific to specific lake type.
Stakeholder Surveys	++	0	+	+	Very relevant for aesthetics. Highly subjective and variable results.

B.5.3.6 Step 6 – Weigh the body of evidence

The document describes different approaches for evidence integration (see Section 5.1.2). In Example 3, published literature had the "weightiest" evidence (Table B-23), so it could be used as the main resource for basing numeric nutrient criteria. However, since urban streams are often stressed, additional data could be collected to strengthen the stressor-response line of evidence to inform criteria selection and/or select new criteria where none has been previously set for lakes. This additional data from stressed urban lakes would increase confidence that stressor-response relationships were inclusive of the lake type.

Table B-23. Example 3- Evidence Summary

Line of Evidence	TP µg/L	TN mg/L	Chl-a µg/L	Notes
Reference	12	0.60	3.0	0

Line of Evidence	TP µg/L	TN mg/L	Chl-a µg/L	Notes
Conditions				
Stressor Response	20-72	0.51-2.1	4.0-24	+
Scientific Literature	18-30	n/a	7-9	++
Stakeholder Surveys	n/a	n/a	14-45	+
Candidate criteria	n/a	n/a	n/a	Additional stressor-response data will be collected to strengthen that line of evidence before NNC derivation

B.6 Summary

The examples above were illustrative of the NNC development process. They were based on evidence assembled within an existing exercise, but they should give a sense of the variety of decisions that could be made of NNC teams and how they affect conclusions. Still, they may seem "too easy" relative to the real world. Selecting data, analyzing data to generate evidence, weighting evidence, and integrating evidence in your own state may not be as straightforward. There will be additional factors such as ecoregional differences, overlapping trophic levels, lack of data, temporal distinctions, overabundance of data, evidence that seems out of date, etc. All of this will need to be considered as your state works through the NNC development process.

References

Canfield, DE; Bachmann, RW. (1981). Prediction of total phosphorus concentrations, chlorophyll a, and Secchi depths in natural and artificial lakes. Can J Fish Aquat Sci 38: 414-423. http://dx.doi.org/10.1139/f81-058

Dillon, PJ; Rigler, FH. (1974). The phosphorus-chlorophyll relationship in lakes. Limnology 19: 767-773. http://dx.doi.org/10.4319/lo.1974.19.5.0767

Elliott, JM; Fletcher, JM; Elliott, JA; Cubby, PR; Baroudy, E. (1996). Changes in the population density of pelagic salmonids in relation to changes in lake enrichment in Windermere (northwest England). Ecol Freshwater Fish 5: 153-162. http://dx.doi.org/10.1111/j.1600-0633.1996.tb00128.x

Heiskary, SA; Walker, WW, Jr. (1988). Developing phosphorus criteria for Minnesota lakes USA. Lake Reserv Manag 4: 1-10.

Hoyer, MV; Brown, CD; Canfield, DE. (2004). Relations between water chemistry and water quality as defined by lake users in Florida. Lake Reserv Manag 20: 240-248. http://dx.doi.org/10.1080/07438140409354247

Johnston, IA; Strugnell, G; McCracken, ML; Johnstone, R. (1999). Muscle growth and development in normal-sex-ratio and all-female diploid and triploid Atlantic salmon. J Exp Biol 202: 1991-2016. http://dx.doi.org/10.1242/jeb.202.15.1991

OECD. (1982). Eutrophication of waters: Monitoring, assessment and control. Paris, France: Organisation for Economic Co-operation and Development.

Schupp, D; Wilson, CB. (1993). Developing lake goals for water quality and fisheries. LakeLine 13: 18–21.

Smith, VH. (1982). The nitrogen and phosphorus dependence of algal biomass in lakes: An empirical and theoretical analysis. Limnology 27: 1101-1111. http://dx.doi.org/10.4319/lo.1982.27.6.1101

Soranno, PA; Bacon, LC; Beauchene, M; Bednar, KE; Bissell, EG; Boudreau, CK; Boyer, MG; Bremigan, MT; Carpenter, S. R.; Carr, JW; Cheruvelil, KS; Christel, ST; Claucherty, M; Collins, SM; Conroy, JD; Downing, JA; Dukett, J; Fergus, CE; Filstrup, CT; ... Yuan, S. (2017). LAGOS-NE: A multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of US lakes. Gigascience 6: 1-22. http://dx.doi.org/10.1093/gigascience/gix101

U.S. EPA. (2010). Using stressor-response relationships to derive numeric nutrient criteria [EPA Report]. (EPA-820-S-10-001). Washington, DC: U.S. Environmental Protection Agency, Office of Water. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P100IK1N.txt

U.S. EPA. (2019). National Aquatic Resource Surveys: Rivers and streams 2013-2014 (data and metadata files). Retrieved from https://www.epa.gov/national-aquatic-resource-surveys/data-national-aquatic-resource-surveys

U.S. EPA. (2022). Nutrient Scientific Technical Exchange Partnership & Support (N-STEPS) Online. Available online at https://nsteps.epa.gov/

EPA
United States
Environmental Protection
Agency

PRESORTED STANDARD
POSTAGE & FEES PAID
EPA
PERMIT NO. G-35

Office of Research and Development (8101R)
Washington, DC 20460

Official Business
Penalty for Private Use
$300