# Utah Lake Water Quality Study—Uncertainty Guidance

# FINAL

March 13, 2020
Version 8.0

## PRESENTED TO

**Utah Department of Environmental Quality**
**Division of Water Quality**
PO Box 144870
Salt Lake City, UT 84114

## PREPARED BY

**Tetra Tech**
1 Park Drive, Suite 200
Research Triangle Park, NC 2709

**Cover image**: Aerial View of Provo Utah with River Valley and Utah Lake, by Aqua Mechanical. Source file available at https://www.flickr.com/photos/aquamech-utah/24776739750/in/photostream/

# TABLE OF CONTENTS

# ABBREVIATIONS

| Abbreviation | Definition |
|---|---|
| ULWQS | Utah Lake Water Quality Study |
| SP | Science Panel |
| S-R | Stressor-response |

# 1.0 BACKGROUND

The goal of the Utah Lake Water Quality Study (ULWQS) is to evaluate the role of excess nutrients in the support of designated uses in Utah Lake, with a focus on development of in-lake water quality criteria that are protective of the lake's designated uses. ULWQS Science Panel (SP) is tasked with guiding water quality criteria development on Utah Lake by overseeing targeted scientific studies. The panel is working under a Charter, a set of operating principles including six significant tasks, and a set of high-level specific initial charge questions which are, at a distilled level: 1) What was the historic ecological and nutrient condition of Utah Lake pre-settlement and how has it changed?; 2) What is the current ecological and nutrient condition?; and 3) What additional information is needed? Any recommendation of a numeric value(s) to protect the lake or a response to a specific charge question should be accompanied by an estimate of certainty.

In support of this, the SP charter provides a list of specific objectives that define the SP duties. These objectives are primarily to: develop a scientifically defensible approach for criteria development, identify gaps in understanding, provide recommendations for scientific studies to fill any gaps, recommend and prioritize studies/analyses of existing data, review study workplans, guide the scientific research, oversee peer review of the studies, **develop a process to characterize uncertainty**, and finally to recommend science-based water quality criteria options to the Steering Committee.

As noted above, Objective 4 of the SP Charter specifically tasks the SP with developing a process to characterize scientific uncertainty including confidence of scientific findings and quantified measures of uncertainty, where possible. Uncertainty is inherent to any scientific study, and it is important to evaluate and communicate uncertainty to scientists, decision-makers, and the public in consistent, transparent, traceable, and understandable ways. It is also an important part of the weight of evidence process for evaluating relevant, strong, and reliable information. Considering multiple lines of evidence makes the decision process robust, but also presents a challenge for communicating different types of uncertainty associated with literature, mechanistic model and statistical model output.

This document provides guidance to help the SP identify, characterize and communicate uncertainty.

# 2.0 A STRUCTURE FOR EVALUATING UNCERTAINTY

Following guidance developed for high-profile environmental decision-making applications[1], the ULWQS SP will rely on two main metrics for communicating uncertainty associated with responses to charge questions and derivation of numeric targets:

- Qualitative expressions of confidence based on the type, amount, quality, and agreement (consistency) of evidence where that evidence may be literature, statistical analysis, mechanistic model output, or expert judgment. Such expressions could include "the SP has medium to high confidence in this finding given the high agreement among the medium amount of studies";
- Quantitative measures of uncertainty expressed probabilistically (based on statistical analysis, model results, or expert judgment). Such metrics take forms such as "90-100% probable".

These metrics will rely on three principal considerations of evaluating uncertainty: evidence and agreement, confidence, and likelihood.

[1] Mastrandrea, M.D., C.B. Field, T.F. Stocker, O. Edenhofer, K.L. Ebi, D.J. Frame, H. Held, E. Kriegler, K.J. Mach,P.R. Matschoss, G.-K. Plattner, G.W. Yohe, and F.W. Zwiers, 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC).

## 2.1 EVIDENCE AND AGREEMENT

The first consideration in communicating the validity of any statement of finding (e.g., a response to a charge question) is to characterize the evidence (as to type, amount, and quality) as well as the agreement among evidence underlying the finding or conclusion. The type of evidence refers to its derivation (e.g., literature, mechanistic model output, field observations, experimental evidence, or expert judgment). The amount of evidence refers to the quantity of independent evidence types. The quality of evidence refers to the rigor with which the evidence was derived. In previous applications of this approach, the terms “limited”, “medium”, and “high” have been used to describe the evaluating of evidence. The SP can decide how to weigh or combine these three elements into an assessment of the evidence. For example, one large, comprehensive, high quality study of the lake itself may constitute more evidence than results from several observational studies of dissimilar lakes. Finally, agreement refers to how results or conclusions among the different lines of evidence differ or concur and the terms “low”, “medium”, and “high” are used to describe agreement. Once again, the SP can decide what constitutes these qualitative statements of agreement. The amount and agreement of evidence form axes that define a space that informs estimates of confidence (Figure 1).

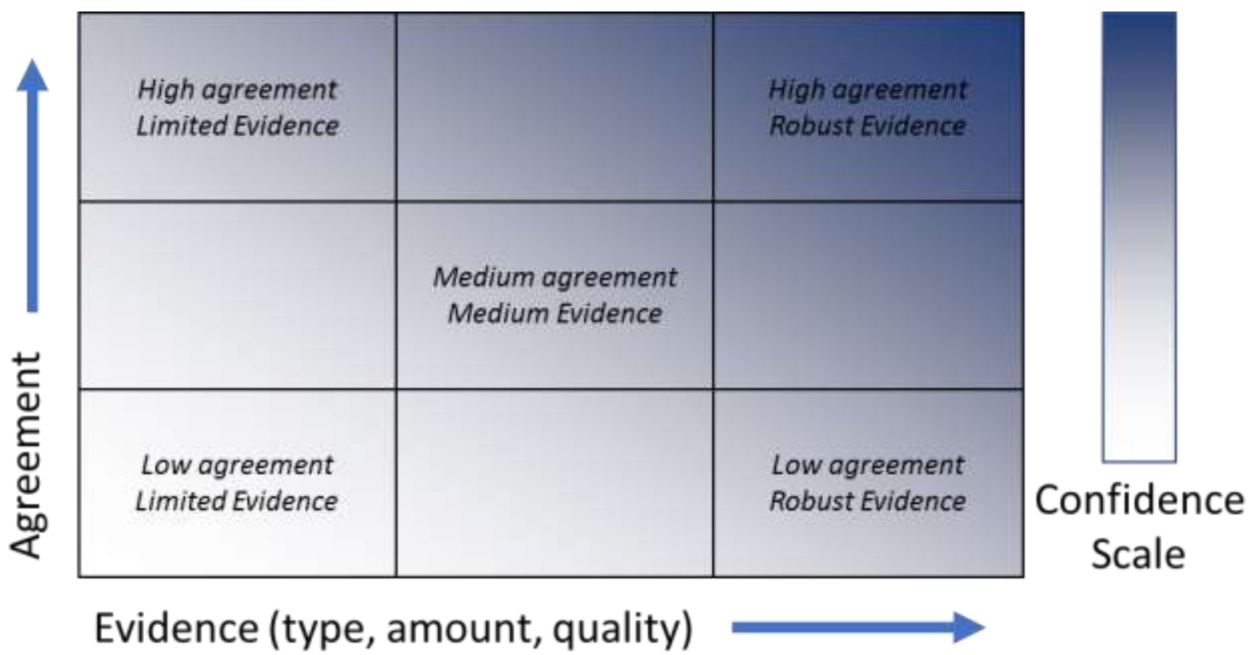


**Figure 1. A matrix for combining information on evidence and agreement for use in evaluating confidence in findings. From Mastrandea et al. 2010[1].**

The SP can deliberate about how best to quantify evidence and agreement, but some alternative options may inform this decision. As one alternative, the SP could develop very specific rules on the amount and quality of evidence (Table 1) and agreement (Table 2). Each line of evidence could be evaluated against these rules and a consensus decision made by the SP as to the quality of the evidence for any statement derived from these different lines.

**Table 1. Example of discrete rules for evaluating evidence quality**

| | | **Evidence Quality** | | |
|---|---|---|---|---|
| | | **Limited** | **Medium** | **High** |
| **Type** | | Other Scientific Studies of Lakes | • Mechanistic Model of Similar Systems<br>• S-R analyses for similar systems<br>• Reference based data<br>• Scientific Studies from similar systems | • Mechanistic Models of Utah Lake<br>• S-R analysis for Utah Lake |
| **Amount** | ***Mechanistic Model*** | 1 model run | 2-3 model runs | >3 model runs |
| | ***S-R Analyses*** | 1 independent analysis | 2-3 independent analyses | >3 independent analyses |
| | ***Scientific Literature*** | 1-2 studies | 2-4 studies | >4 studies |
| **Quality** | ***Mechanistic Model*** | 75% Variables meet Very Good calibration criteria | 75-90% Variables meet Very Good calibration criteria | >90% Variables meet Very Good calibration criteria |
| | ***S-R Analyses*** | • $p<0.20$<br>• Variance explained <30% | • $P<0.10$<br>• Variance explained 30 to 50% | • $P<0.05$<br>• Variance explained >50% |
| | ***Scientific Literature*** | • $p<0.20$<br>• Variance explained <30% | • $P<0.10$<br>• Variance explained 30 to 50% | • $P<0.05$<br>• Variance explained >50% |

**Table 2. Example of discrete rules for evaluating agreement.**

| | | Agreement | | |
|---|---|---|---|---|
| | | **Low** | **Medium** | **High** |
| **Amount** | | Half the lines of evidence agree | 75% of the lines of evidence agree | All lines of evidence agree |

Alternatively, the SP can decide to forego discrete rules and use their professional judgment along with a description of the lines considered to make an ad hoc determination of evidence quality and agreement for any recommendation or conclusion. Such a determination could be required to be based on some pre-set level of agreement among SP members (e.g., 100% consensus or a majority of SP members).

Having defined evidence and agreement, the next consideration is confidence.

## 2.2 CONFIDENCE

Any SP finding should be accompanied by a statement of confidence that is a qualitative expression based on the type, amount, quality and agreement among the evidence. For this context, confidence does not refer to probabilistic or statistical confidence (see likelihood below). Confidence is an expression of agreement among the SP as to the validity of a statement related to a finding. In other applications, confidence has been split into five levels: “very low”, “low”, “medium”, “high” and “very high”. As mentioned, the amount, quality, type, and agreement of evidence should inform statements of confidence in findings (Figure 1), but there is flexibility in this relationship because the content underlying any evidence and agreement statement may differ and, therefore, affect the resulting confidence statement. In addition, consensus among the SP members as to the amount, quality, type, and agreement of evidence is also a factor in determining confidence. In general, however, more evidence, greater agreement among that evidence, and greater consensus among the SP members would increase confidence.

As a concrete example, one could take the discrete rubrics of Table 1 and Table 2 and derive a specific matrix for Figure 1 (Figure 2). SP consensus could either be incorporated into the decisions on agreement and amount of evidence or alternatively added as another dimension to the decision (e.g., Figure 3).

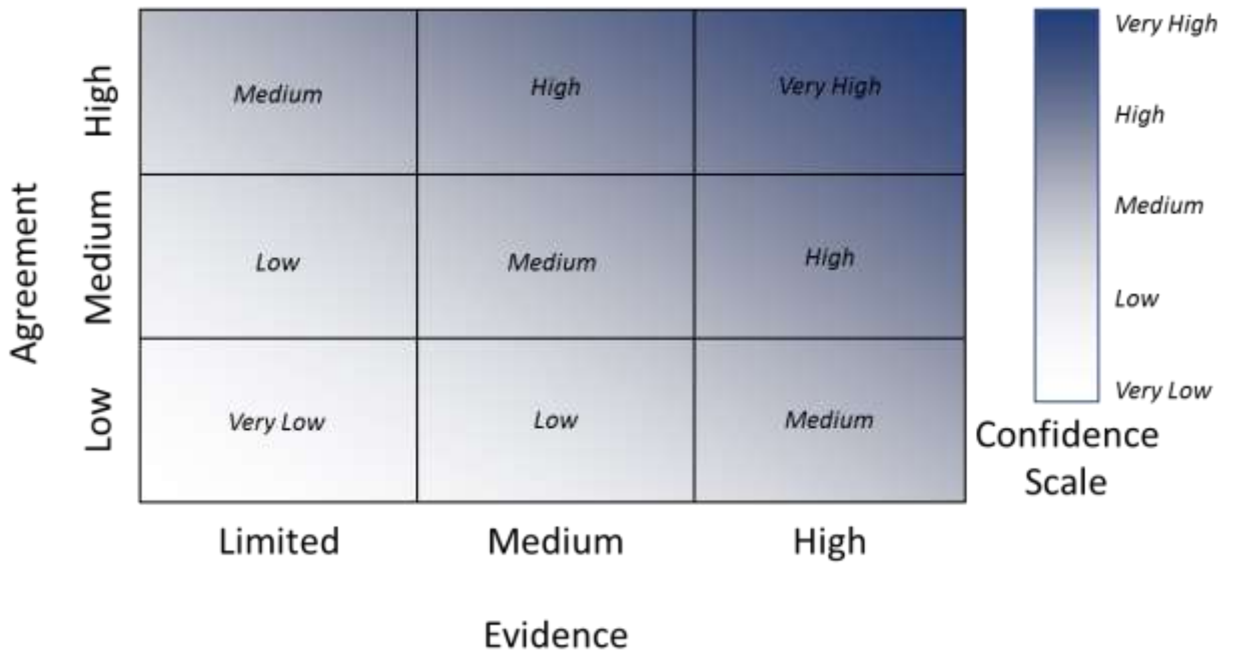


**Figure 2. Example of a discrete rule matrix for deciding confidence.**

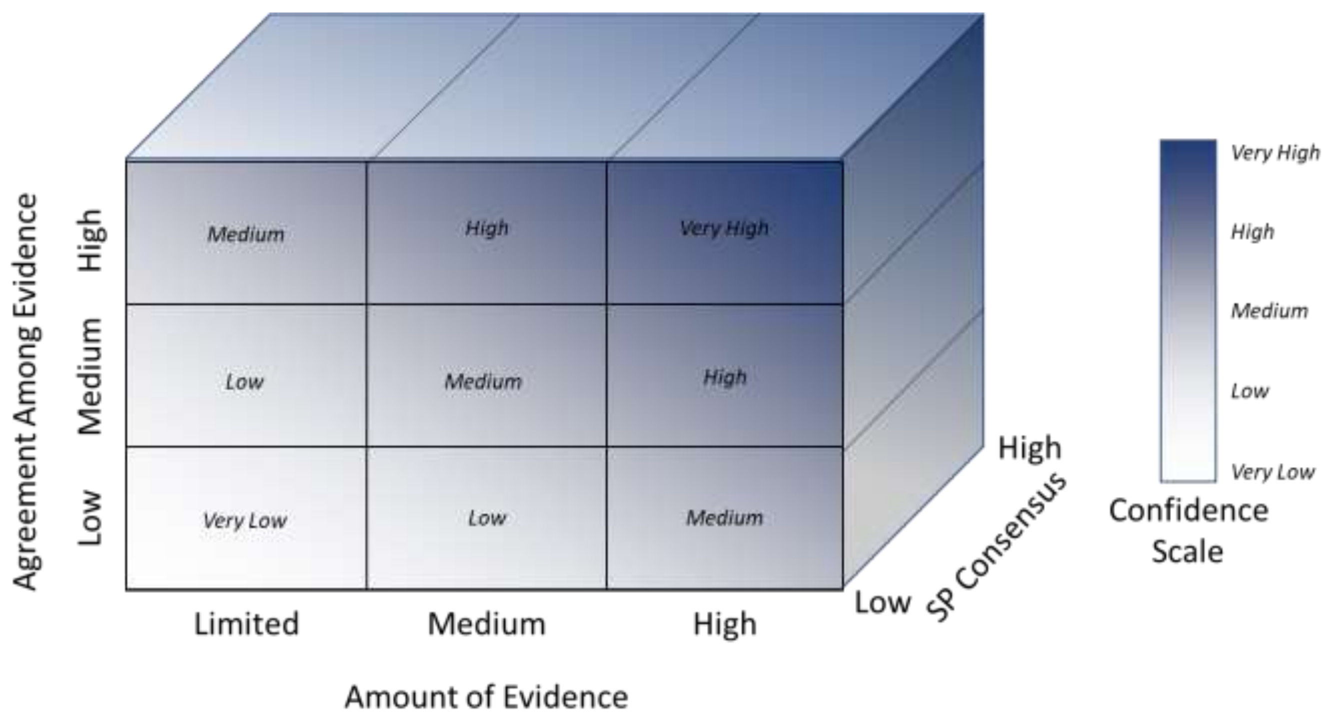


**Figure 3. Example of a discrete rule matrix for deciding confidence with consensus added as another dimension.**

In contrast, the SP can choose to make this judgment based on ad hoc decisions for each statement being considered using a vote of the SP members as described above for evidence and agreement.

## 2.3 LIKELIHOOD

The SP will communicate quantified uncertainty associated with a specific finding as a likelihood statement where quantitative estimates exist. Such quantified likelihood may be derived from statistical models, mechanistic models (where model confidence allows), or even expert elicitation[1]. One model for how likelihood statements can be related to probabilistic data is shown in Table 3.

**Table 3. Description of likelihood terms associated with specific probabilistic values. After Mastrandea et al. 2010[1].**

| Language | Probability |
|---|---|
| Virtually certain | 99-100% Probability |
| Very likely | 90-100% Probability |
| Likely | 66-100% Probability |
| About as likely as not | 33 to 66% Probability |
| Unlikely | 0-33% Probability |
| Very unlikely | 0-10% Probability |
| Exceptionally Unlikely | 0-1% Probability |

# 3.0 EVALUATING UNCERTAINTY IN INDIVIDUAL LINES OF EVIDENCE

The SP will rely on several independent lines of evidence in responding to charge questions or in making specific findings. In this section, we briefly discuss how these individual lines may be interpreted in terms of the structure described above.

## 3.1 EMPIRICAL STATISTICAL MODELS

Empirical statistical models are being applied by the SP to analyze observational data from Utah Lake and in the scientific studies being directed by the SP, all of which are being conducted in the context of answering specific charge questions. To the extent these analyses produce quantifiable estimates of uncertainty, they will inform likelihood estimates. Moreover, the type, quality, and amount of information from these studies as well as agreement among them related to specific findings will inform confidence statements. Again, the SP will decide how to specifically organize this information into the narrative statements described above, but they should be aware that such models will produce this type of evidence and be prepared to communicate this information to decision-makers using the structure above or similar construct. The SP will request that all scientific studies directed by the SP present their results in the context of the uncertainty elements discussed in this document.

## 3.2 MECHANISTIC MODELS

Mechanistic or process models are deterministic models built on algorithms that simulate natural processes. Because such models are deterministic, traditional statistical uncertainty parameters do not necessarily apply because these models do not have error independent of the input data.

However, there are well formed concepts for evaluating mechanistic model uncertainty and here we rely on those produced by the Council for Regulatory Environmental Modeling[2]. This framework defines model uncertainty as "*incomplete knowledge about specific factors, parameters (inputs), or models"*. This uncertainty is driven by natural, structural and conceptual uncertainty; the first being natural variability in important drivers (aleatory uncertainty), the second being uncertainty in the mathematical equations used to model phenomena, and the latter to a lack of knowledge about model pathways (the latter two forms of epistemic uncertainty). In the guidance, this uncertainty is further elaborated into three distinct but interrelated components:

- Model framework uncertainty: the soundness of the underlying scientific foundation
- Model input uncertainty: measurement errors, analytical imprecision and limited sample size during collection and treatment of calibration or validation (corroboration) data
- Model niche uncertainty: use of model outside its developed use.

Model framework and niche uncertainty are addressed through expert judgment and peer review as well as comparison of different model structures. Model input or parametric uncertainty is more quantitative and the guidance provides three primary approaches for evaluating model input/parametric uncertainty: corroboration, sensitivity analysis, and uncertainty analysis.

Corroboration assesses how well a model corresponds to reality and has both quantitative and qualitative approaches. Quantitative approaches principally involve confronting a model with data and are largely numeric and even statistical approaches that quantify how well model output fits observed data. Formal quantitative corroboration involves hypothesis testing for model acceptance, use of validation data, and quantitative tests against pre-determined performance criteria. Robust corroboration involves testing or validating the model with substantially different data then that used to calibrate the model. Qualitative corroborative approaches can be based on expert elicitation of model accuracy.

Sensitivity analysis and uncertainty analysis are sometimes used interchangeably and are very similar, but they do reflect unique elements. Sensitivity analysis refers to how well changes in input or assumptions affect model output and is described as a principal evaluation tool for characterizing the most and least important sources of uncertainty. This tends to be algorithm specific in respect to model variables. Sensitivity analysis is typically conducted using systematic variation in input conditions followed by output evaluation.

Uncertainty analysis, in the EPA guidance, refers to how a model is affected by lack of knowledge about a certain population or real parameter values. It may refer to uncertainty in the relationship of input variables to responses or also to uncertainty in what variables even influence responses. This type of analysis is parameter specific and can sometimes be reduced with additional study and data. This type of analysis is, however, more difficult to quantify and may involve more expert judgment to evaluate. There is a growing literature on methods to improve uncertainty analysis for deterministic models which include expert assessment, model sensitivity, model emulation, and data-based approaches[3]. Probabilistic sampling of input parameters using Monte Carlo

[2] USEPA. 2009. Guidance on the development, evaluation, and application of environmental models. Council for Regulatory Environmental Modeling, United States Environmental Protection Agency, Washington, DC. EPA/100/K-09/003

[3] Uusitalo, L., A. Lehikoinen, I. Helle, and K. Myrberg. 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. Environmental Modelling & Software 63: 24-31.

simulations (including Markov Chain and Bayesian versions) and Bayesian model averaging are all methods to improve model performance as well as improve evaluation of deterministic model uncertainty[4]. They do, however, often come with substantial computational demands, especially as model complexity increases. In any case, the SP will provide guidance on whether and which of these approaches would be appropriate to use for the resulting ULWQS model structure and how it will be used to characterize model uncertainty by any modeling team. The model team should present this uncertainty information in the context of the elements discussion in this document.

## 3.3 LITERATURE

There are many potential unknowns related to elements of the charge questions already provided and likely to be provided to the SP in the future. Existing data and targeted studies will provide one body of information from which to draw conclusions. In addition, there is a wealth of information stored in the scientific literature that is likely relative and informative to making findings in response to specific questions. This literature may corroborate or refute Utah Lake specific work or expert judgment. This literature will contribute evidence and will affect agreement among evidence; in so doing, the scientific literature will inform confidence in findings. Also, many of the studies in the scientific literature have quantified uncertainty in probabilistic terms and will therefore inform likelihood. The SP will be prepared to incorporate the information from the scientific literature into these measures of uncertainty.

The ecological risk assessment community has provided some useful models for how to weigh literature-based evidence[5]. These include weighing the relevance, strength, and reliability of literature deemed relevant to a question (Figure 4). Relevance refers to the correspondence between the system of interest (Utah Lake) and target system treated in the literature study. Strength refers to the degree of differentiation from control, reference or randomness provided in the study results. Literature that produces large effect signals with a high degree of effect across many study elements would be considered strong. Finally, reliability is a more difficult element to evaluate and includes consideration of: design, abundance of information, minimization of confounding effects, specificity of study, potential for bias, corroboration, peer review, and transparency. Some consideration of all three of these elements should be used to help evaluate the quality of literature-based evidence. Of special importance is that any search and review of literature attempt to be exhaustive and unbiased and that dissenting literature be considered, addressed and if, how, and why it affects uncertainty communicated.

---

[4] Camacho, R.A., Martin, J.L., McAnally, W., Díaz-Ramirez, J., Rodriguez, H., Sucsy, P., Zhang, S., 2015. A comparison of bayesian methods for uncertainty analysis in hydraulic and hydrodynamic modeling. JAWRA J. Am. Water Res. Assoc. 51 (5): 1372e1393.; Camacho, R.A., J.L. Martin, T. Wool, and V.P. Singh. 2018. A framework for uncertainty and risk analysis in total maximum daily load applications. Environmental Modelling & Software 101: 218-235; Gudimov, A., M. Ramin, T. Labencki, C. Wellen, M. Shelar, Y. Shimoda, D. Boyd, and G.B. Arhonditsis. 2011. Predicting the response of Hamilton Harbour to the nutrient loading reductions: a modeling analysis of the “ecological unknowns”. Journal of Great Lakes Research 37, no. 3: 494-506; Ramin, M., Stremilov, S., Labencki, T., Gudimov, A., Boyd, D., Arhonditsis, G.B., 2011. Integration of mathematical modeling and Bayesian inference for setting water quality criteria in Hamilton Harbour, Ontario, Canada. Environmental Modelling & Software 26, 337–353; Ramin, M., T. Labencki, D. Boyd, D. Trolle, and G.B. Arhonditsis. 2012.A Bayesian synthesis of predictions from different models for setting water quality criteria. Ecological Modelling 242: 127-145.

[5] USEPA. 2016. Weight of Evidence in Ecological Assessment. Risk Assessment Forum, United States Environmental Protection Agency, Washington, DC. EPA/100/R-16/001.

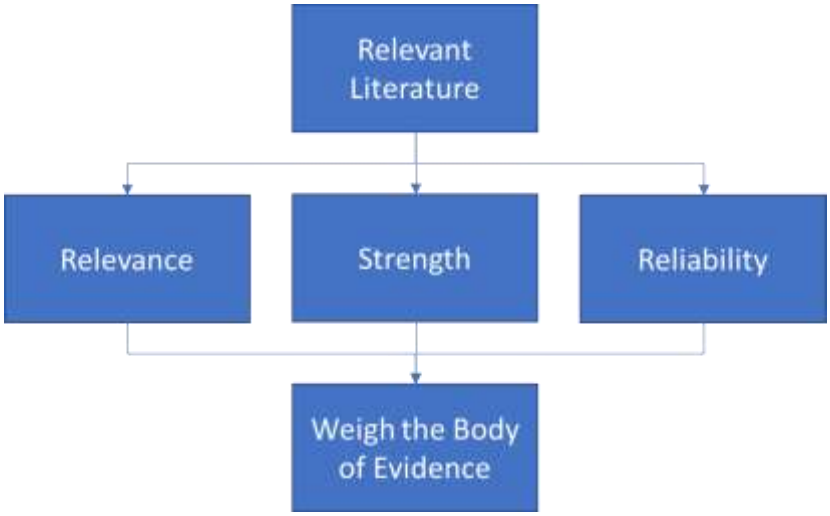


**Figure 4. Elements for weighing literature-based evidence. After USEPA 2016[5].**

## 4.0 COMMUNICATING OVERALL UNCERTAINTY

The SP will evaluate uncertainty associated with different lines of evidence using the elements described above as a facilitated discussion and will communicate the results of this discussion to decision makers consistently using the language on confidence and likelihood. This applies both to any communication involving individual lines of evidence or observations as well as to synthesis across lines of evidence in making specific findings, such as a response to a specific charge question, or on the protectiveness of a specific nutrient concentration, such as in recommending numeric criteria. In the case of specific charge questions to the SP, the SP will endeavor to get clarity for any question that is not formulated in a way that it can be addressed as to confidence or likelihood. For example, "*what was the historic condition of the lake*?" does not lend itself to an expression of confidence or likelihood, even though an answer can be given without an expression of certainty. In contrast, "*was the lake historically eutrophic defined using Carlson's TSI and thresholds for trophic state?*" is a statement that can be answered with a level of certainty. In the case of recommending numeric criteria, incorporation of uncertainty is discussed in the Framework document.

In addition to developing a consistent language in communicating uncertainty, the SP will endeavor to provide traceable accounts of their reasoning. A traceable account is a description of the type, amount, quality and degree of agreement among the individual pieces of evidence that constitute a specific finding. They provide transparency about conclusions that makes it possible for others to replicate and for decision-makers to clearly understand the basis for any finding. A traceable account could include the types of evidence used, standards for consideration and evaluation, how the evidentiary lines were combined, any assumptions, and an explanation of important conditions or factors influencing the finding.