Application of Weight-of-Evidence Methods to Water Quality Criteria

Glenn Suter
U.S. EPA
National Center for Environmental Assessment

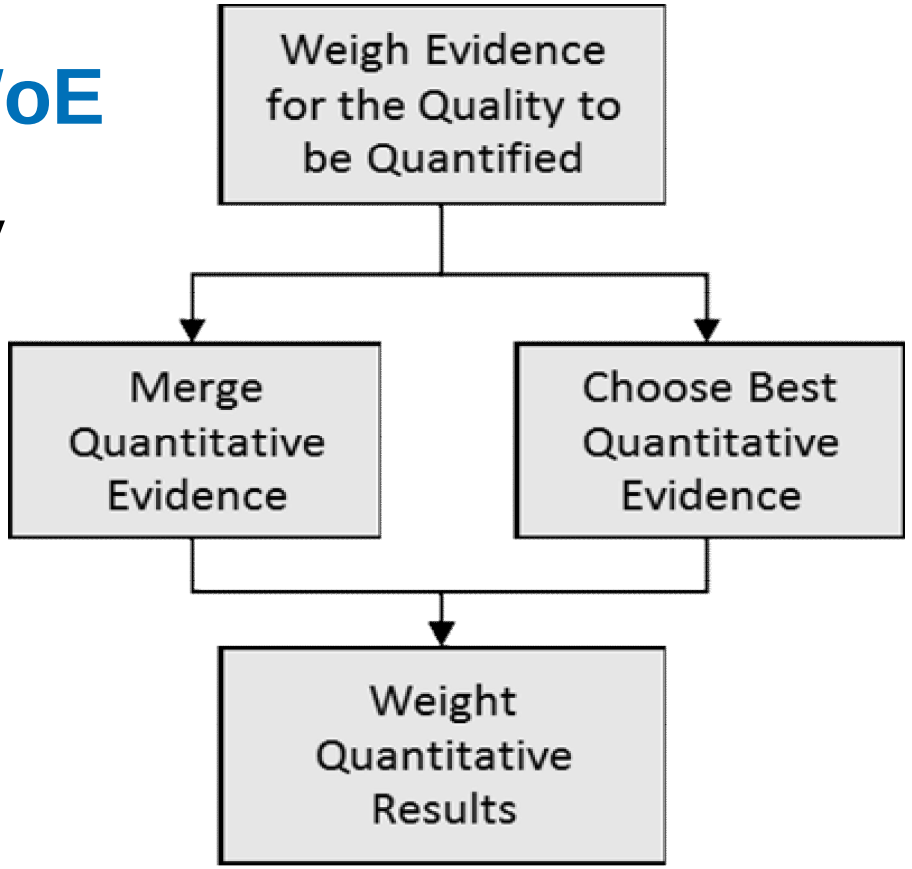
EPA United States Environmental Protection Agency

Weight of Evidence is:

- A summary interpretation of multiple pieces of evidence
- Based on the metaphor of the scales of justice
- Useful for assessing environmental **qualities**
  - Acts through dietary exposure?
  - Teratogen?
  - Cause of observed impairments?
- Also environmental **quantities**
  - Estimate given multiple values
- Risk Assessment Forum guidelines under development

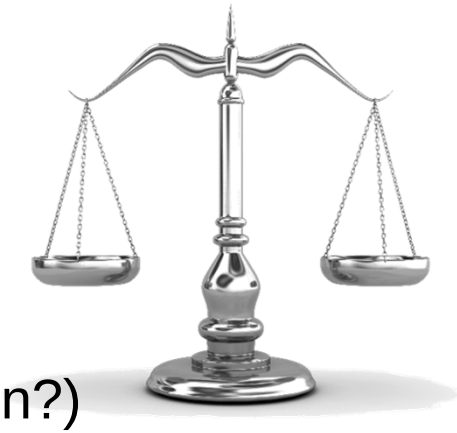
Qualitative with Quantitative WoE

- Qualitative WoE to determine what to quantify
  - Problem formulation
- Quantitative WoE to determine an estimate
  - Merge estimates
    - Meta-analysis
  - Weight to choose among estimates
- Qualitative WoE to determine confidence
  - Not just scatter



EPA United States Environmental Protection Agency

WoE for Problem Formulation



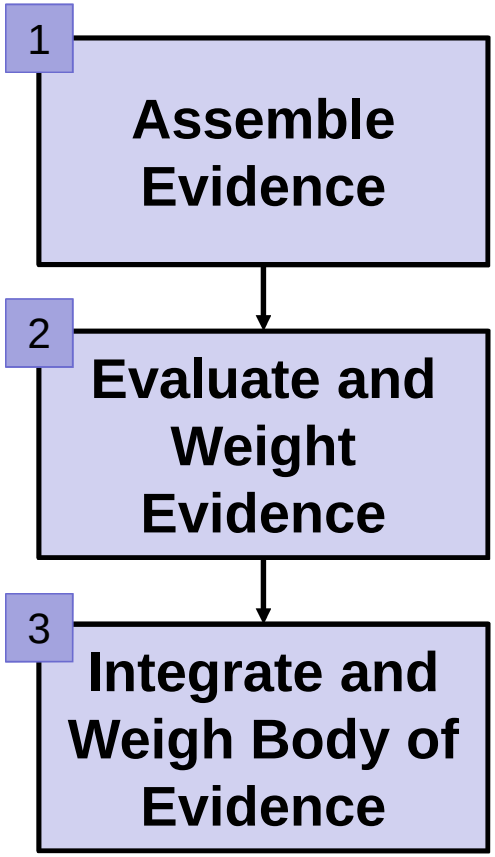
- Routine in Human Health Risk Assessment (Is it a carcinogen?)
- Endpoints in Integrated Science Assessments for Air Quality Criteria
  - Does ozone reduce timber production at ambient levels?
- Currently, not in Aquatic Life Criteria (ALC) Guidelines because they
  - Assume direct aqueous exposure
  - Assume critical effects are survival, growth, and reproduction
  - Until, as with Se, violations of assumptions cannot be ignored
- But WoE could be used in problem formulation for ALC

WoE and Field-Based Criteria

- Field-based benchmarks are realistic but
  - may not be causal
  - may be confounded
- OW guidance on causal assessment based on WoE (CADDIS)
  - Adapted to demonstrate that invertebrate extirpation is caused by major ions (SC)
  - Adapted to determine whether other potential causes are significant confounders
  - Favorably reviewed by the SAB and ET&C peer reviewers
- WoE may also be used to determine whether applicable to another region

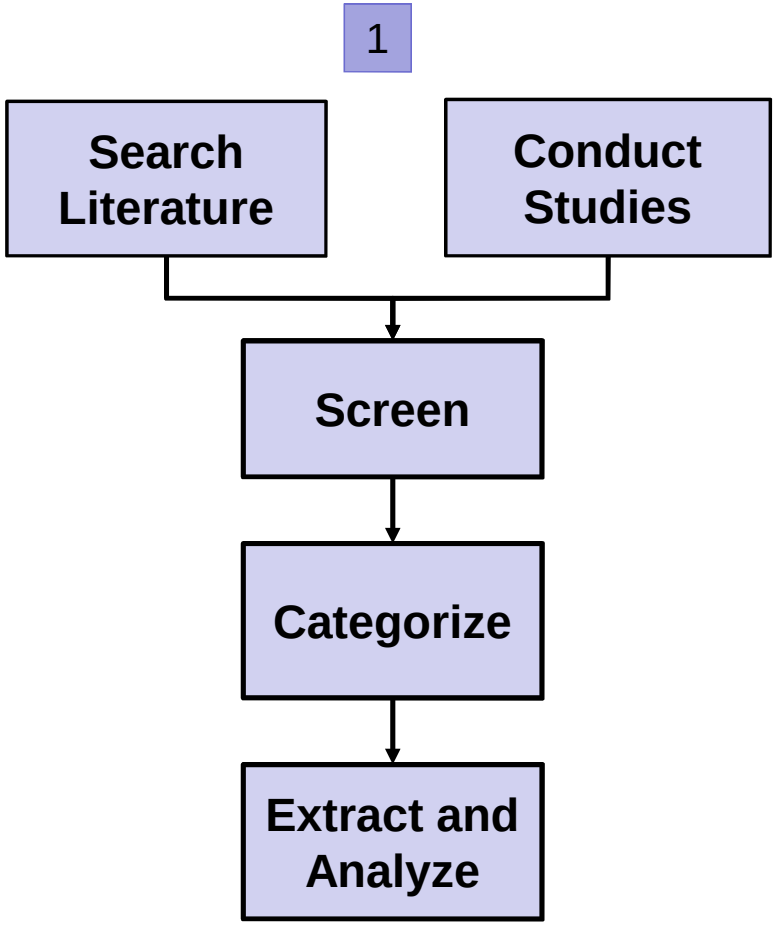
Framework for WoE

- Must be adapted to particular uses
- Needed for defensibility
  - Consistency
  - Transparency
  - Rigor
- Needed for practicality:
  - Minimum process complexity for the particular purpose and application



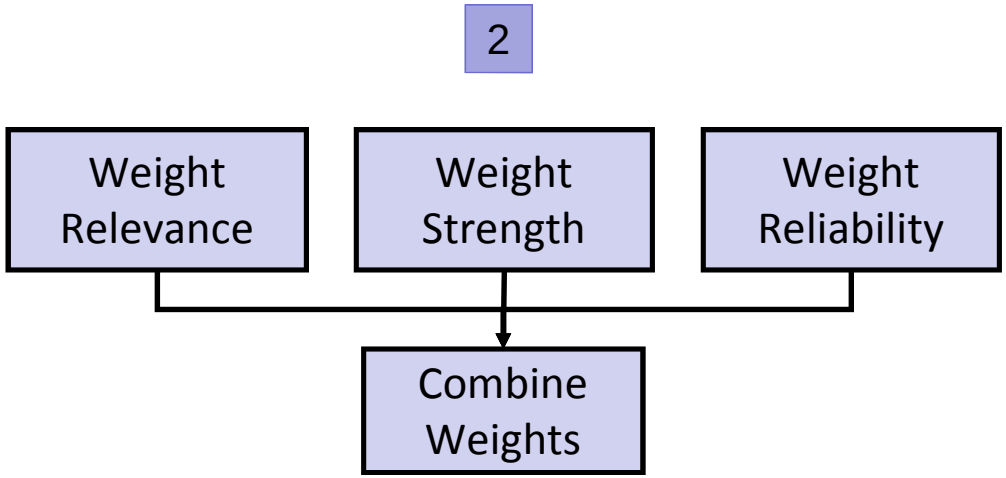
Step 1. Assemble Evidence

- Systematic review
  - Find all potentially relevant literature
  - Fill gaps with new studies
  - Screen for relevance & quality
  - Categorize so comparable
  - Extract data and analyze
    - Data are not evidence
    - Investigator's analysis often not usef

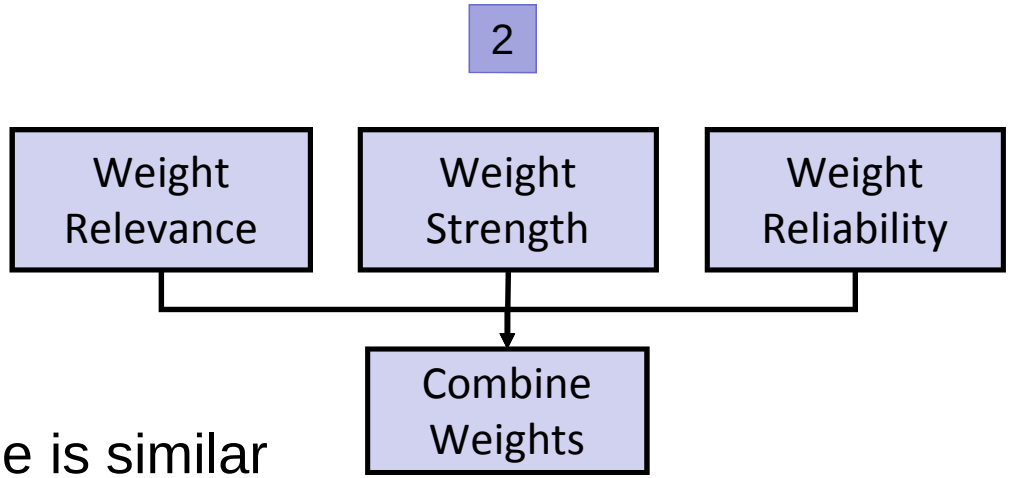


Step 2. Evaluate and Weight Evidence

- Not all evidence should be equally influential, so apply weights
- Three generally useful attributes
  - Relevance
  - Strength
  - Reliability



Step 2. Evaluate and Weight Evidence: How to



- Combine into weight for
  - a piece of evidence
  - a category of evidence
  - evidence for a criterion
- Weighting may be skipped if all evidence is similar
  - e.g., all evidence generated ad hoc or strict screening

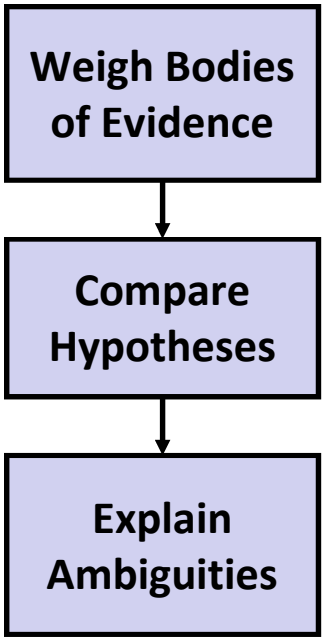
Example of a priori weighting rules:
field correlations of agents and biological responses

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**
	$\|r\| < 0.1$	**–**
	r has the wrong sign	**– –**

Step 3. Integrate and Weigh Evidence

3

- How well does evidence support a hypothesis?
  - Relative to other hypotheses
  - Relative to a standard of evidence
    - causal, likely, suggestive, not likely
- If body of evidence is ambiguous or discrepant:
  - List inconsistencies
  - Ask what would explain them
  - Find and evaluate evidence relative to the explanations



United States Environmental Protection Agency

Ideal WoE summary table

Summary of evidence for lead as a cause of mass mortality of tundra swans in the Coeur d'Alene River watershed (Norton et al., 2014) based on evidence from (URS Greiner Inc and CH2M Hill, 2011).	
Causal Characteristic	**Evidence**
Co-occurrence	Swan kills occurred in Pb-contaminated lakes and wetlands and not elsewhere in the region.
Sufficiency	Mortality occurred in laboratory tests at Pb doses and body burdens seen in dead or moribund swans in the field. Consistent mortality in the field at blood Pb levels >0.5 µg/g.
Time order	No evidence—no pre-mining information on swan mortality
Interaction	Dead and moribund swans had high blood and liver Pb levels. Pb-contaminated sediments were found in swan guts and excreta.
Specific alteration	Swans had pathologies characteristic of Pb, particularly enlarged gall bladders containing viscous dark green bile.
Antecedents	Spills of Pb mine tailings and atmospheric deposition from smelters account for the high sediment Pb levels.

WoE Summary Table

Weighing General Causation for SC and invertebrate extirpation (USEPA, 2011)		
Characteristic	**Body of evidence**	**Scores**
Co–occurrence	Loss of genera occurs where conductivity is high but is rare where conductivity is low	+++
Preceding causation	Sources of the ionic mixture are present and are shown to increase stream conductivity in the region	+++
Interaction	Aquatic organisms are directly exposed to dissolved ions. Based on first principals of physics, ionic gradients in high conductivity streams would not favor the exchange of ions across gill epithelia. Physiological studies over the last 100 years have documented the many ways that physiological functions of organisms are affected by the relative amounts and concentrations of ions (i.e., combinations of ions that some genera do not have mechanisms or the capacity to regulate;	++
Alteration	Some genera and other response metrics and assemblages are affected at sites with higher conductivity, whereas others are not. These differences are characteristic of high conductivity	+++
Sufficiency	Laboratory analyses report results of effects for a tolerant species, but test durations and most ionic compositions are not representative of exposure in streams. However, regular increases in effects on invertebrates with increased exposure to ions, based on field observations, indicate that exposures are sufficient	+++
Time order	Conductivity is high and extirpation has occurred after mining permits are issued, but conductivity and biological data before and after mining began are not available	NE
Summary of Body of evidence	Five characteristics supported and none weakening the body evidence that increases in conductivity causes extirpation of freshwater benthic invertebrates	Very Probable

EPA United States Environmental Protection Agency

WoE and Quantitative Criteria

- Concentration based on WoE
- Durations and recurrence frequencies based on WoE
- Weighted and Weighed SSDs
- Weighted and Weighed alternative methods
  - Microcosms, QSARs, Statistical models, Simulation models, …
  - SABS precedent but could do better

Weighing Evidence in SSDs—Meta-analysis

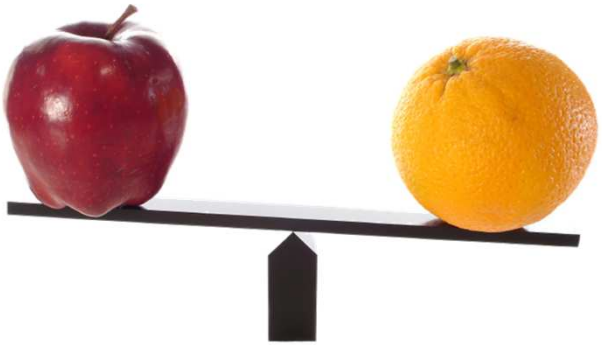
- Equal weight for tests and species using geometric means
  - SMAV = exp [(∑ log LC50)/n]
  - GMAV = exp [(∑ log SMAV)/n]
- Could weight tests based on quality, # of partial responses, etc.
  - SMAV = exp [(∑ (w_t log LC50))/ ∑ w_t]
- Could weight species based on number of tests
  - GMAV = exp [(∑ w_s log SMAV)/ ∑ w_s]

Weighted regression for SSDs

- Weighted GMAVs
  - Could weight by number of tested species
  - Could weight by representativeness
    - Tested species are not random draw from a community
  - Could weight by importance (opposite of representativeness)
    - Weight loss of a fish species more than an insect species
      - Larger proportion of taxon
      - Public values
  - Could weight by test quality, number of species, etc.
  - See Semenzin et al. 2015 in ET&C

Combining Multiple Types of Estimates

- Field and Lab, Experimental and Observational, Single and Multiple Species, SSD, and Aquatox models, …
- In theory, could derive weighted mean across types, but more likely
  - Provide a range of plausible estimates, or
  - Validate best estimate, or
  - Elucidate processes and factors behind variance in estimates



EPA United States Environmental Protection Agency

What WoE can Provide for Criteria

- Best or better estimate
  - By choosing from or combining multiple estimates
- Complete variance
  - By considering the distribution of estimates
- Relevance and Reliability of the estimate
  - By considering qualities of the evidence
- Determination of relevant hypotheses
  - By weighing evidence for route of exposure and mode of action
- Evaluation of model reliability and level of confounding