Overview of Evidence-Integration Systems

Lorenz Rhomberg, PhD, Fellow ATS

Gradient

GRADIENT

How did we get to this juncture?

- Older, "rules-based" frameworks (e.g., EPA 1986)
  - Presume relevance
  - Main question: Reliability of observation
  - But increasing MoA understanding and examples of species-specificity, dose-limitation
- Newer, "judgment-based" frameworks (e.g., EPA 2005)
  - Guidance on "factors" or "considerations"
  - Main question: "Sufficiency" of evidence for conclusions
  - But how to justify conclusions? Hold to objective standards?
  - Weed (2005) critique of loose use of "WoE"
- NRC review of EPA Formaldehyde
  - "Roadmap" stressing systematic processes
  - Need for "methodology" for WoE judgments
- NRC review of IRIS Process

GRADIENT

"WEIGHT OF EVIDENCE"

- As a metaphor

- As a method

    - Use all the data

    - Systematic evaluation

    - Aim at objective procedures that lay out the process of scientific professional judgment

        Question: *In view of incomplete and contradictory evidence, how compelling is the case for the existence and potential magnitude of risk?*

Caution: If we drop "WoE" and adopt "Evidence Integration," we need to be sure that the new term means something specific.

GRADIENT

Critical Reviews in Toxicology

informa healthcare

REVIEW

A survey of frameworks for best practices in weight-of-evidence analyses

Lorenz R. Rhomberg[1], Julie E. Goodman[1], Lisa A. Bailey[1], Robyn L. Prueitt[1], Nancy B. Beck[2], Christopher Bevan[3], Michael Honeycutt[4], Norbert E. Kaminski[5], Greg Paoli[6], Lynn H. Pottenger[7], Roberta W. Scherer[8], Kimberly C. Wise[2], and Richard A. Becker[2]

[1]Gradient Corporation, 20 University Road, Cambridge, MA, USA, [2]American Chemistry Council, NE, Washington, USA, [3]CJB Consulting LLC, [4]Texas Commission on Environmental Quality, MC-168, Austin, USA, [5]Michigan State University, East Lansing, USA, [6]Risk Sciences International, [7]The Dow Chemical Company, Midland, USA, and [8]Johns Hopkins School of Public Health

Abstract

The National Academy of Sciences (NAS) *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde* proposed a "roadmap" for reform and improvement of the

GRADIENT

Desiderata:

- Systematic Approach to Literature Inclusion
  - inclusion/exclusion criteria
  - not just positive or "featured" outcomes of studies

- Systematic and Consistent Review of Studies
  - established procedure, tabulation

- Evaluations of Study Strengths and Weaknesses
  - BUT – what to do with "lesser" studies?
    - omit? down-weight? interpret?
  - study design, power, confounders, potential problems

- *An Established Process for Evaluation and Combination into Inferences*
  - *Rigorous review of studies alone is not enough*

GRADIENT

Rules based vs. guided judgment: Sailing between Scylla and Charybdis

"JUDGMENT"   "RULES"

A *"Known Human Carcinogen"* is one for which the evidence is sufficient to conclude that it is a human carcinogen.

A *"Known Human Carcinogen"* is one for which, following the framework, one ends up in the "Known Human Carcinogen" box.

"STRUCTURED JUDGMENT"
- guided evaluations with recorded results
- Judgments are proposed explanations of the array of results
- Judgments are justified by citing basis and showing superiority over alternatives

GRADIENT

"WEIGHT OF EVIDENCE"

HOW MUCH DOES EVIDENCE WEIGH?

i.e., what process for the evaluation of evidence provides a means for judging how compelling it is and how to trade off among apparent contradictions?

- "Evidence" has no meaning except in relation to a specific hypothesis

- The *hypothesis* is evaluated w.r.t. the evidence (not the other way around)

- ***So WoE should be organized around evaluating specific hypotheses against data***

GRADIENT

INTEGRATION:

Two Kinds of Inferences from Multiple Studies

- Multiple observations of the thing of interest itself
  - e.g., multiple epidemiologic studies; Evidence-Based Medicine on studies of treatment efficacy
  - Main question is consistency and reliable observation
  - "Weight" from methodologically and statistically reliable measurements
- Indirect evidence of related or relevant phenomena in other systems
  - e.g., animal bioassays, MoA information
  - Main question is relevance and how to generalize
  - Need to integrate across evidence that is relevant in different ways
  - "Weight" from support of relevance arguments

GRADIENT

Articulate an Hypothesis

What is the *proposed basis* for inferring that a particular phenomenon seen in studies of a chemical's effects will also happen in target populations?

- May be hypothesized MoA -- or general ("animals predict humans")

- A *generalization*, not just an extrapolation (should apply to all cases within its realm)

- What manifestations of the hypothesis are expected and not expected?   Check against <u>all</u> actual observed results.

GRADIENT

"Hypotheses" in HB-WoE

- "Hypotheses" are this kind of reasoning for why lines of evidence constitute "evidence"
    - (=/= the overall question, but rather lines of evidence regarding the overall question)
    - Hypotheses are multi-layered

- Hypotheses can be:
    - Mechanistic (underlying MoA applies to target population)
    - Empirical (positives in this system associated with endpoint of concern in other studies)
    - Hypothetical (reasonable speculation, but lack of positive evidence counts against their credence)

GRADIENT

Key WoE Questions

- Based on observed positives, what hypothesized causal processes are necessary? Sufficient?

- How do they generalize? What *other* manifestations should they have?

- If hypothesis were wrong, how *else* would one explain the array of outcomes?

GRADIENT

For Observed Outcomes that are Candidates for "Evidence"

- Why we think they happened where they did.

- Why we think they *didn't* happen where they *didn't*.

- Why we think the "did-happen" factors would also apply to the target population.
    - Might apply? Probably apply? Known to apply?

- Are there discrepant observations, and if so, how do we account for them?

- Are our "whys"
    - Observable underlying causes?
    - Reasonable guesses based on wider knowledge, other cases?
    - *Ad hoc* assumptions without evidence, needed to explain otherwise puzzling phenomena?

GRADIENT

Evaluate How Compelling the Case is for the Proposed Basis of Human Risk in View of:

"**Predictions**" of hypotheses that are confirmed in the observations

- More weight to "risky" and specific predictions

- Less weight when subsidiary assumptions or explanations are needed

- Both Positive and Negative predictions!

Apparent Refutations (counterexamples)

- Failure to repeat result across studies

- Non-responding sexes or species

- Unpredicted but clearly relevant phenomena

An hypothesis can often be reconciled with apparent refutations by either modifying it or adding subsidiary assumptions – but this entails a weight "penalty"

GRADIENT

Relative Credence in Competing Accounts

- "Account" = an articulated *set* of proposed explanations for the *set* of observations
  - Relevant Causation – but also chance, error, confounding factors, general-knowledge possibilities, plausible assumptions, assertions of irrelevance, and "unknown reasons"

| Certain Findings Indicate Target-Population Risk	Those Findings Do <u>Not</u> Indicate Target-Population Risk
• reasoning why • how contradictions resolved • why assumptions reasonable	• reasoning why *not* • how findings are *otherwise* explained • why assumptions reasonable

Can we measure the weights?

GRADIENT

NRC IRIS Process Review

"Rather than organize the narrative around a checklist of criteria, such as the Hill criteria, EPA might consider organizing the narrative as an *argument* for or against hazard on the basis of available evidence. It should be qualified by explicitly considering alternative hypotheses, uncertainty, and gaps in knowledge." (NRC, 2014, p.105)

GRADIENT

Sir Austin Bradford Hill on the Hill Criteria
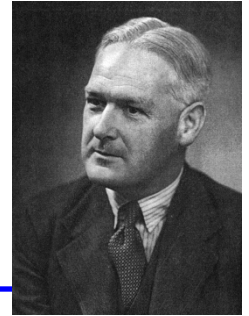
"... *the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?"* A. Bradford Hill (1965) *Proc Roy Soc Medicine* 58:295.

"set of facts" =
- all the epi (+ and -)
- mode of action
- animal studies
- other potential explanations

GRADIENT

Some Pitfalls:

- Overreliance on Conventions, Heuristics

- Narrative, citing "consistent" information

- Failure to address null or contradictory findings on same endpoints (especially when combining endpoints)

- Multiple Comparisons (choosing from among many parallel alternative analyses of same data)

- Mutually contradictory explanations of individual study results

- *Ad Hoc* Explanations (distinguish discovery from support)

- Accommodation

GRADIENT

Combining Realm-Specific <u>Judgments</u> (about epi, about animal bioassays, about mechanistic information) vs. Making and Evaluating <u>Inferences Across </u>All Realms of Evidence

Bringing Individual Study Quality to Bear on Evidence Integration

Applications
of Hypothesis-Based Weight of Evidence (HBWoE)

- Chlorpyrifos neurodevelopmental toxicity
  › Prueitt, RL; Goodman, JE; Bailey, LA; Rhomberg, LR. 2011. *Crit. Rev. Toxicol.* 42(10):822-903.
  › Goodman, JE; Prueitt RL; Rhomberg, LR. 2012. Dose Response 11(2):207-219.

- Methanol carcinogenicity
  › Bailey, LA; Prueitt, RL; Rhomberg, LR. 2012. Re*gul. Toxicol. Pharmacol.* 62:278-291.

- Dioxins thyroid hormone perturbation
  › Goodman, JE; Kerper, LE; Petito Boyce, C; Prueitt, RL; Rhomberg, LR. 2010. *Regul. Toxicol. Pharmacol.* 58(1):79-99.

- Formaldehyde as a leukemogen
  › Rhomberg, LR; Bailey, LA; Goodman, JE; Hamade, AK; Mayfield, DB. 2011. *Crit. Rev. Toxicol.* 41(7):555-621.

- Naphthalene carcinogenicity
  › Rhomberg, LR; Bailey, LA; Goodman, JE. 2010. *Crit. Rev. Toxicol*. 40(8):671-696.

- Methylmethacrylate nasal toxicity
  › Pemberton, M; Bailey, EA; Rhomberg, LR. 2013. Regul. Toxicol. Pharmacol. 66(2): 217-233.

- Toluene Diisocyanate carcinogenicity
  › Goodman, JE; Prueitt, RL; Rhomberg, LR. 2013. *Crit. Rev. Toxicol*. 43(5):391-435.

GRADIENT

Publications on the HBWoE Method

- Rhomberg, LR. 2014. "Hypothesis-Based Weight of Evidence: An Approach to Assessing Causation and its Application to Regulatory Toxicology." *Risk Analysis*. doi: 10.1111/risa.12206.

- Rhomberg, LR; Goodman, JE; Bailey, LA; Prueitt, RL; Beck, NB; Bevan, C; Honeycutt, M; Kaminski, NE; Paoli, G; Pottenger, LH; Scherer, RW; Wise, KC; Becker, RA. 2013. "A Survey of Frameworks for Best Practices in Weight-of-Evidence Analyses." *Crit. Rev. Toxicol.* 43(9):753-784.

- Lutter, R; Abbott, L; Becker, R; Borgert, C; Bradley, A; Charnley, G; Dudley, S; Felsot, A; Golden, N; Gray, G; Juberg, D; Mitchell, M; Rachman, N; Rhomberg, L; Solomon, K; Sundlof, S; Willett, K. 2014. "Improving Weight of Evidence Approaches to Chemical Evaluations." In Press in *Risk Analysis*.

GRADIENT