Hypothesis-Based Weight of Evidence:

An Approach to Assessing Causation and its Application to Regulatory Toxicology

Lorenz Rhomberg, PhD FATS
Gradient

Methods for Research Synthesis:
A Cross-Disciplinary Workshop
Harvard Center for Risk Analysis
October 3, 2013

GRADIENT

RISK ASSESSMENT: Bringing to bear existing scientific information—*and critical interpretation of that information*—on questions of the existence, nature and magnitude of risks that may be posed by exposure to an agent.

PROBLEMS:

- Incomplete information
- Indirect information (extrapolation)
- Inherently unobservable phenomena are of concern
- Contradictory information
- Alternative explanations (with different risk consequences) are possible
- Public process conducted in the face of contending interests

The general WoE Question is as big as "Science," but the application to Chemical Regulation entails some special features:

- The regulatory process cannot sustain pure science's suspension of judgment until ultimate resolution. A basis for actions is needed.
  - Skepticism and consideration of alternatives as part of the scientific method
  - diversity in interpretations among scientists
- The regulatory process needs "**findings**," and judgment is **delegated** to particular people tasked with representing the larger body of informed scientific opinion.
  - So, whose judgments and how they are justified become key

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"
- NAS review of EPA Formaldehyde
  - "Roadmap" stressing systematic processes
  - Need for "methodology" for WoE judgments

Critical Reviews in Toxicology

http://informahealthcare.com/txc
ISSN: 1040-8444 (print), 1547-6898 (electronic)

Crit Rev Toxicol, Early Online: 1–32
 DOI: 10.3109/10408444.2013.832727

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

Keywords

Data integration, human relevance, mode of action, risk assessment, systematic review

GRADIENT

WoE "Frameworks" aimed at Specific Evaluations

- Guidance-like, procedural, specified operations and structured evaluations based on stated rules
- Aim at capturing principles of valid scientific inference into rules that apply to the question at hand
    - Rules become standards that analysts can be held to
    - Aim at objective, operational analysis independent of the judge
    - Often with lists of "principles" or "considerations"
- Challenge: Automating "judgment"
    - Too prescriptive → lose credibility, become conventionalized
    - Too unstructured → lose warrant, question whose judgment?

GRADIENT

Systematic Presentation and Review of Relevant Data

- Not just positive results from positive studies
  - Also null results from same and other studies
  - Selection / Omission criteria explicit
- Consistent evaluation criteria
  - Design soundness, rigor, statistical power
  - Reliability (aka “internal validity”)
    - According to standards of field
    - According to needs of the application
  - Relevance (aka “external validity”)
    - … largely a question of interpretation, so intermediate between Phase 1 and Phase 2
- Other “relevant” data – historical controls, understanding of endpoints and MoA, basis for understanding biology, similar agents, etc.

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

General Kinds of Evidence

- Observed **toxicity process that represents an instance** of a more general one that would operate in parallel in the target population
- Observed **biological perturbation** or effect that represents a candidate element of a possible MoA that might operate in the target population
- Evidence by **correlation** of the study outcome with the target population toxicity of concern in other cases
- Evidence by **analogy** with other similar cases

The Span of Generalization

- We observe particular instances, but what makes them relevant is the potential for *generalization* – that other settings (including the target population) might have similar causal processes.
- What is the span of generalization? What are its limits? Assessing this is part of the WoE.

Some of our questions:

- What makes data "evidence"?
- What does evidence "weigh"?
- How much evidence is "enough"?

GRADIENT

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

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?

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?

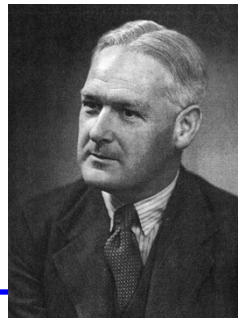
Relative Credence in Competing "Accounts"

- "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 Not 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?

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

Applications

of Hypothesis-Based Weight of Evidence (HBWoE)

- Chloryprifos 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.