“Uncertainty Characterization: The Role of Hypothesis-Based Weight of Evidence"

Lorenz Rhomberg, PhD
Gradient Corporation
lrhomberg@gradientcorp.com 617-395-5552
Outline

1. Uncertainty in QRA
2. WoE
3. Hypothesis-Based
4. WoE
5. Epistemology
QRA Challenges

- Many choices of datasets, dose measures, models, low-dose extrapolations, …
- Just doing “most sensitive” no longer acceptable
  - calls for uncertainty characterization
  - most sensitive may not be most supportable or representative
  - why should one dataset represent the whole of the agent’s potential carcinogenicity?
  - ranges (as with TCE) are problematic with risk managers
- How to weigh alternatives? How much information to move from defaults?
Main Components of Uncertainty

- Is each potential endpoint a human-relevant hazard?
- Choice of dataset to represent human risk

- D-R model fit
  - Uncertainty in response measure
  - Uncertainty in dose measure
- Low-dose extrapolation of high-dose effects
- Toxicologic equivalency of exposures across species
Expert Judgment Assessment of Chloroform Carcinogenic Potency Probability Tree

Evans et al. (1994) Reg Toxicol Pharmacol 20:15-36
Individual Expert’s Distributions
for excess cancer risk from 100 ppb Chloroform in Drinking Water

- Expert Opinion Varies Among Experts
- We Seek an Objective, Transparent, Data-Driven Means to Assign Weights

Evans et al. (1994)
Reg Toxicol Pharmacol 20:15-36
Where QRA Uncertainty Gets Stuck

Significant Qualitative Uncertainties that do not fit easily into “statistical distribution” characterization (questions of bearing, relevance, extrapolability)

Qualitative questions complicate even straightforward uncertainty approaches (through questions about model misspecification)

And they are the most consequential elements of the overall uncertainty!
FOR HUMAN RISK ESTIMATION:

What specific **endpoints** are:

Expected in Humans? Likely? Possible?

What **datasets** best represent them?

What **models** provide an appropriate basis for generalizing from studied responses to human risk projections?
Outline

1. Uncertainty in QRA
2. WoE
3. Hypothesis-Based WoE
4. Epistemology
Weight-of-Evidence for Human Carcinogenicity (Hazard ID)

- Douglas Weed – “Weight of Evidence” has several definitions. Need to specify what is meant and specify criteria
- Christina Rudén – Different organizations when evaluating the same set of (TCE) data come to different conclusions about hazard
Weight-of-Evidence for Human Carcinogenicity (Hazard ID)

- An instance in which expert judgment about qualitative questions is routinely used

- 1986 Guidelines – Too Prescriptive (need to consider MoA, synthesis across studies)

- 2005 Guidelines – Too Unstructured?
  - “Factors” to consider, but how to synthesize? How to weigh + / - ?
## Weight-of-Evidence Factors

### Increase Weight
- Number of independent studies with consistent results
- Same site across species, structural analogues
- Multiple observations
  - Species
  - Sites
  - Sexes
- Severity and progression of lesions
  - Early in life tumors/malignancy
  - Dose response relationships
  - Lesion progression
  - Uncommon tumor
- Route of administration like human exposure

### Decrease Weight
- Single study
- Inconsistent results
- Single site/species/sex
- Benign tumors only
- High background of incidence tumors
- Route of administration unlike human exposure

---

**BIOASSAY FACTORS**
Weight-of-Evidence for Human Carcinogenicity (Hazard ID)

- An instance in which expert judgment about qualitative questions is routinely used

- 1986 Guidelines – Too Prescriptive (need to consider MoA, synthesis across studies)

- 2005 Guidelines – Too Unstructured?
  - “Factors” to consider, but how to synthesize? How to weigh + / - ?
  - Role of negative studies
  - Several weak endpoints → “Likely”?

- How to argue conclusions against evidence?
INTERNAL VALIDITY
(Rigor)

EXTERNAL VALIDITY
(Bearing)
“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 of human cancer risk?
“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 its relation to a specific hypothesis
- It is the hypothesis that is evaluated w.r.t. the evidence (not the other way around)
- So WoE should be organized around evaluating specific hypotheses against data
## Endpoint-by Endpoint Evaluation of Data

**EPIDEMIOL.**
- A
- B
- B
- C
- D
- E

**BIOASSAY**
- B
- B
- B
- F
- -

**MoA**
- X
- -
- Y
- Z

**WoE for Endpoint “B”**
Endpoint-by Endpoint Evaluation of Data

- Systematic, endpoint-by-endpoint evaluation
- Note all results, not just “supportive” ones
- Evaluate w.r.t. the hypothesized basis for human prediction of risk
- Result: a WoE evaluation for each endpoint of interest

WoE for Endpoint “C”
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 environmentally (or occupationally) exposed humans?

- animal studies *and* human studies
- general scientific understanding and experience with other agents
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 environmentally (or occupationally) exposed humans?

what commonality of:

- material bases?
- causal processes?
- progressions of events?
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 environmentally (or occupationally) exposed humans?

- *a generalization*, not just an extrapolation (should apply to all cases within its realm)
- issue is to define that realm and what manifestations of the hypothesis are expected and not expected – which can then be checked against all the actual observed results (not just the source study to human extrapolation)
Evaluate how compelling is the case for the proposed basis for human risk in view of:

"predictions" of hypothesis that are confirmed in the observations

- more weight to "risky" predictions
- more weight to specific predictions
- less weight when subsidiary assumptions or explanations needed
Evaluate how compelling is the case for the proposed basis for human risk in view of:

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 modifying it or by adding subsidiary assumptions – but this entails a weight "penalty"
Hypotheses (and the human risk projections they entail) are compelling to the extent that:

- they are compatible with existing data
- they explain patterns among existing data
- they comport with and emerge from our knowledge of carcinogenesis and biology in general
- they are predictive and not just accommodative
- competing hypotheses aren't also compatible/explanatory
- apparent refutations can be reconciled using plausible, likely explanations
### Endpoint-by Endpoint Evaluation of Data

**EPIDEMIOL.**
- A
- B
- C
- D
- E

**BIOASSAY**
- B
- B
- F

**MoA**
- X
- Y
- Z

- Systematic, endpoint-by-endpoint evaluation
- Note all results, not just "supportive" ones
- Evaluate w.r.t. the hypothesized basis for human prediction of risk
- Result: a WoE evaluation for each endpoint of interest

**WoE for Endpoint “C”**
Evaluating Hypotheses Against the Data

- Point is not to prove or disprove – which is impossible
- Rather, to gauge plausibility relative to other endpoints and other cases
- Weights inversely proportional to how much one must fill in gaps with defaults and/or *ad hoc* explanations of apparent refutations
- How much stretching of credibility is needed? How much faith in assumptions is required? How much does the hypothesis explain that would be unexplained without it?
Can (and should) also evaluate hypotheses about lack of human carcinogenic risk

- how to reconcile assertion of lack of human risk with positive studies and the general presumption that animals and humans share fundamental physiology and control of cell division and differentiation?

- how compelling is the basis for assertion of human risk versus the basis for assertion of lack of risk – gauged by their abilities to account for the array of all the relevant data
For Each Hypothesis…

Kidney tumors secondary to high-dose cytotoxicity due to reactive thiol

<table>
<thead>
<tr>
<th>PREDICTED</th>
<th>SUPPORTING</th>
<th>SUPPORTING WITH SIGNIFICANT AD HOC ASSUMPTIONS</th>
<th>NEITHER SUPPORTING NOR REFUTING</th>
<th>REFUTING UNLESS ACCEPT SIGNIFICANT AD HOC EXPLANATIONS</th>
<th>REFUTING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bioassay Rats w kidney tumors have kidney cytotoxicity</td>
<td>Some Strains of Rat have fewer tumors and females have fewer (sensitivity differences? metabolic differences?) Human studies w kidney tumors only at very high exposures (high enough for cytotoxicity?) Humans seem to produce reactive thiol (enough?)</td>
<td></td>
<td></td>
<td>Bioassay Mice have kidney toxicity but no kidney tumors (less sensitive to cytotoxicity? why?)</td>
<td></td>
</tr>
<tr>
<td>Humans seem to produce reactive thiol (enough?)</td>
<td></td>
<td></td>
<td></td>
<td>Most human studies are w/o kidney tumors, but have lower exposures (low enough to avoid cytotoxicity?)</td>
<td></td>
</tr>
</tbody>
</table>

- note key assumptions (general and agent-specific)
- parallel evaluation of alternative MoA (e.g., mutagenic thiol)
- need an “account” of the whole array of observations
# Mode of Action – Assessing Relevance to Humans

## TABLE 1
Concordance of Key Events in Rats and Humans: Atrazine

<table>
<thead>
<tr>
<th>Key event</th>
<th>Evidence in animals (rats)</th>
<th>Evidence in humans</th>
</tr>
</thead>
<tbody>
<tr>
<td>decreased GnRH pulses</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>suppression of LH surge</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>change in cyclicity</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>prolonged increase in estrogen/prolactin</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Mammary Tumors</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Table from Cohen SM et al. 2004. Toxicol Sci 78:181
Does THIS ENDPOINT constitute a Human Hazard?

Reasoned Ranking:

| “Compelling” |  |
| “Likely”    |  |
| “Plausible” (w further assumptions) | Liver, Kidney |
| “Not Impossible” | MCL, Esophageal |
OVERALL WoE is Secondary to the Set of Endpoint-Specific WoE

Does THIS ENDPOINT constitute a Human Hazard?

Reasoned Ranking:

- "Compelling"
- "Likely"
- "Plausible" (w further assumptions) Liver, Kidney
- "Not Impossible" MCL, Esophageal

- Human Carcinogen
- Likely
- Possible
- Cannot Be Determined
- Unlikely

This (along with the reasoning to develop it) becomes part of the “narrative”
Advantages of Hypothesis-Based WoE

- Shows which endpoints are most compelling
- Makes reasoning explicit – shows both strengths and weaknesses
- Allows debate about particular data and interpretations
- Frames WoE classifications as scientific statements (falsifiable, points to tests)
- Does not blend all endpoints together in an uninterpretable generalized “carcinogenicity” statement
- Informs the QRA process
QRA

Does THIS ENDPOINT constitute a Human Hazard?

Reasoned Ranking:

- “Compelling”
- “Likely”
- “Plausible” (w further assumptions)
- “Not Impossible”

Liver, Kidney

MCL, Esophageal

Based on animal liver

Based on human liver meta-analysis

Based on human high-dose kidney

Based on animal kidney

Based on rat MCL

Based on human esophageal meta-analysis
QRA Uncertainty Characterization

- Endpoint WoE categories give relative weighting
- Hypothesis informs quantitative uncertainty analysis (e.g., variation in sensitivity of rats and mice may indicate uncertainty about human sensitivity)
- Still have statistical uncertainties for each dataset analysis (model fit, extrapolation, x-spp), but the qualitative questions have been removed to the WoE category question
- Representativeness of alternative potencies tied to Hazard ID understanding
Advantages of Hypothesis-Based WoE

• Gives some measure of how compelling the evidence is for each endpoint
• Gives structure and substance to the overall WoE
• Gives context for model and dataset choices; ties QRA to WoE
• Not just one endpoint to represent all possible human risks
• Lays out reasoning; transparent
Is This Just...

- What we do already, but reorganized?
- The MoA human-relevance framework?
- The Bradford Hill Criteria?
- Evidence-Based Toxicology? (Guzelian)
Help from Epistemology in Gauging WoE and Evaluating Hypotheses Against Data

Epistemology, or theory of knowledge, is the branch of philosophy that studies the nature and scope of knowledge and belief.

- What is the nature of knowledge?
- By what processes of inference can we acquire knowledge?
- How are we justified in claiming knowledge?
Considerations in Evaluating a Theory

- falsifiability
- parsimony
- predictivity (and “retrodiction”)
- risky predictions
- penalties for *ad hoc* addenda to reconcile underdetermined theories with discordant observations
- penalties for accommodation
Thank you for the opportunity to present these ideas.