Systematic Reviews, Risk of Bias and Evidence Evaluation for Characterization of Environmental Chemicals

Tracey J. Woodruff, PhD, MPH
Professor and Director
UCSF Program on Reproductive Health and the Environment

Risk Assessment Specialty Section (RASS)

January 14, 2015
"Eureka! More information!"
Environmental health literature is vast, diverse, and of variable quality

How do we evaluate scientific evidence to make decisions?

And shorten the time between science and decision?
Clinical sciences have faced and addressed these same challenges.
What is Evidence Based Medicine?

• Defined as, the “conscientious, explicit, and judicious use of the current best evidence in making decisions about the care of individual patients”

• In general, systems of EBM combine:
  
  (1) existing scientific evidence;
  
  (2) clinical expertise; and
  
  (3) patient values and preferences

to make diagnosis, prognosis and treatment decisions.
But Evidence Based Medicine Methodologies Are Not Directly Transferable to Environmental Science!
Navigation Guide Methodology

A systematic and transparent method to evaluate the quality of evidence and to support evidence-based decision making, bridging the gap between clinical and environmental health

Developed by UCSF’s Program on Reproductive Health and the Environment in collaboration with the Navigation Guide Working Group in 2009
Overview of the Methodology

1. Specify Study Question
   Is human environmental exposure to a chemical a reproductive health risk?

2. Select Evidence

3. Rate Quality & Strength of the Evidence:
   Strength of Evidence in Non-Human Systems
   - Known to be Toxic to Human Reproduction
     - Sufficient
     - Limited
     - Inadequate
   - Possibly Toxic
     - Sufficient
     - Limited
     - Inadequate
   - Not Classifiable
     - Sufficient
     - Limited
     - Inadequate
   (Strength of human & non-human evidence are combined into 1 of 5 possible strength of evidence summary statements)

Feedback

4. Grade Strength of Recommendation:
   Strength of Evidence (from Step 3 above)
   - High
     - Known to be Toxic
     - Probably Toxic
     - Possibly Toxic
     - Not Classifiable
   - Medium
     - Known to be Toxic
     - Probably Toxic
     - Possibly Toxic
     - Not Classifiable
   - Lower
     - Known to be Toxic
     - Probably Toxic
     - Possibly Toxic
     - Not Classifiable
   Is a Less Toxic Alternative Available?
   Patient Values and Preferences
   Strong or Discretionary Recommendation

S = Strong Recommendation
  - denotes “we recommend”
D = Discretionary Recommendation
  - denotes “we suggest”

1. High Exposure =
   - Exposure at any level that occurs during critical or sensitive windows of development or during other periods of heightened vulnerability (i.e., nutritional deficiencies, chronic disease/immunosuppressed state, etc.);
   - Exposure at high level for any duration;
   - Exposure of moderate or low level for long (chronic) duration
2. Medium Exposure =
   - Exposure at moderate level for short or intermittent duration
3. Lower Exposure =
   - Exposure at low level for short or intermittent duration
“...systematic-review standards provide an approach that would substantially strengthen the IRIS process...” NAS 2014

“EPA should consistently use a more systematic approach to evaluating the literature .........” NAS 2014
Weight of Evidence

“too vague and is of little scientific use”

National Academy of Sciences 2014
The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Integration of Animal and Human Evidence for PFOA Effects on Fetal Growth

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Nonhuman Evidence for PFOA Effects on Fetal Growth

The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes
Tracey J. Woodruff and Patrice Sutton

Systematic Review and Evidence Integration for Literature-Based Environmental Health Science Assessments
Andrew A. Rooney, Abee L. Boyles, Mary S. Wolfe, John R. Bucher, and Kristina A. Thayer
Evaluate each evidence stream separately using systematic and transparent approaches
Systematic Review Approach for Each Evidence Stream

A pre-specific analytic plan (protocol) is developed and applied consistently to the evidence.

- **“PECO” Statement**
- **Systematic search**
- **Select Studies**
- **Extract Data & Data Analysis**
- **Rate Quality of Evidence**
- **Rate the Strength of Evidence**

**Overall Conclusion**

- **Human Data**
  - “PECO” Statement
  - Systematic search
  - Select Studies
  - Extract Data & Data Analysis
  - Rate Quality of Evidence
  - Rate the Strength of Evidence

- **Non Human Data**
  - “PECO” Statement
  - Systematic search
  - Select Studies
  - Extract Data & Data Analysis
  - Rate Quality of Evidence
  - Rate the Strength of Evidence
A pre-specific analytic plan (protocol) is developed and applied consistently to the evidence.

- Population
- Exposure
- Comparator
- Outcome
Rate the Quality and Strength of the Evidence

- “PECO” Statement
- Systematic search
- Select Studies
- Extract Data & Data Analysis
- Rate Quality of Evidence
- Rate the Strength of Evidence

Rate Quality of Evidence:
- High
- Moderate
- Low

Rate Strength of Evidence:
- Sufficient evidence of toxicity
- Limited evidence of toxicity
- Inadequate evidence of toxicity
- Evidence of lack of toxicity
• Risk of Bias

• Rating quality of evidence
Figure 1. Evaluating Study Quality and Strength of Evidence

**Risk of Bias**
Risk of bias is determined for each individual study.

**Quality of Evidence**
Quality is rated across all studies separately for human and non-human evidence streams. Human evidence begins as 'moderate quality' and may be downgraded (-1 or -2) or upgraded (+1 or +2) according to factors. Non-human evidence begins as 'high quality' and may be downgraded (-1 or -2) according to factors.

**Strength of Evidence**
Strength is rated across all studies separately for human and non-human evidence streams. The final ratings represent the level of certainty of toxicity.

**Considerations**
- Quality of body of evidence
- Direction of effect
- Confidence in effect
- Other compelling attributes of the data that may influence certainty

**Downgrade Factors (Human and Non-human)**
- Risk of bias across studies
- Indirectness
- Inconsistency
- Imprecision
- Publication bias

**Upgrade Factors (Human only)**
- Large magnitude of effect
- Dose response
- Confounding minimizes effect

**Rating (based on all quality factors)**
- High quality
- Moderate quality
- Low quality

**Considerations (based on all strength considerations)**
- Sufficient evidence
- Limited evidence
- Inadequate evidence
- Evidence of lack of toxicity

Lam et al. EHP 2014
Risk of Bias vs Random Error

1. Bias

2. Random Error

True Effect

0
“the committee emphasizes the need for EPA to assess the “risk of bias” in individual studies.” [for all study types]

“The committee notes that assessing the quality of the study is not equivalent to assessing the risk of bias in the study. An assessment of study quality evaluates the extent to which the researchers conducted their research to the highest possible standards and how a study is reported. Risk of bias is related to the internal validity of a study and reflects study-design characteristics that can introduce a systematic error (or deviation from the true effect) that might affect the magnitude and even the direction of the apparent effect. An assessment of risk of bias is a key element in systematic-review standards; potential biases must be assessed to determine how confidently conclusions can be drawn from the data.”

The committee emphasizes the importance of assessing risk of bias for all study types.

National Academy of Sciences 2014
Evaluate Risk of Bias

- Recruitment strategy
- Blinding
- Confounding
- Exposure assessment
- Incomplete outcome data
- Selective reporting
- Conflict of interest
- Other bias

Sequence generation
- Allocation concealment
- Blinding
- Incomplete outcome data
- Selective reporting
- Conflict of interest
- Other bias
Factors for downgrading/upgrading evidence were derived directly from factors used in GRADE and Cochrane.
Factors that **DECREASE** quality level

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

*For Human and Animal evidence*

1. Risk of bias (study limitations)
2. Indirectness
3. Inconsistency
4. Imprecision
5. Publication bias
Factors that **INCREASE** quality level

Possible ratings: 0=no change; -1 or -2 downgrade 1 or 2 levels

*For Human evidence only*

1. Large magnitude of effect
2. Dose response
3. Confounding minimizes effect
<table>
<thead>
<tr>
<th></th>
<th>Hill</th>
<th>GRADE</th>
<th>Navigation Guide</th>
<th>NTP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Downgrading confidence or weakening recommendation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of bias</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Indirectness(^d)</td>
<td>X</td>
<td>X(^b)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Imprecision</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Publication bias</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Financially conflicted sources of funding</td>
<td>X(^c)</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Upgrading confidence or strengthening recommendation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large effect</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Dose-response relationship(^d)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>No plausible confounding</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Cross-species, population, or study consistency</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Serious or rare end points, such as teratogenicity</td>
<td>X</td>
<td></td>
<td>X(^b)</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) Indirectness is the extent to which a study directly addresses the study question (Higgins and Green 2011).
\(^b\) Indirectness might arise from the lack of a direct comparison or if some restriction of the study limits generalizability.
\(^c\) Includes Hill criteria of specificity, biologic plausibility, and coherence.
\(^d\) Rated under “other.”
\(^d\) A formal dose-response assessment is typically performed, depending on the outcome of the hazard identification. However, at this stage, a potential dose-response relationship provides evidence of a hazard and should be used in a hazard-identification process.
## Summary of Quality of Evidence for PFOA

<table>
<thead>
<tr>
<th>Evidence Stream</th>
<th>Human</th>
<th>Non-human mammalian</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Starting rating</strong></td>
<td>Moderate</td>
<td>High</td>
</tr>
<tr>
<td><strong>Downgrade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk of Bias</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>Indirectness</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Inconsistency</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Imprecision</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Publication bias</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Upgrade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large magnitude effect</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>Dose response</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>All possible confounding would confirm negative result</td>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>Grade</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td><strong>Final rating</strong></td>
<td>Moderate</td>
<td>Moderate</td>
</tr>
</tbody>
</table>
Rate the Strength of Evidence

CRITERIA:
1. Quality of evidence:
2. What is the direction of effect?
3. What is the confidence in the effect?
4. Are there other compelling attributes of the data that influence certainty?

QUALITY OF EVIDENCE
- High
- Moderate
- Low

STRENGTH OF EVIDENCE (LEVEL OF CERTAINTY REGARDING TOXICITY)
- Sufficient evidence of toxicity
- Limited evidence of toxicity
- Inadequate evidence of toxicity
- Evidence of lack of toxicity
Methodological Needs

• Criteria for moving from quality to strength of evidence
• Methods to include all potential types of evidence, i.e., assessing chickens, flies and *in vitro* data
• Improved methods of animal toxicity testing – high ROB may be prevalent for key domains
• Mechanistic data is considered under other considerations…. Further development needed
• Consider the nature and extent of consensus that is needed for a decision
Conclusion

• Doable
• Rigorous, systematic, transparent and doable
• Capacity to evolve with changes in evidence stream
Acknowledgements

PFOA Case Study Authors

Dylan Atchley
UCSF
Research Assistant

Paula Johnson
UCSF
Post-doctoral Fellow

Saunak Sen
UCSF
Associate Professor

Patrice Sutton
UCSF
Research Scientist

Tracey Woodruff
UCSF
Professor, Director, PRHE

Daniel Axelrad
EPA
Environmental Scientist

Erica Koustas
EPA
ORISE Post-doctoral Fellow

Juleen Lam
EPA
ORISE Post-doctoral Fellow

Karen Robinson
Johns Hopkins
Associate Professor
Acknowledgements

Navigation Guide Funders (since 2009)

- USEPA
- US Environmental Protection Agency STAR (RD83467801)
- National Institute for Environmental Health Sciences (ES018135)
- Cal-EPA
- Fred Gellert Foundation
- Clarence Heller Foundation
- New York Community Trust
- Forsythia Foundation
- Passport Foundation
- Johnson Family Foundation
- Heinz Endowments
- Rose Foundation
- Kaiser Permanente
- Planned Parenthood Federation of America
- UCSF Phillip R Lee Institute for Health Policy Studies
Thank you

Program on Reproductive Health and the Environment

UCSF
University of California
San Francisco
Sir Austin Bradford Hill - incompleteness of science...

“does not confer upon us a freedom to ignore the knowledge we already have, or to postpone the action that it appears to demand at a given time” (Hill 1965).