PUBLIC STAKEHOLDER WORKSHOP
TO INFORM EPA'S UPCOMING IRIS

TOXICOLOGICAL REVIEW OF INORGANIC ARSENIC

SESSION 1:

APPLYING SYSTEMATIC REVIEW TO THE iAs ASSESSMENT

Tuesday, January 8 &
Wednesday, January 9
RTP, North Carolina

HOSTED BY EPA'S NATIONAL CENTER
FOR ENVIRONMENTAL ASSESSMENT

ELEMENTS OF SYSTEMATIC REVIEW

NTP
National Toxicology Program

Elements of Systematic Review

Andrew Rooney, Ph.D.
Deputy Director Office of Health Assessment and Translation
National Institute of Environmental Health Sciences

Public Stakeholder Workshop to Inform EPA's upcoming IRIS Toxicological Review of Inorganic Arsenic
January 8, 2013

Systematic Review

- A scientific investigation that focuses on a specific question, and uses explicit, pre-specified methods to identify, select, summarize, and assess the findings of similar studies
- Provides greater transparency
- Used to:
  - reach evidence-based conclusions
  - develop clinical or public health recommendations
  - clarify need for additional research
  - may or may not result in quantitative meta-analysis
- Existing methodologies are generally used for assessment of healthcare interventions

Preparing the Topic

- **Scope and focus the topic** to answer specific questions
- **Develop protocol** to detail project-specific procedures used throughout the evaluation
  - Literature search strategy
  - Procedures for selection of relevant studies
  - Outcomes considered
  - Data extraction methods
  - Approach for assessment of study quality (risk of bias)
  - Methods for evaluation of confidence
- **Protocol** contains enough details so that the process and procedures could be reconstructed
- **Opportunities to obtain input** from experts and public

Prepare Topic → Search for and Select Studies for Inclusion → Extract Data from Studies → Assess Quality of Individual Studies → Rate Confidence in Body of Evidence

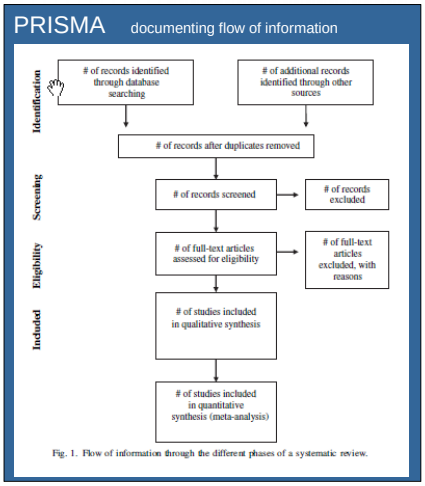
Searching for and Selecting Studies for Inclusion

- **Literature search**
  - Perform comprehensive search
  - Documented strategy in enough detail so that it could be replicated
- **Screen studies for inclusion/exclusion**
  - Select relevant studies based on pre-defined criteria
  - Generally recommended methods
    - Evaluate each study by 2 reviewers independently
    - Plan how conflicts between reviewers will be resolved
    - Document reasons for exclusion

Prepare Topic → Search for and Select Studies for Inclusion → Extract Data from Studies → Analyze and Synthesize Studies → Report Systematic Review

From Searching to Screening (continued)

- Literature identification
  - Database searching
  - Other sources
    - Bibliographies of good studies
    - Experts, public, etc.
- Screening
  - Title/abstract relevance screen
    - Exclusion or retrieval of full text
  - Full-text eligibility screen
  - Ability to document reasons for exclusion
    - "review, no new data"
    - "no data on outcome of interest"



*Moher D *et al.* 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology 62(10): 1006-1012.*

Prepare Topic → **Search for and Select Studies for Inclusion** → Extract Data from Studies → Analyze and Synthesize Studies → Report Systematic Review

Extract Data from Studies

- **Extract data**
  - Individual study information collected systematically
  - Generally recommended methods
    - Standardized data collection forms
    - Quality assurance of data
    - Training/testing of the approach

Prepare Topic → Search for and Select Studies for Inclusion → **Extract Data from Studies** → Analyze and Synthesize Studies → Report Systematic Review

Analyze and Synthesize Studies

- **Assess study quality (risk of bias) of individual studies**
  - Are you confident in the study findings?
  - There are a number of reporting quality tools
  - There are some established risk of bias tools
  - Decide how risk of bias assessments will be used
    - Will studies be excluded?
    - Is a narrative discussion of risk of bias planned?
  - Generally recommended methods
    - Single summary scores for "study quality" are discouraged
    - Reporting quality checklists are not risk of bias tools
- **Rating confidence in a body of evidence**
  - Most existing methods (e.g., GRADE and AHRQ) are primarily used to assess health care interventions
- **Present findings**

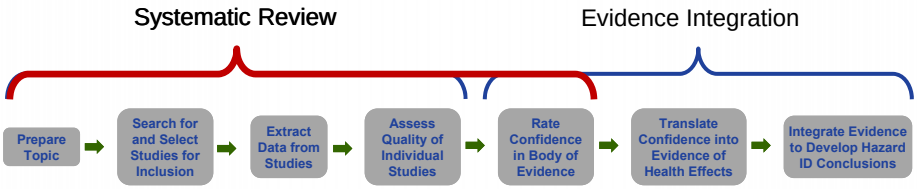
Prepare Topic → Search for and Select Studies for Inclusion → Extract Data from Studies → **Analyze and Synthesize Studies** → Report Systematic Review

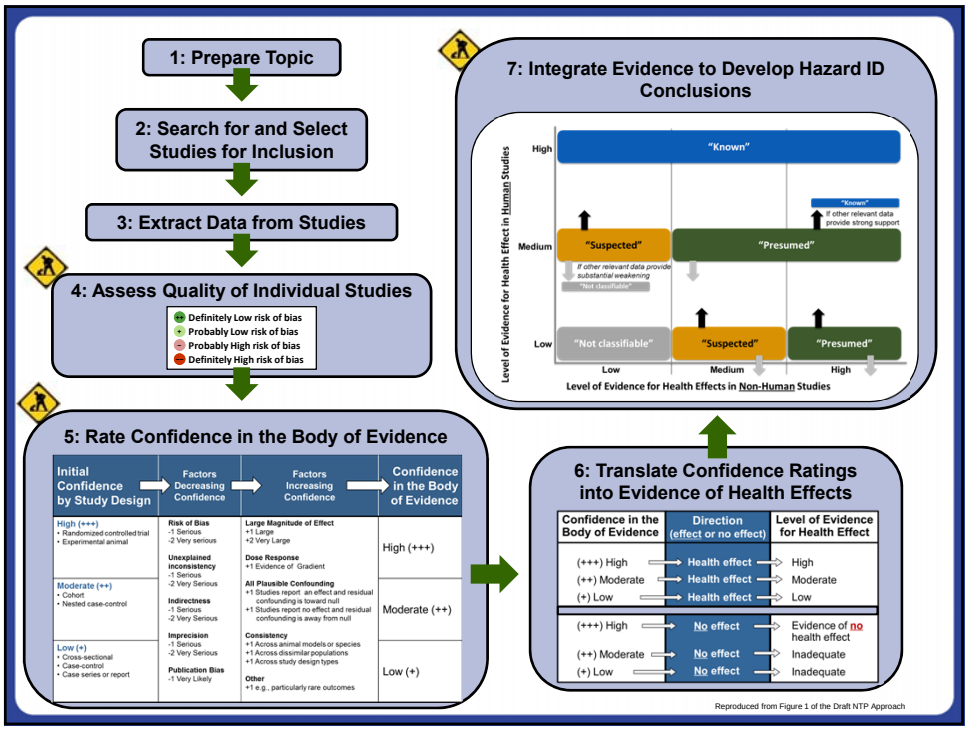
What Does A Systematic Review Not Do?

- Does not eliminate the need for expert judgment
- Does not guarantee reproducibility of conclusions
  - Increased transparency does not necessarily eliminate differences in scientific judgment
- Most methods do not provide guidance on how to
  - Integrate evidence across human, animal, and mechanistic studies
  - Reach hazard identification conclusions
  - Select key studies for dose-response analysis/set reference values

Draft NTP Approach for Systematic Review and Evidence Integration for Literature-Based Health Assessments

- **Evidence integration** is the process for reaching conclusions on the NTP's confidence across a body of studies within an evidence stream (i.e., human and animal data separately) and then integrating those conclusions across the evidence streams with consideration of other relevant data such as supporting evidence from mechanistic studies
- Why not "Weight of Evidence"?
  - Lack of consensus on meaning (Weed et al., 2005)





OPTIONS FOR LITERATURE SEARCH STRATEGIES

EPA United States Environmental Protection Agency

Inorganic Arsenic Literature Search and Evaluation



Janice S. Lee
USEPA / ORD / NCEA
Research Triangle Park, NC

Office of Research and Development
National Center for Environmental Assessment

EPA United States Environmental Protection Agency

Systematic Review

Literature search

↓

Screening for relevance

↓

Evaluation of study strengths and limitations

↓

Evaluating data

↓

Synthesizing data

13

EPA United States Environmental Protection Agency

Literature Search Strategies for Arsenic

- Goal is to identify **relevant** literature
- Possible approaches:

1. Cast a wide net
   - Manual evaluati
   - Clustering
2. Citation mapping

Acknowledgements: Ryan Jones and Ray Antonelli

14

EPA United States Environmental Protection Agency

Core Databases for Primary Literature

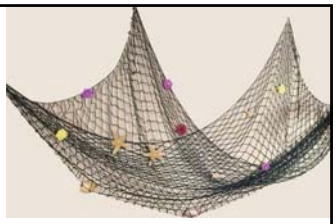
Database	Description
PubMed *	Approximately 5,600 medical, biology, and other life sciences journals (through MEDLINE), most back to 1966. www.pubmed.com
Web of Science *	12,000 science and social science journals, back to 1970. Also includes conference abstracts. Maintained by Thompson Reuters. http://apps.webofknowledge.com
TOXLINE *	Toxicology journals, including developmental and reproductive toxicology (DART), technical reports and research projects, and archival collections; back to 1965 (a few citations dating back to the 1940's); run by NLM. http://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?TOXLINE

* Accessible through Health and Environmental Research Online (HERO) database

15

EPA United States Environmental Protection Agency

Literature Search Approach: Cast a Wide Net



Arsenic 2012

Initial Lit Search - 43795 references
- PubMed - 21489 references
- WOS - 18208 references
- ToxNet - 26998 references

Exclusion - 16779 references
- Non-English - 3753 references
- Reviews - 1357 references
- Non Peer-Reviewed - 13831 references

Considered - 27016 references

hero.epa.gov/arsenic

16

EPA United States Environmental Protection Agency

Manual Evaluation

- Manual review of all references
- Is it feasible to manually evaluate the considered references?
- Is it a valuable use of resources?

17

EPA United States Environmental Protection Agency

Clustering

References are placed in groups based on their textual similarity

Relevant groups identified by presence of references from prior assessments

Preliminary results:

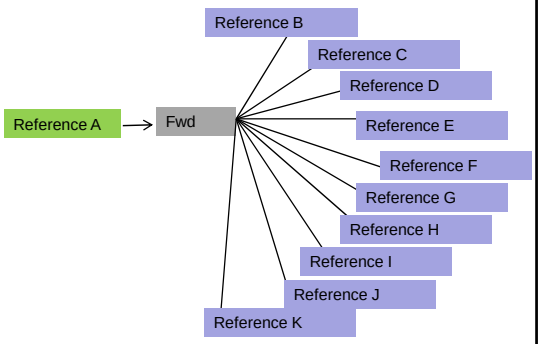
Seed	**References Identified**
NCEA Past Efforts (800)	13461
Other Assessments* (2200)	20944

*WHO (2011), IARC (2004, 2012), NRC (1999, 2001), Health Canada (2006), ATSDR (2007), OPP (2006), CalEPA (2004)

18

EPA United States Environmental Protection Agency

Literature Search Approach: Citation mapping



- Identifies publications that have cited a reference from previous health assessments
- References that cite many seed items have a higher incidence of relevance
- Preliminary results - 108,059 references (unrestricted search)

19

EPA United States Environmental Protection Agency

Literature Search Approach: comparisons

Approaches	Pros	Cons
Manual Evaluation	• thorough • human expertise applied to every item individually • possible to record reason for each exclusion	• labor and resource intensive • each reference must be double checked to account for cognitive exhaustion • time consuming
Clustering	• computer does the work • objective • able to analyze large data sets	• will not group relevant items from alternate fields with different vocabularies
Citation Mapping	• draws attention to items overlooked by traditional database searches, as it does not rely on metadata but on connections formed by expert evaluation of relevance	• limited to items in databases that index citations, like Web of Science • does not fully overlap PubMed

20

EPA United States Environmental Protection Agency

Goals

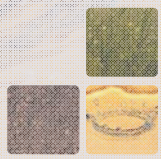
- How to conduct systematic review of a large database?
- How to identify relevant literature?
- How to evaluate studies?
- How to handle new studies from literature search updates?

21

Methods for Identifying, Evaluating, and Synthesizing Literature

1.1 What approaches could EPA use to identify relevant literature for the development of a Toxicological Review of iAs? What approaches could EPA use to transparently communicate results of its literature search and screening strategy?

Lead Discussants: Beth Owens, Andy Rooney



Methods for Identifying, Evaluating, and Synthesizing Literature

Andy Rooney
Beth Owens

1.1 Literature Search and Screening

- What approaches could EPA use to identify relevant literature for the development of a Toxicological Review of iAs?
- What approaches could EPA use to transparently communicate results of its literature search and screening strategy?

1.1 Literature Search and Screening

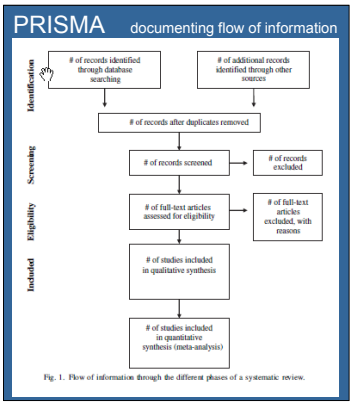
What approaches could EPA use to identify relevant literature for the development of a Toxicological Review of iAs?

- Comprehensive search
  - How many databases should you search? (MEDLINE, TOXNET, etc.)
- Collaborate with trained librarian
  - Training in searching for systematic review?
- Grey literature
  - Produced by industry, government, business and academics not managed by commercial publishers
- Utility of past reviews as source of references
  - Reference lists and HERO
- Role for experts and public
  - FR notice, stakeholder workshop

1.1 Literature Search and Screening

What approaches could EPA use to transparently communicate results of its literature search and screening strategy?

- Communication tools
  - EPA IRIS chemical-specific Website
  - FR notices
- Systematic review reporting standards
  - Sufficient detail to
    - Replicate literature search
    - Recreate screening process and results
  - Examples - Cochrane, AHRQ, etc.
  - PRISMA statement (Moher et al., 2009*)

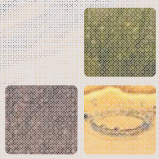


*Moher D *et al.* 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Journal of Clinical Epidemiology 62(10): 1006-1012.*

Methods for Identifying, Evaluating, and Synthesizing Literature

1.2 What approaches are available to evaluate the quality of individual studies? What aspects of epidemiological studies could be considered in such an evaluation?

Lead Discussants: Andy Rooney, Craig Steinmaus



Methods for Identifying, Evaluating, and Synthesizing Literature

Andy Rooney
Craig Steinmaus

1.2 Quality of Individual Studies

- What approaches are available to evaluate the quality of individual studies?
- What aspects of epidemiological studies could be considered in such an evaluation?

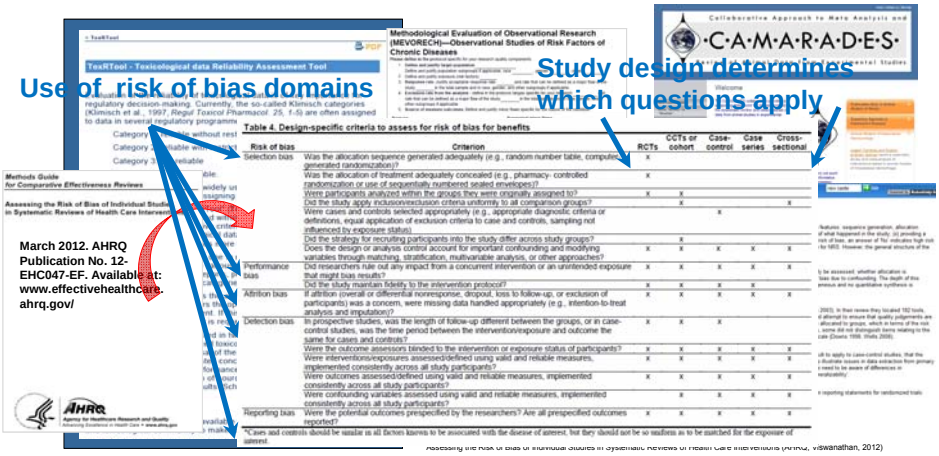
1.2 Quality of Individual Studies

What approaches are available to evaluate the quality of individual studies?

- » Study quality ≈ risk of bias ≈ internal validity
- » Reporting quality checklist ≠ risk of bias tool
- » Judge whether the design and conduct of individual studies compromise credibility of the link between exposure and outcome
- » Risk of bias approaches within systematic review methods
  - − Single summary scores for "study quality" are discouraged
  - − Established tools for randomized controlled trials
  - − No consensus on how to assess risk of bias for observational human studies, animal studies, or *in vitro* studies

Available Study Quality Methods

- Tools for animal studies are generally reporting quality checklists
- There are a number of risk of bias methods for human studies
- The 2012 AHRQ method* addresses a range of human study types
- Current draft NTP Approach adapts the AHRQ questions to address both human and animal studies (http://ntp.niehs.nih.gov/go/38138)



1.2 Quality of Individual Studies

- What aspects of epidemiological studies could be considered in such an evaluation?

SPECIFIC CRITERIA FOR EPIDEMIOLOGIC STUDIES ON ARSENIC
Format based on The Cochrane Collaboration's tool for assessing risk of bias

IA. EXPOSURE MISCLASSIFICATION

GENERAL ISSUES	
Was an appropriate metric of exposure used?	• Drinking water concentrations • Urine • Nails, hair, blood
Was an appropriate latency period considered? Were exposures in all relevant time periods assessed?	Examples: • Cancer: latency may be 40 years or more, exposure from birth • Heart disease: latency may be 10-20 years • Biochemical changes: may be even less
Were all **major exposure** sources included?	• Was drinking water at all residences during all relevant exposure windows assessed? Role of migration (*US studies: 35% move every five years*) • Were other water sources assessed: school, work, filters, bottled water…(*e.g., some studies show skin lesions at very low exposures: were higher exposures missed?)* • Are food or work exposures important? (*These may be minor if drinking water concentrations are high*)
Was exposure assessment independent of disease status?	• Blinded exposure assessment? • Was exposure assessed similarly regardless of disease status?

IB. EXPOSURE MISCLASSIFICATION

SPECIFIC METRICS	
Drinking water concentrations	• Completeness of all sources during all relevant time windows • If past exposures are important: *Are records of historical drinking water concentrations available?* *Current measurements? Are water concentrations stable over time?* • Impact of using ecologic exposure data *Taiwan: use of village medians, lots of variability from well to well* *Chile studies: few water sources, less misclassification*
Urine	• Only inorganic (vs. total) arsenic and metabolites assessed? • Appropriate for suspected latency: urine only reflects past 1-2 weeks If not, are exposures likely stable over time? (*day to day variability may be especially important in low exposure studies*) • Appropriate adjustment for urine dilution Problems with InAs/creatinine: Barr et al., EHP 2005, 113:192-200. *Urine creatinine related to diet, muscle, illness, gender, age…* *Urine creatinine related to some diseases: kidney, diabetes?*
Nails and hair	• Appropriate for the suspected latency: few months? • Inter-individual variability: are they well correlated with actual intake? • Possible impact of external contamination?
Blood	• Short half -ife: may be only good for acute effects

ESTIMATE THE IMPACTS OF EXPOSURE MISCLASSIFICATION ON RELATIVE RISKS: QUANTIFY THE LIKELY DIRECTION AND MAGNITUDE
Modern Epidemiology II (Greenland and Rothman)

II. CONFOUNDING

What are the **major** determinants of the outcome/disease?	• Provide a list of all major causes of each outcome • Are they in the causal chain between arsenic and disease?
Were the potential major confounders directly assessed in the study: If yes…	Were they appropriately measured? Were they appropriately accounted for or examined: • Matching, stratification, adjustment • Shown that the levels of the confounder doesn't differ between exposed and unexposed areas
If no…	• Are there reasons to believe they are related to arsenic exposure? • Are they strongly enough related to the disease to cause major confounding? *(e.g., RRs for diet & bladder cancer are mostly low)* • Are they prevalent enough to cause major confounding *(e.g., rare genetic disorders*)
Statistical adjustments: multivariate analyses	• Were appropriate methods used? • Are both adjusted and unadjusted results given? If a large change is seen following adjustments, are these differences adequately and rationally explained

ESTIMATE IMPACTS ON RELATIVE RISKS: QUANTIFY LIKELY DIRECTION & MAGNITUDE *Axelson 1978. Scand J Work Environ Health 4, 85-89. Quantitatively shows that even some confounders (e.g., smoking and lung cancer) may have only small effects*

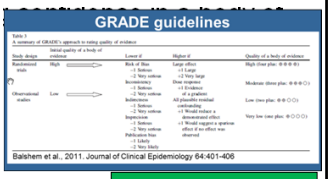
III. SELECTION BIAS AND OTHER ISSUES

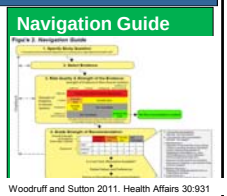
Participation rates	• Did rates include those who declined, could not be found, provided inadequate data, other exclusions…? • Were the participation rates adequate (*e.g., >70%)* • Any major differences based on disease, exposure, and both? • Were participants similar to non-participants?
Case-control studies	• How were cases and controls ascertained? Was this similar in exposed and unexposed areas? • Do controls represent the population from which the cases were selected?
Prospective studies	• Were follow-up rates adequate? • Any major differences in follow-up rates based on disease and exposure status?
Cross-sectional studies	• Is appropriate latency, past exposures considered (see above)
Ecologic studies	• Variability within an exposure area (*e.g., Chile vs. Taiwan*)
Other	• Dose-response: adequate range of exposures? • Magnitude of RRs consistent with other research? • Plausible?

Extra Slides

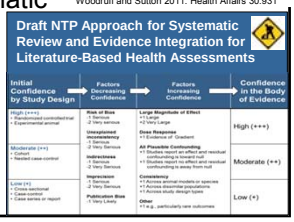
Analyze and Synthesize Studies

- Existing methods provide guidance on rating evidence (e.g., GRADE, AHRQ, Cochrane)
  - Separate "quality of evidence" and confidence in a body of evidence or "strength of evidence"
- Published methods do not provide guidance on how to
  - Integrate evidence across human, animal, and mechanistic studies
  - Reach hazard identification conclusions
- Recent efforts have been made to adapt systematic review approaches to environmental questions
  - Navigation Guide
  - Draft NTP Approach adapts GRADE to consider the range of data relevant for addressing environmental health questions (see http://ntp.niehs.nih.gov/go/38138)





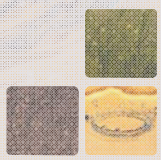
Woodruff and Sutton 2011. Health Affairs 30:931



Methods for Identifying, Evaluating, and Synthesizing Literature

1.3 What approaches are available to synthesize the available evidence on iAs?

Lead Discussants: Warner North, Roberta Scherer



JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Approaches to data synthesis

Roberta W. Scherer, PhD

Toxicological Review of Inorganic Arsenic

January 8, 2013

Data synthesis for systematic reviews

- Goal is to present and synthesize data related to question posed by review
  - Aim for transparency and rigor
  - Provide the evidence to make decisions or recommendations
- First step - write protocol for data synthesis before data collection to reduce bias

Features of systematic review data synthesis

- **Evidence tables** describes included studies:
  - Population
  - Intervention/exposure
  - Outcome measures
  - Quality of evidence
  - Sample sizes
- **Qualitative synthesis**
  - Narrative summary that may include additional data tables, graphs, charts
  - Addresses the strength of the evidence in context

Features, cont'd

- **Quantitative synthesis** – meta-analyses or other statistical testing and examination of heterogeneity
  - Statistically pools results to obtain a single summary result with confidence intervals using weighted values
  - Assesses strength of evidence to determine whether an effect exists in a particular direction
  - Investigates heterogeneity to examine reasons for different results

Sensitivity Analyses

- Assesses effects of including or excluding studies
- Used to re-analyze data
  - Within a range of results
  - Imputing values for missing data
  - Different statistical approaches
- Data from all analyses would be presented to ensure transparency

GRADE approach assess confidence in results

- **GRADE** approach: addresses the ***confidence*** that can be placed in the study findings
- Studies are "graded" up or down, based on:
  - risk of bias (internal validity)
  - external validity
  - heterogeneity across studies
  - sample size
- Results presented in a *Summary of Findings* table
- Does not make decisions or recommendations – only provides the evidence

Comments for the Arsenic/IRIS Workshop Jan 8-9, 2013

D. Warner North

President and Principal Scientist,
NorthWorks, Inc.;

E-mail: northworks@mindspring.com

Web: www.northworks.net

Section 1.3

Section 1.3 – North comments

- Inorganic arsenic is **unusual** as an IRIS entry.
- Not all risk assessments should be done the same way – for IRIS, or in other contexts. (Reference: the NAS "Color" books)
- Focus should be on **health risk at potential low-dose human exposure.**
  - Is the health risk potentially significant?
  - If so, want quantitative estimate(s) and uncertainty disclosure.

Section 1.3 – North comments, 2

- For **important** cases/IRIS entries, may want **more** than a review of published papers.
  - Convene a gathering of the **best experts** for discussion and debate. Publish the proceedings.
  - **Frame** the problem first. What is included?
  - Want a **transparent process**, understandable by stakeholders.
  - **Don't preclude evidence**: Assemble it, then evaluate it.

  Reference: *Public Participation in Environmental Assessment and Decision Making*, National Academy Press, 2008, See esp. Chapter 6, "Integrating Science," particularly page 141 on defaults and guidelines.

Quotes – Red*+Blue Books**

"Risk assessment policy consists of the analytical choices that must be made in the course of a risk assessment. Such choices are based on both scientific and policy considerations." (p. 38)

An inference guideline is an explicit statement of a predetermined choice among the options that arise in inferring human risk from data that are not fully adequate or not drawn directly from human experience. A guideline might, for example, specify the mathematical model to be used to estimate the effects of exposure at low doses from observations based on higher doses. (page 51)

[The term "default" (option or assumption) has replaced "inference guideline." "Default options … are essentially policy judgments of how to accommodate uncertainties" Blue Book** page 5.]

**Risk Assessment in the Federal Government: Managing the Process*, National Academy Press, 1983. http://www.nap.edu/catalog.php?record_id=366

Red* + Blue Books**

From *Findings and Recommendations*, page 266:

"EPA and others often interpret the term *risk assessment* as a specific methodological approach to extrapolating from sets of human and animal carcinogenity data, often obtained in intense exposures, to quantitative estimates of carcinogenic risk associated with the (typically) much lower exposures experienced by the human population.

- EPA should recognize that the conduct of a risk assessment does not require any specific methodological approach and that it is best not seen as a number or even a document, but as a way to organize knowledge regarding potentially hazardous activities or substances to facilitate the systematic analysis of the risks that those activities or substances

****Science and Judgment in Risk Assessment**, National Academy Press, 1994. http://www.nap.edu/catalog.php?record_id=2125

Mark Powell, *Science at EPA: Information in the Regulatory Process*, Washington, DC: Resources for the Future, 1999

"… [Society for Environmental Geochemistry and Health] Arsenic Task Force members Willard Chappell and Warner North met with ORD Assistant Administrator Robert Huggett to urge additional arsenic research. Page 214.

(There follows a lengthy description by other meetings and workshops to develop a research agenda . As far as I know, little of this research was subsequently funded and done. Instead, EPA went to an NAS committee, which produced the 1999 and 2001 NAS reports.) - WN)

"The major frustration of a former drinking water official concerning arsenic was the lack of new research available when the time for decision-making arrived: 'The political appointees should never have been put in that type of position.' Interviewees offered a variety of reasons why substantial new research had not been done over the past ten years." Page 214

(Further discussion on why is on page 215-18. What research has been done since 2001?)

Mark Powell, *Science at EPA: Information in the Regulatory Process*, Washington, DC: Resources for the Future, 1999

" … many observers associate Warner North with the 1989 SAB report. At the time, North was vice chair of the SAB Environmental Health Subcommittee through which the Drinking Water Subcommittee reported. When the subcommittee's EPA staffer Richard Cothern approached North with the panel's report from its 1988 meeting in Cincinnati, the controversial issue of a possible detoxification pathway for arsenic was framed as a false dichotomy between the linear, no-threshold dose-response model and the threshold or "hockey stick" model. By pointing out that there are any number of scientifically plausible non-linear dose-response curves lying between these extremes, North helped negotiate the report through the internal SAB review process. (Note, however, that despite the carefully crafted language of the 1989 SAB report, the false dichotomy between the linear, no-threshold, and threshold models has proved to be a hearty perennial.) "

page 219