Methods for Evidence Synthesis and Integration

Jonathan M. Samet, MD, MS
Dean and Professor
Colorado School of Public Health

National Academies Committee on Guidance on PFAS Testing and Health Outcomes

July 14, 2021

colorado school of public health

Statement of Task

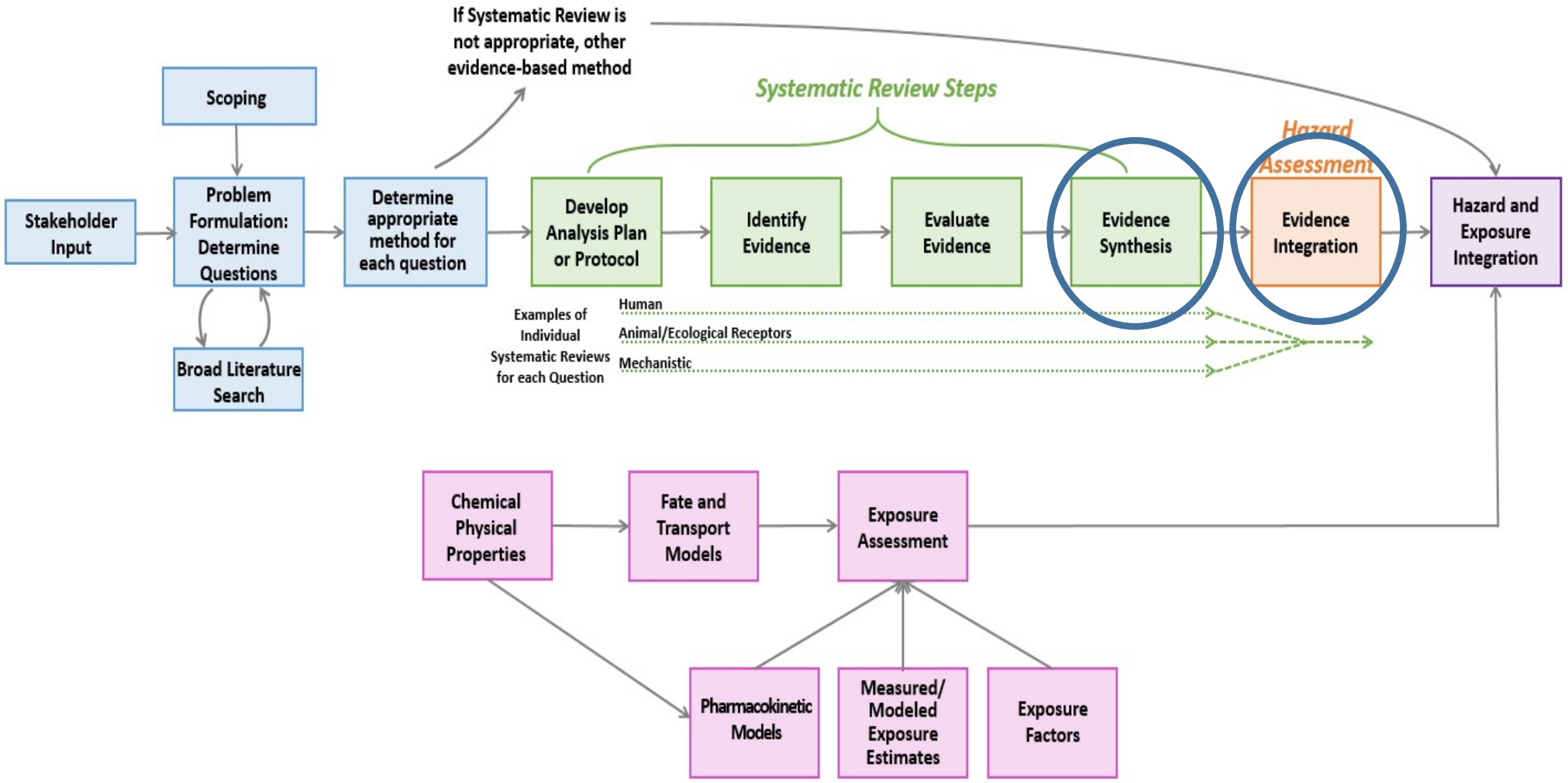
"An ad hoc committee appointed by the National Academies of Sciences, Engineering, and Medicine (the National Academies) **will consider current evidence** regarding human health effects of the most widely studied per- and polyfluoroalkyl substances (PFAS). The National Academies will provide the Centers for Disease Control and Prevention and the Agency for Toxic Substances and Disease Registry (CDC/ATSDR) and the National Institutes of Environmental Health Sciences (NIEHS) **an objective and authoritative review of current evidence** regarding human health effects of those PFAS being monitored in the CDC's National Report on Human Exposure to Environmental Chemicals."

Specific Tasks--Selected

1. “Assess the strength of evidence for the spectrum of putative health effects suggested by human studies…This assessment should characterize the likelihood of those health effects occurring…”
2. “Develop general principles for clinical evaluation or biological testing given substantial scientific uncertainty…”
3. “Review current knowledge about the contribution of PFAS exposure sources (i.e., drinking water, diet, the indoor environment, etc.) to human exposure…”

Sounds like risk assessment to me.

Systematic Review (and not) in Risk Assessment



How systematic review has been applied in other risk assessment contexts

Source: NASEM TSCA Report

Evidence Synthesis (TSCA Report)

Definition: “According to the Cochrane Systematic Review Handbook, evidence synthesis is a process of bringing together data from a set of included studies with the aim of drawing conclusions about a body of evidence (Higgins, Thomas et al. 2019). The process consists of summarizing study characteristics, quality, and effects, and combining results and exploring differences among the studies (e.g., variability of findings and uncertainties), using qualitative and/or quantitative methods.”

Note: Each relevant evidence stream needs to be separately synthesized: mechanistic, bioassay, and human—clinical and epidemiological.

Source: NASEM TSCA Report

Evidence Integration (TSCA Report)

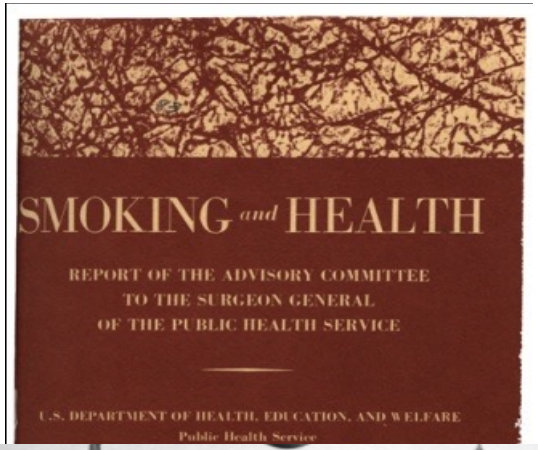
Definition: “Evidence integration is typically considered outside of the systematic review process itself and, in the context of risk evaluations, only employed when an evaluation reviews different evidence streams that have to be reconciled and, as the name suggests, integrated. The outcome of the integration step is an overall conclusion that is based on the holistic consideration of the various evidence streams.”

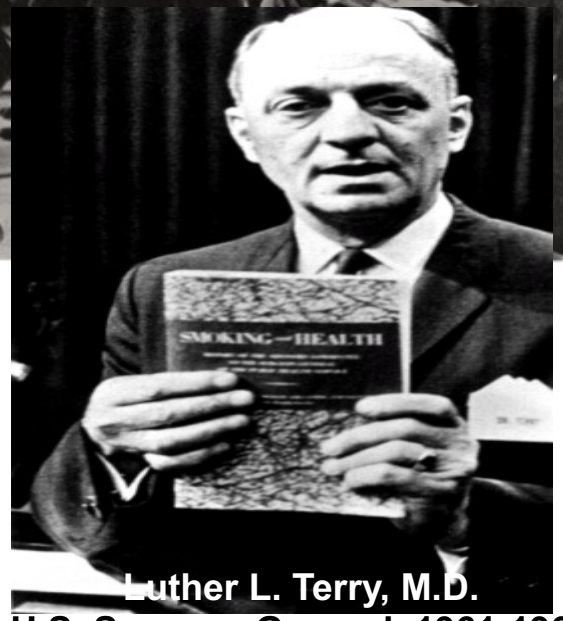
Note: There are multiple schemes for evidence integration, all grounded in expert judgement

Source: NASEM TSCA Report

Relevant History

1964 Surgeon General's Report





Luther L. Terry, M.D.
U.S. Surgeon General, 1961-1965

The 1964 SG report: methods and a protocol for evidence integration

- “A plan was adopted at the first meeting…”
- “…a major general requirement was that of making the information available…”
- “…made decisions or judgments at three levels…”: 1) validity of a publication or report; 2) validity of interpretations and conclusions of authors; and 3) conclusions of the committee.
- Criteria for causal inference

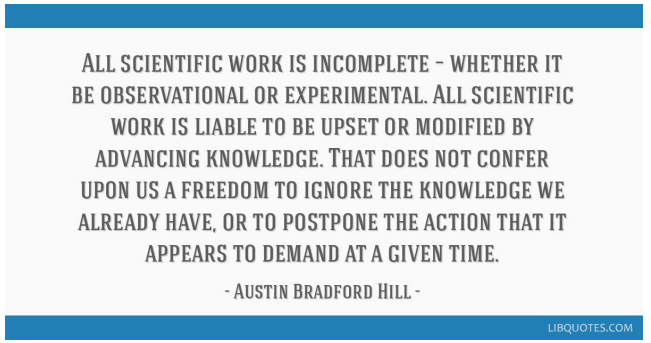
Statistical methods cannot establish proof of a causal relationship in an association. The causal significance of an association is a matter of judgment which goes beyond any statement of statistical probability. To judge or evaluate the causal significance of the association between the attribute or agent and the disease, or effect upon health, a number of criteria must be utilized, no one of which is an all-sufficient basis for judgment. These criteria include:

a) The consistency of the association
b) The strength of the association
c) The specificity of the association
d) The temporal relationship of the association
e) The coherence of the association

Hill's nine "viewpoints"

1. **Strength of the association**
2. **Consistency of the association**
   - "Has it been repeatedly observed by different persons in different places, circumstances, and times?"
3. **Specificity**
   - "If…the association is limited to specific workers and to particular sites and types of disease and there is no association between the work and other modes of dying, then clearly that is a strong argument in favor of causation."
4. **Temporality**
   - … "which is the cart and which is the horse?"
5. **Biological gradient (dose-response)**
   - If the association is one which can reveal a biological gradient, or dose-response curve, then we should look most carefully for such evidence."

Hill's nine "viewpoints"



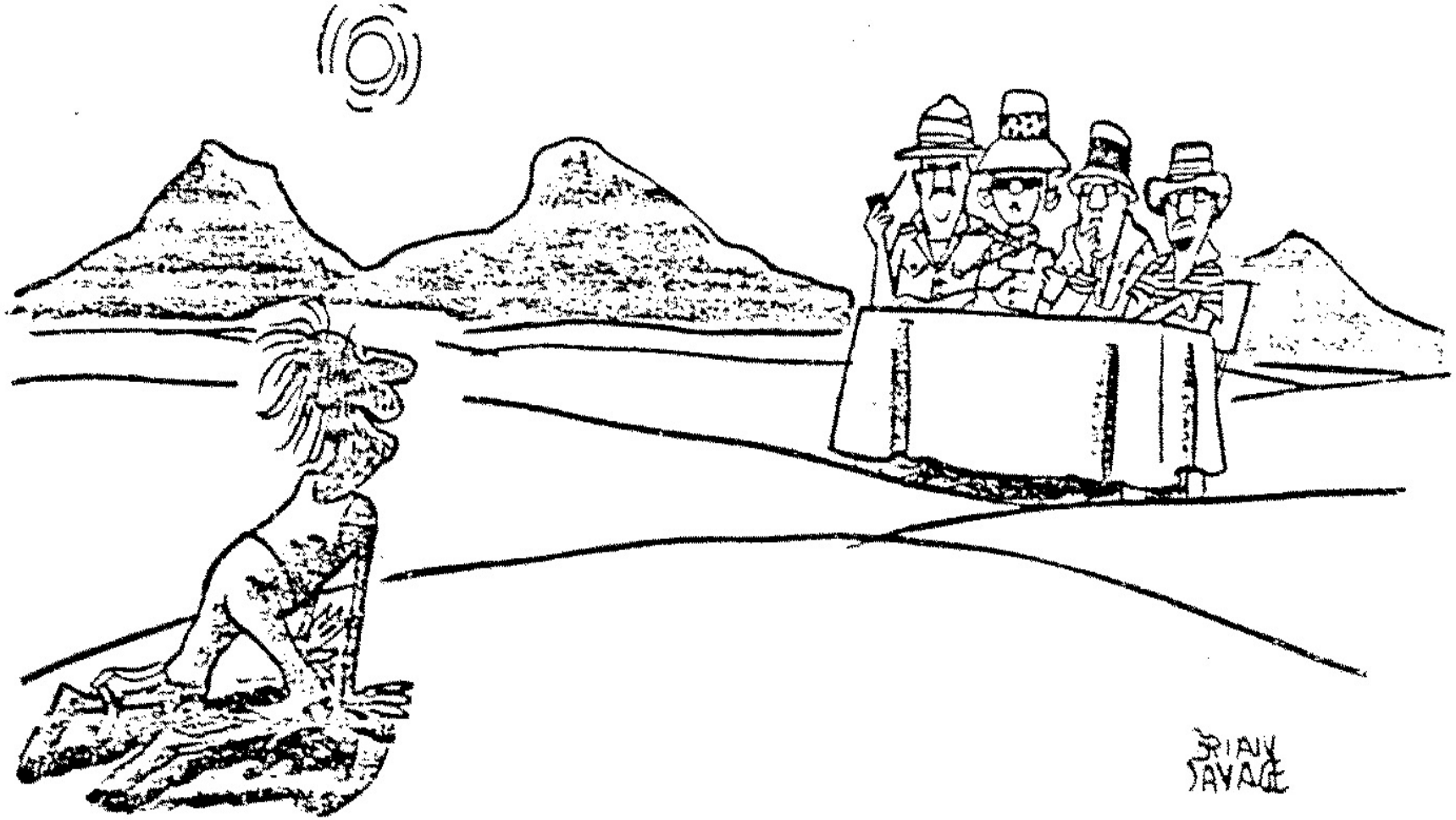
6. **Plausibility**
   - "What is biologically plausible depends on the biological knowledge of the day."
7. **Coherence**
   - "…the cause-and-effect interpretation of our data should not seriously conflict with the generally known facts of the natural history and biology of the disease"
8. **Experiment**
   - "Occasionally is it possible to appeal to experimental or semi-experimental evidence?"
9. **Analogy**
   - "With the effects of thalidomide and rubella before us we would surely be ready to accept slighter but similar evidence with another drug or another viral disease in pregnancy."

Table 1 Guidelines for causal inference. Data from the 1964 *Smoking and Health: Report of the Advisory Committee to the Surgeon General* (71) and from Hill 1965 (33)

US Surgeon General Report's criteria	Hill's criteria
Consistency of association	Strength
Strength of association	Consistency
Specificity of association	Specificity
Temporal relationship of association	Temporality
Coherence of association	Biological gradient
	Coherence
	Experiment

Glass et al. ARPH 2013;34:61-75

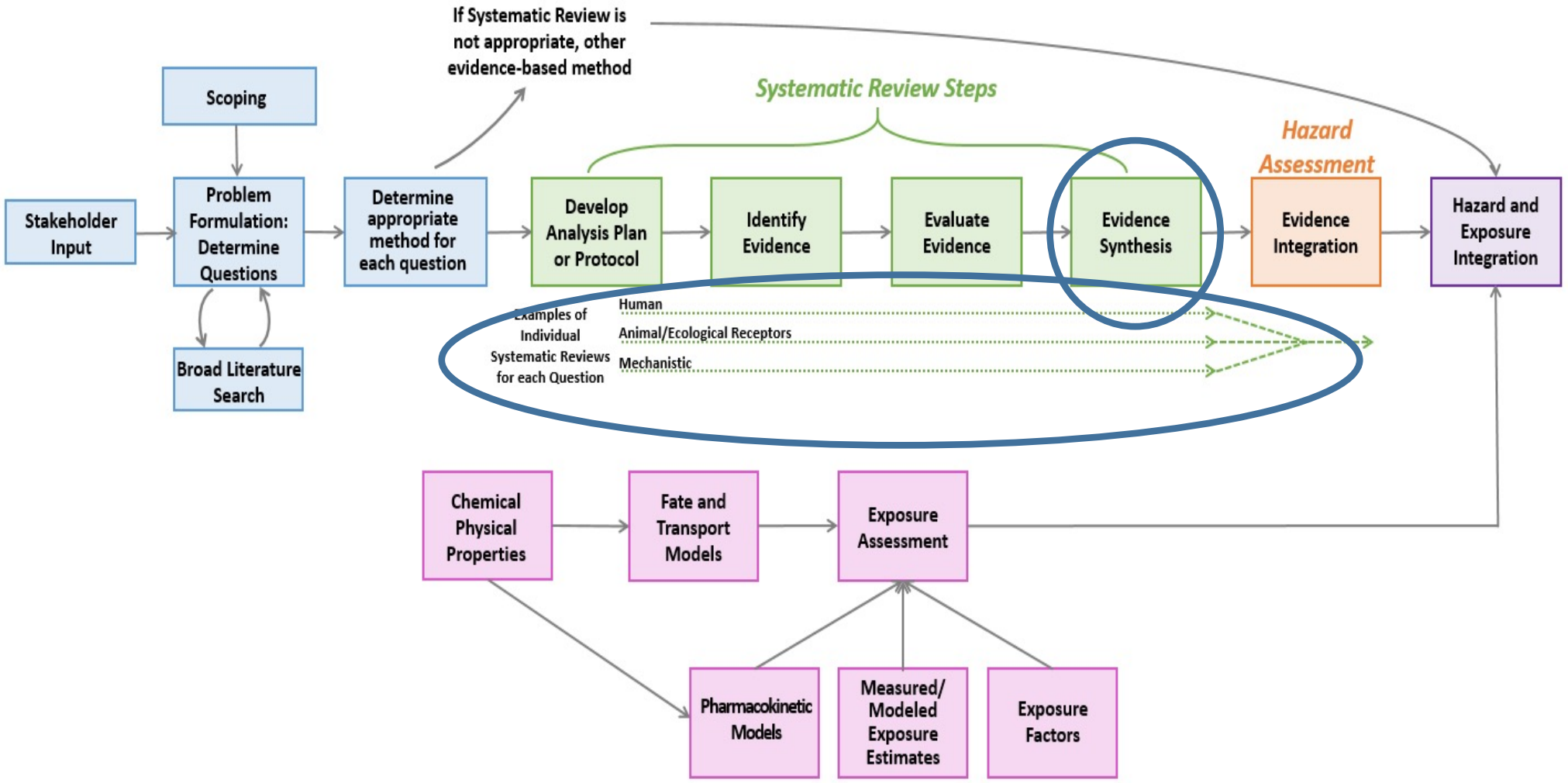
These frameworks for integration are based in expert judgment



"Thank God! A panel of experts!"

Evidence Synthesis

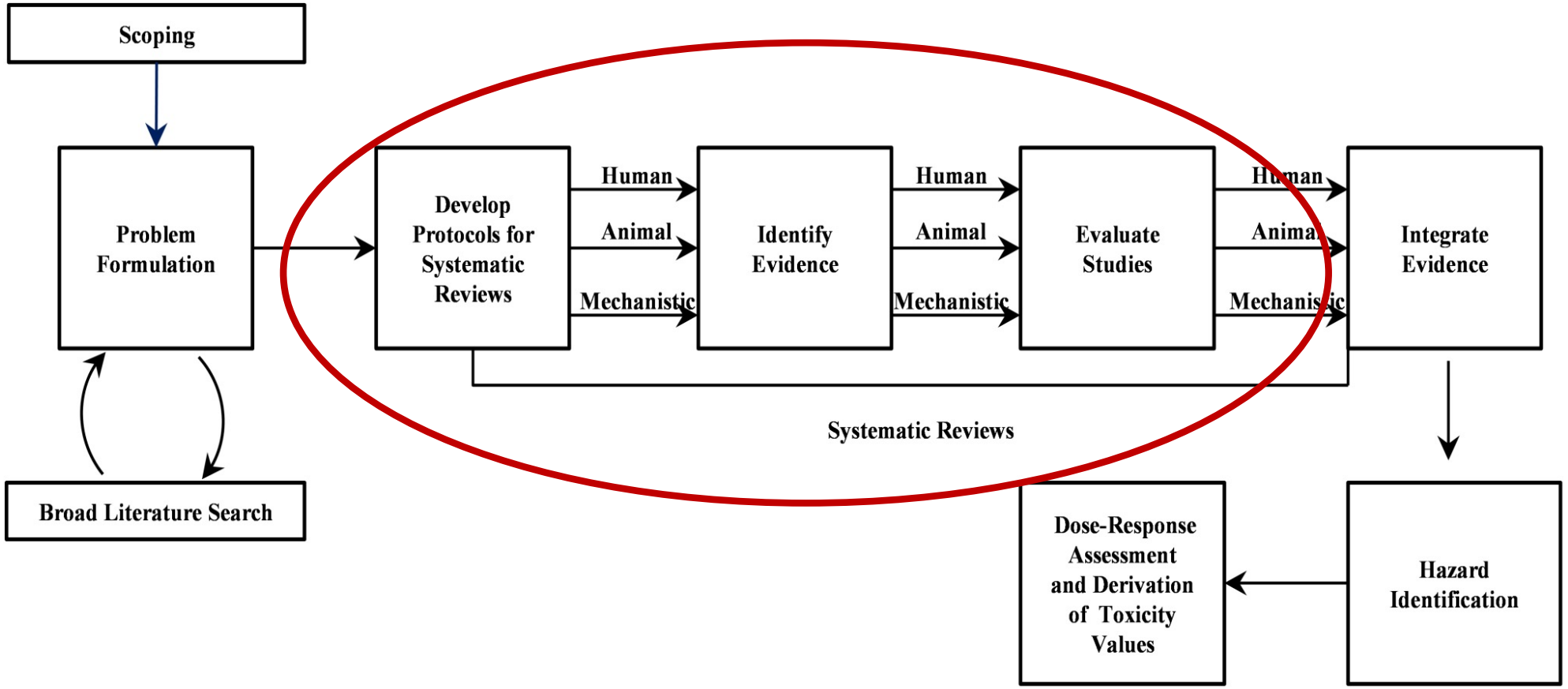
Systematic Review in Risk Assessment



How systematic review has been applied in other risk assessment contexts

Source: NASEM TSCA Report

IRIS process: separate evidence streams



NRC 2014

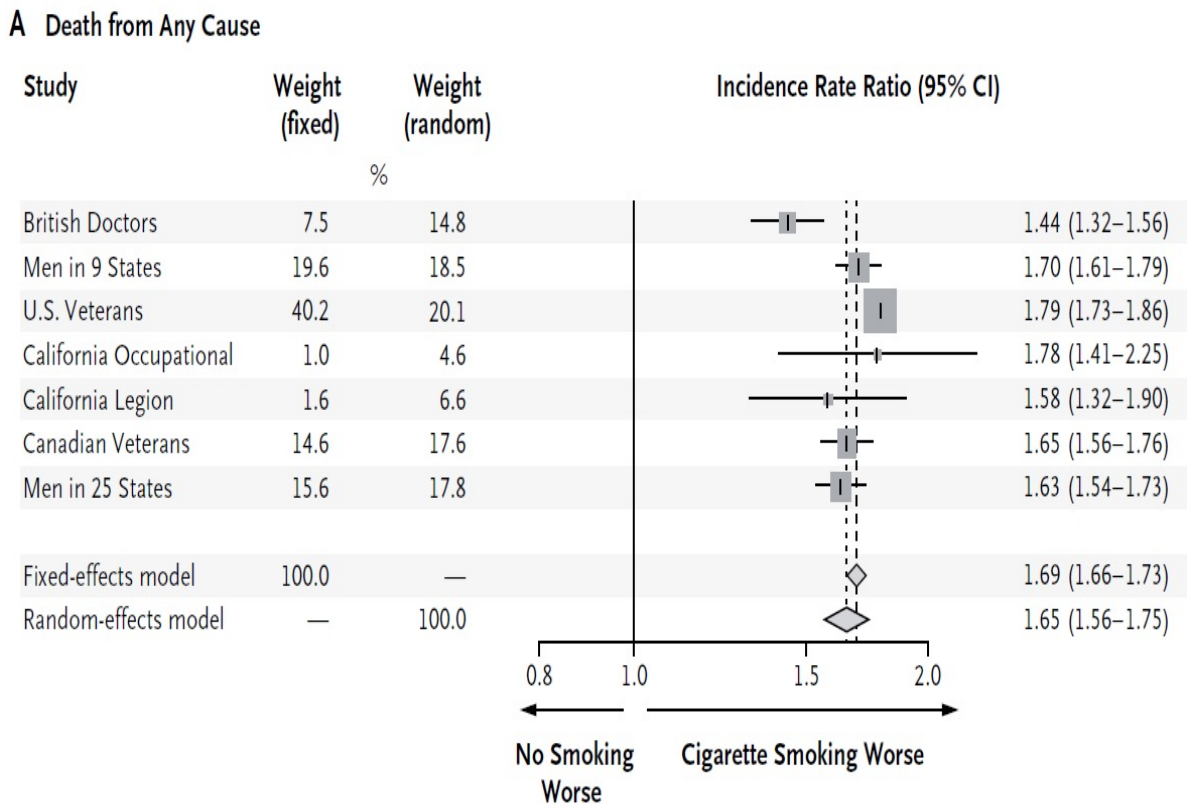
Evidence synthesis for hazard assessment

- Relevant evidence streams for PFAS and human health:
  - Mechanistic toxicology
  - Animal bioassays
  - Human clinical studies
  - Human epidemiological studies
- Emerging challenges
  - Development of methods for “evidence-based toxicology”
  - Using “21st Century Science”

Meta-analysis: a tool for quantitative synthesis

- Quantitative pooling complementary to qualitative synthesis.
- Characterize variation and its causes.
- Used for dose-response assessment

Smoking and mortality: 1964 SG Report

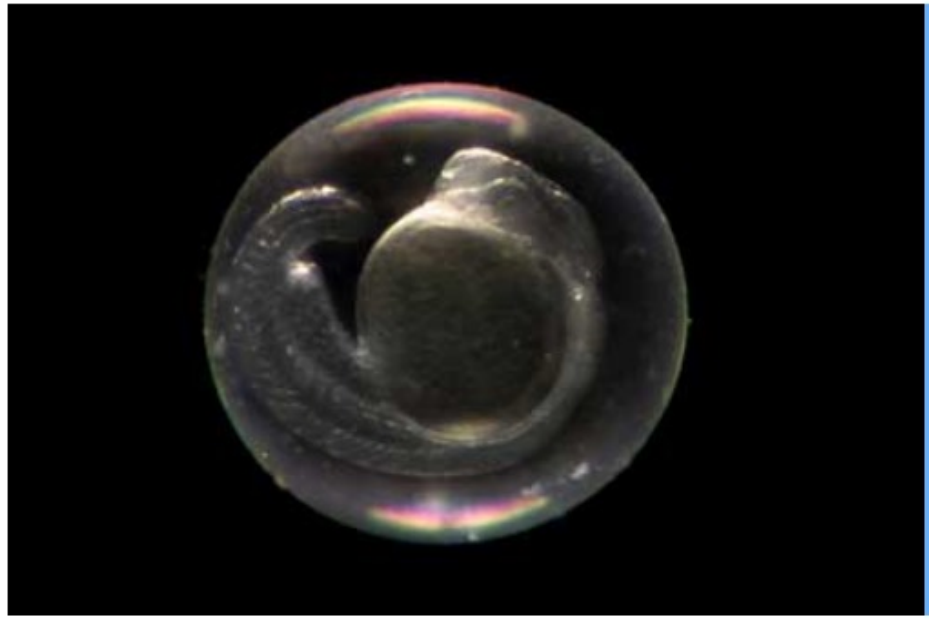


ABOUT US
SELECT PROJECTS
PUBLICATIONS
EVENTS
NEWSLETTER
CONTACT

Sign up for our email newsletter!

ebtc
Evidence-based Toxicology Collaboration

Evidence-Based Toxicology Collaboration



ZEBRAFISH

EBTC is conducting a systematic review to determine how well the Zebrafish Embryotoxicity Test (ZET) can predict developmental malformations in conventional guideline studies using rats and rabbits.

LEARN MORE

New challenges to systematic review: Mechanistic evidence

2019 IARC Preamble

Table 4. Integration of streams of evidence in reaching overall classifications (the evidence in *bold italic* represents the basis of the overall evaluation)

Stream of evidence			Classification based on strength of evidence
Evidence of cancer in humans[a]	**Evidence of cancer in experimental animals**	**Mechanistic evidence**	
Sufficient	Not necessary	Not necessary	**Carcinogenic to humans (Group 1)**
Limited or Inadequate	***Sufficient***	***Strong (b)(1) (exposed humans)***	
Limited	***Sufficient***	Strong (b)(2–3), Limited, or Inadequate	**Probably carcinogenic to humans (Group 2A)**
Inadequate	***Sufficient***	***Strong (b)(2) (human cells or tissues)***	
Limited	Less than Sufficient	***Strong (b)(1–3)***	
Limited or Inadequate	Not necessary	***Strong (a) (mechanistic class)***	
Limited	Less than Sufficient	Limited or Inadequate	**Possibly carcinogenic to humans (Group 2B)**
Inadequate	***Sufficient***	Strong (b)(3), Limited, or Inadequate	
Inadequate	Less than Sufficient	***Strong b(1–3)***	
Limited	***Sufficient***	***Strong (c) (does not operate in humans)***[b]	
Inadequate	***Sufficient***	***Strong (c) (does not operate in humans)***[b]	**Not classifiable as to its carcinogenicity to humans (Group 3)**
All other situations not listed above			

[a] Human cancer(s) with highest evaluation

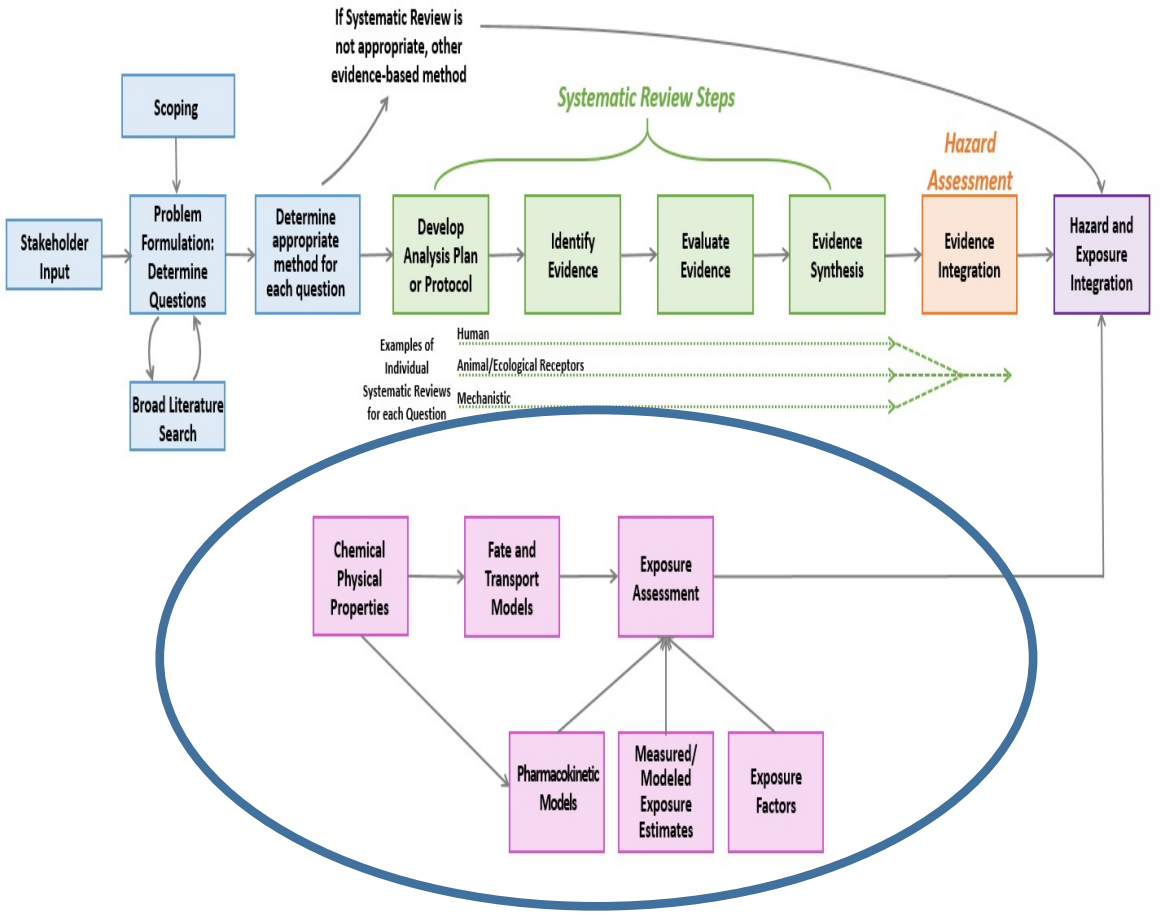
[b] The *strong evidence that the mechanism of carcinogenicity in experimental animals does not operate in humans* must specifically be for the tumour sites supporting the classification of *sufficient evidence in experimental animals*.

Extension to new areas

- Mechanisms
- Exposures
- Models
- Tools lacking for evaluation
- Challenges of synthesis and integration

Synthesis for exposure estimation

- Systematic review has generally not been used for exposure data.
- EPA OPPT has attempted to do so for TSCA with limited success.
- Multiple streams of data are relevant.
- Much not contained in the peer-reviewed literature.



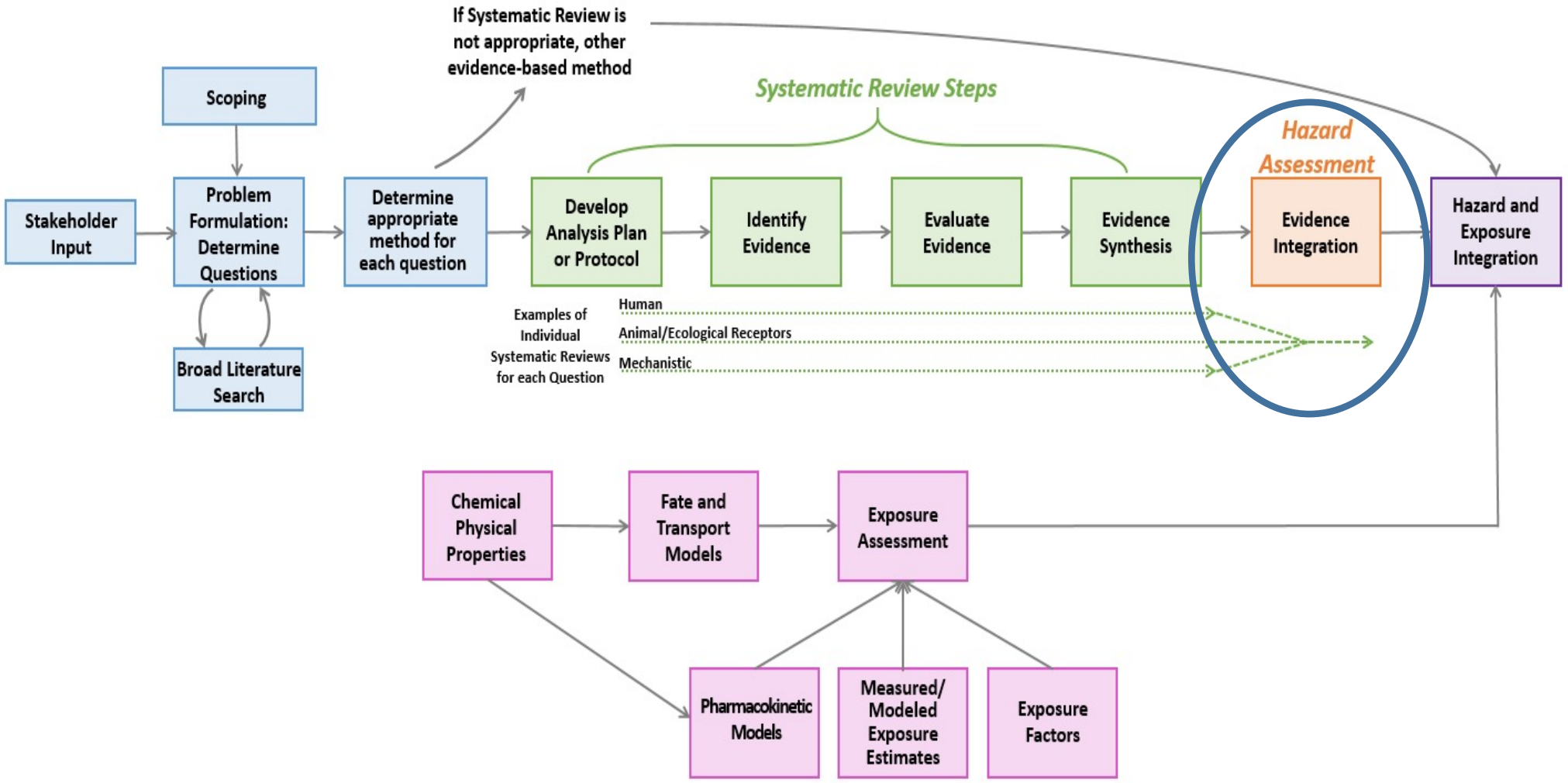
Source: NASEM TSCA Report

Systems for evidence synthesis

- Challenges
  - Dealing with different evidence streams
  - Handling human observational studies that are non-randomized
  - Dealing with the inherent risk of bias in observational studies
- Options
  - GRADE adapted
  - Navigation Guide
  - OHAT
  - IRIS and more

Evidence Integration

Systematic Review in Risk Assessment



How systematic review has been applied in other risk assessment contexts

Source: NASEM TSCA Report

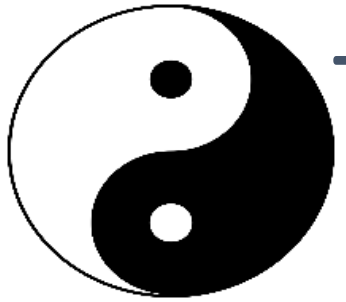
Evidence integration schemes

- Overall purpose is to evaluate the strength of evidence for a conclusion on hazard.
- Bring together the multiple evidence streams in a transparent manner.
- Examples
  - Bradford Hill/Surgeon General's approach
  - International Agency for Research on Cancer (IARC)
  - US EPA: various schemes for "weight-of-evidence" assessment
    - IRIS: Integrated Risk Information System
    - TSCA
    - National Ambient Air Quality Standards (NAAQS)

Potential Advances

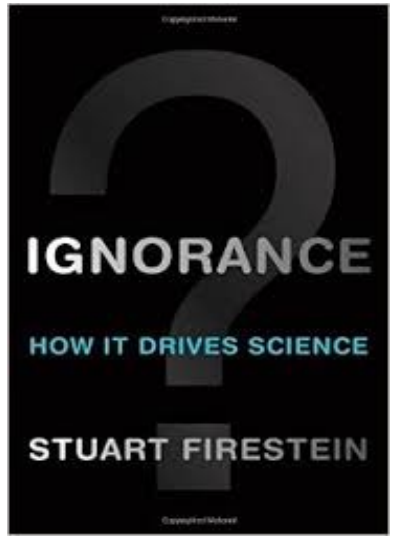
- Bayesian methods for combining information
  - See Biological Effects of Ionizing Radiation (BEIR IV) report 1988: Annex 7A
  - Using 21st Century Science to Improve Risk-Related Evaluations
- Causal modeling to estimate causal effects
- But, reliance on expert judgment remains

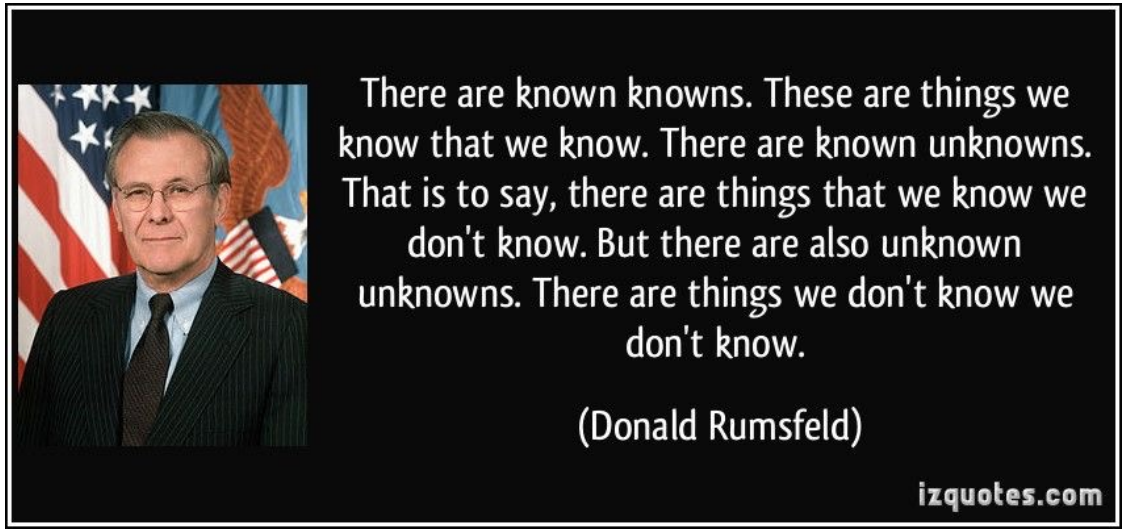
Strength of Evidence



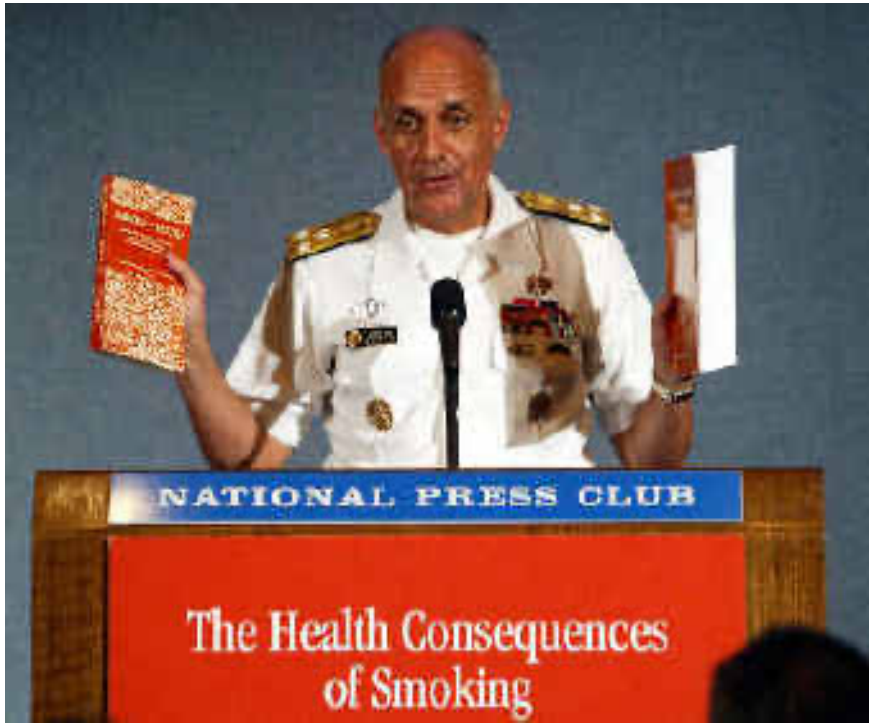
The Yin Yang of evidence

- Evidence: what we know
- Ignorance: what we do not know
- Uncertainty: the consequence of ignorance
- Doubt: lack of belief/confidence in something





The Release of the 2004 U. S. Surgeon General's Report



Richard H. Carmona, MD, MPH, FACS
Surgeon General, 2002-2006

"I hope this Report will inform, galvanize, and inspire our nation, states, and communities to reduce the terrible toll of smoking and to secure a healthy future for America."

Surgeon Generals' Reports: Terminology of Conclusions and the Causal Claims

A four-level hierarchy for classifying the strength of causal inferences based on available evidence:

A. Evidence is **sufficient** to infer a causal relationship.

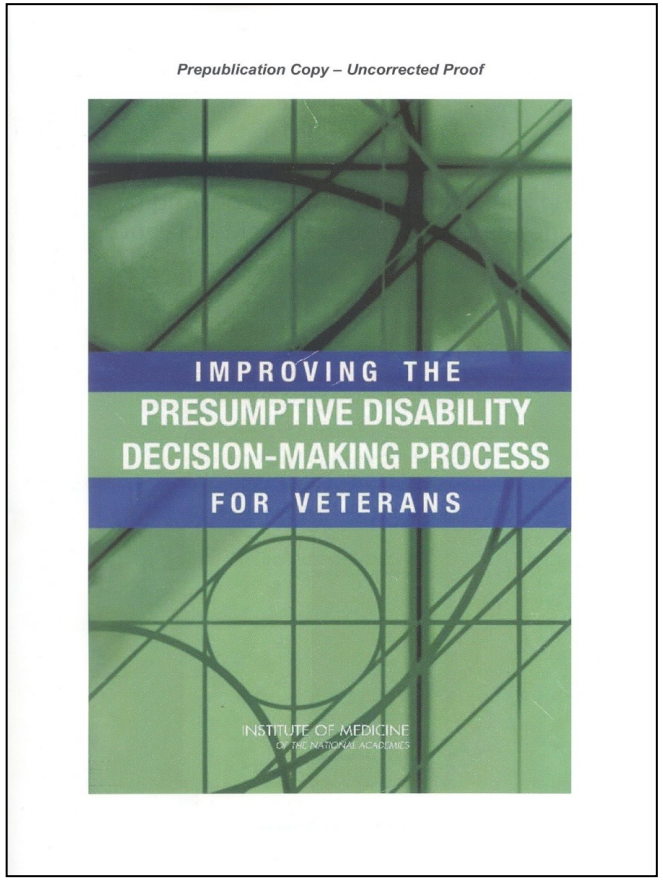
B. Evidence is **suggestive but not sufficient** to infer a causal relationship.

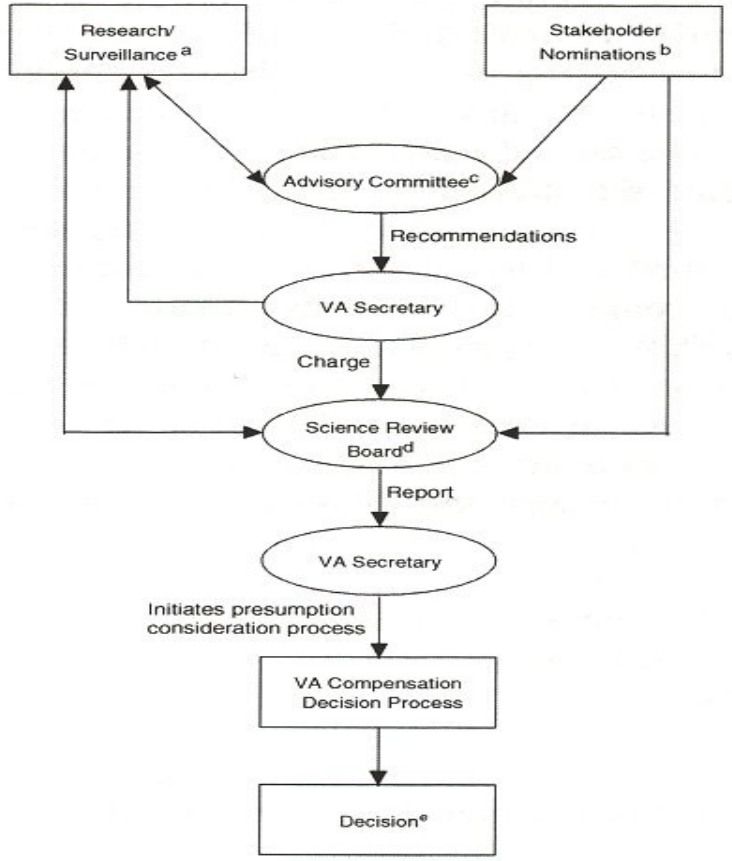
C. Evidence is **inadequate** to infer the presence or absence of a causal relationship (which encompasses evidence that is sparse, of poor quality, or conflicting).

D. Evidence is **suggestive of no causal relationship**.

USDHHS 2004

Setting a decision threshold: Compensating Veterans





IOM. 2007. *Improving the Presumptive Disability Decision-Making Process for Veterans.* Washington, DC: The National Academies Press.

Equipoise and Evidence
What is it?

FOR

AGAINST

The balance point for strength of evidence on causation

Proposed 4-Level Classification Scheme for Causation

Sufficient: the evidence is sufficient to conclude that a causal relationship exists

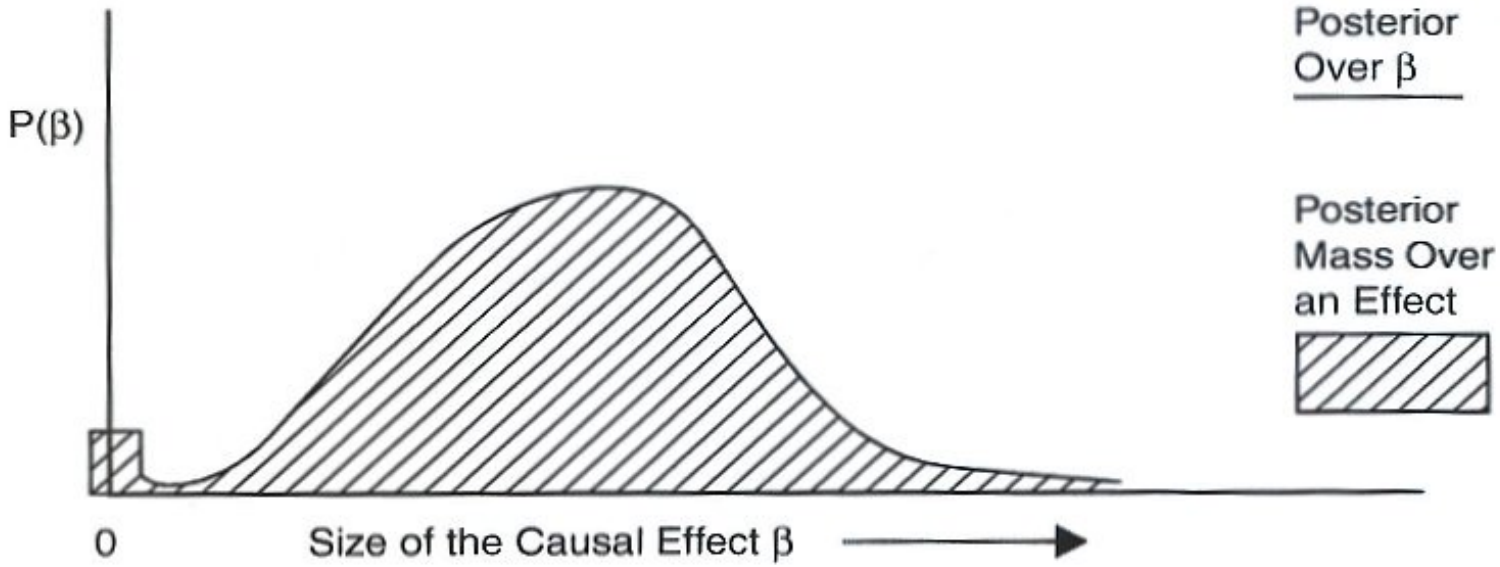
Equipoise and Above: the evidence is sufficient to conclude that a causal relationship is at least as likely as not, but not sufficient to conclude that a causal relationship exists

Below Equipoise: the evidence is not sufficient to conclude that a causal relationship is at least as likely as not, or is not sufficient to make a scientifically informed judgment

Against: the evidence suggests the lack of a causal relationship

Example posterior for *sufficient*

Synthesizing evidence for causation

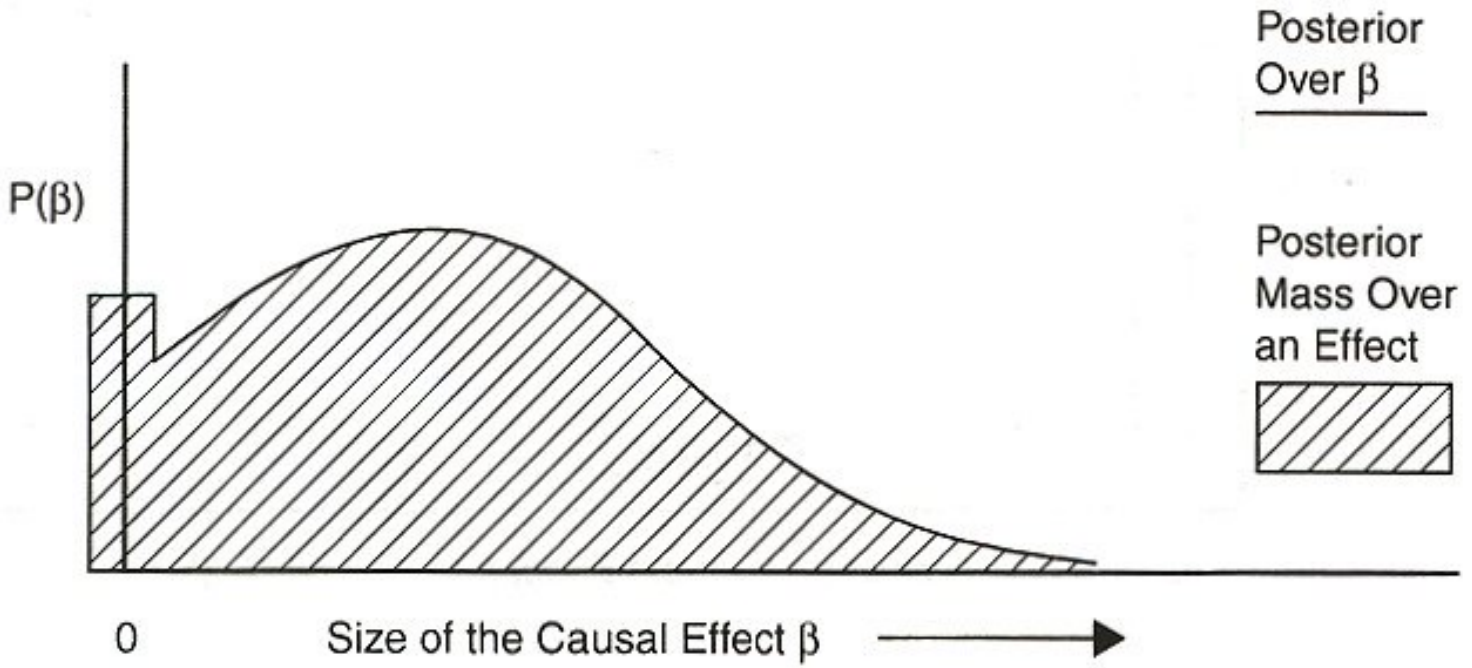


THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

IOM. 2007. *Improving the Presumptive Disability Decision-Making Process for Veterans.* Washington, DC: The National Academies Press.

Example posterior for *equipoise and above*

Synthesizing evidence for causation

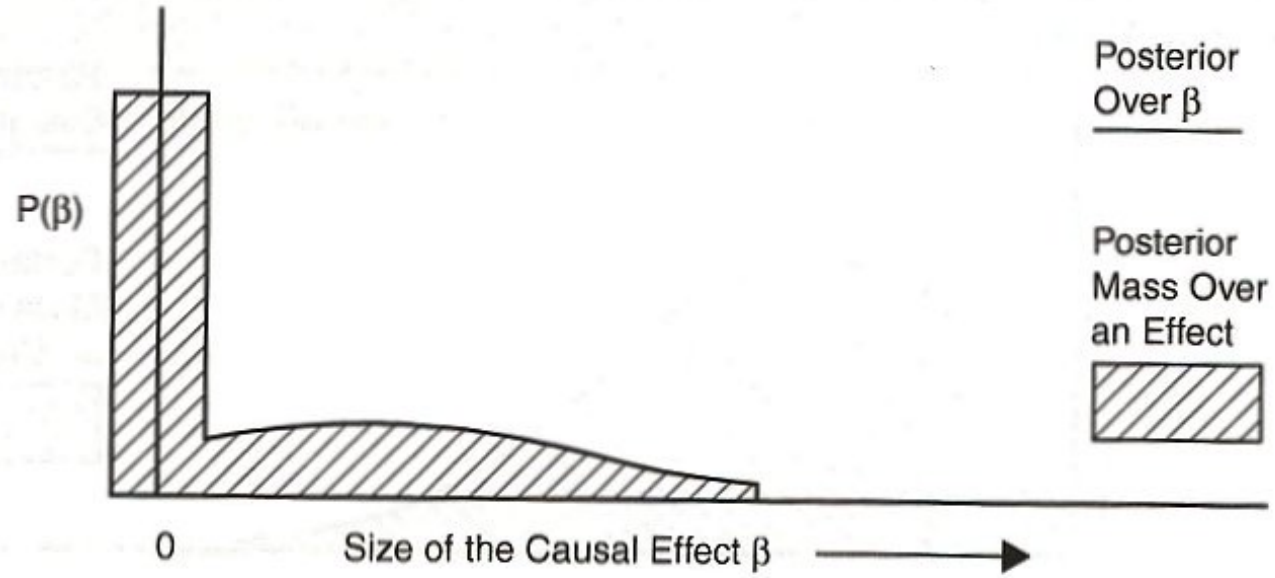


THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

IOM. 2007. *Improving the Presumptive Disability Decision-Making Process for Veterans.* Washington, DC: The National Academies Press.

Example posterior for *below equipoise*

Synthesizing evidence for causation

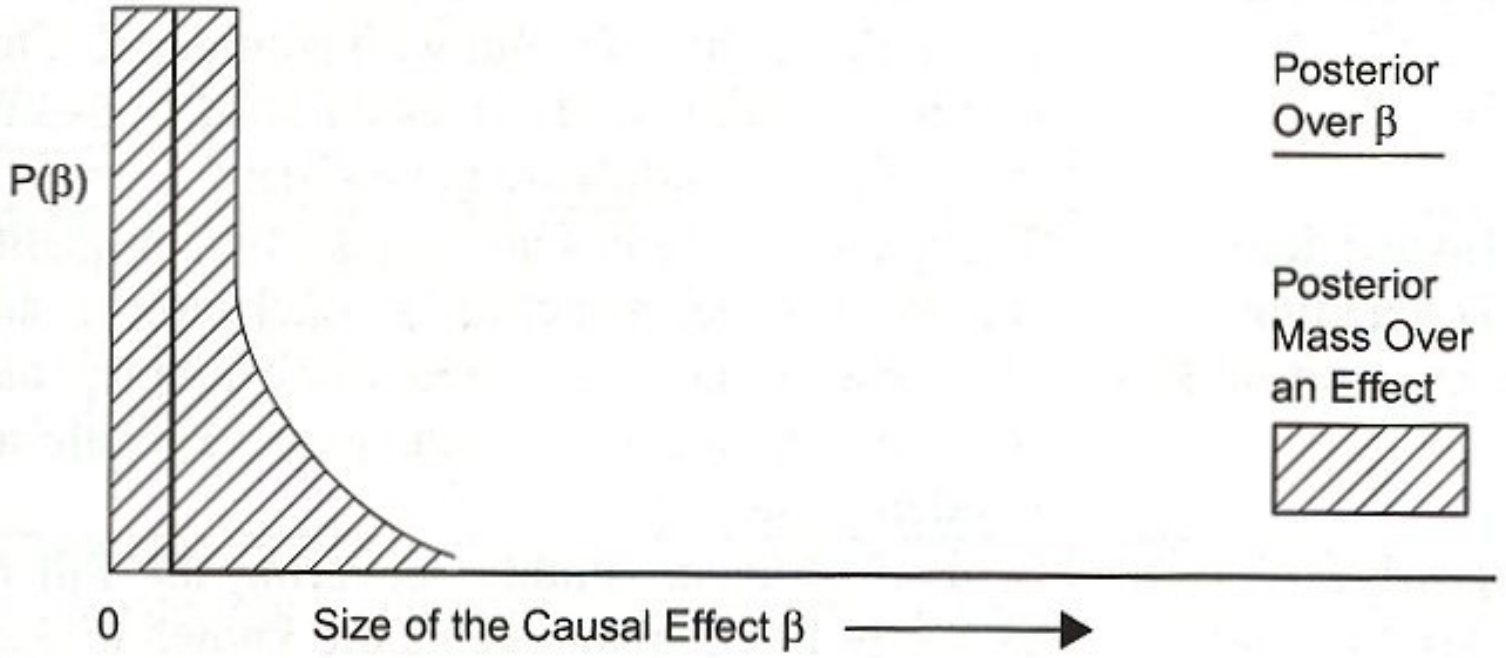


THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

IOM. 2007. *Improving the Presumptive Disability Decision-Making Process for Veterans.* Washington, DC: The National Academies Press.

Example posterior for *against*

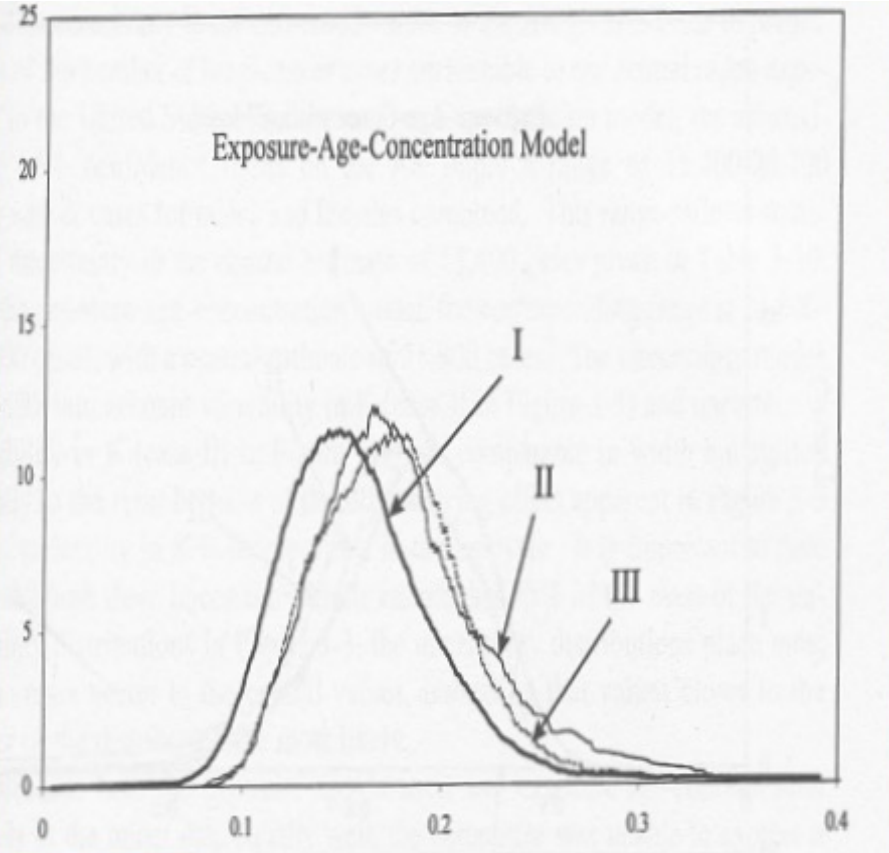
Synthesizing evidence for causation

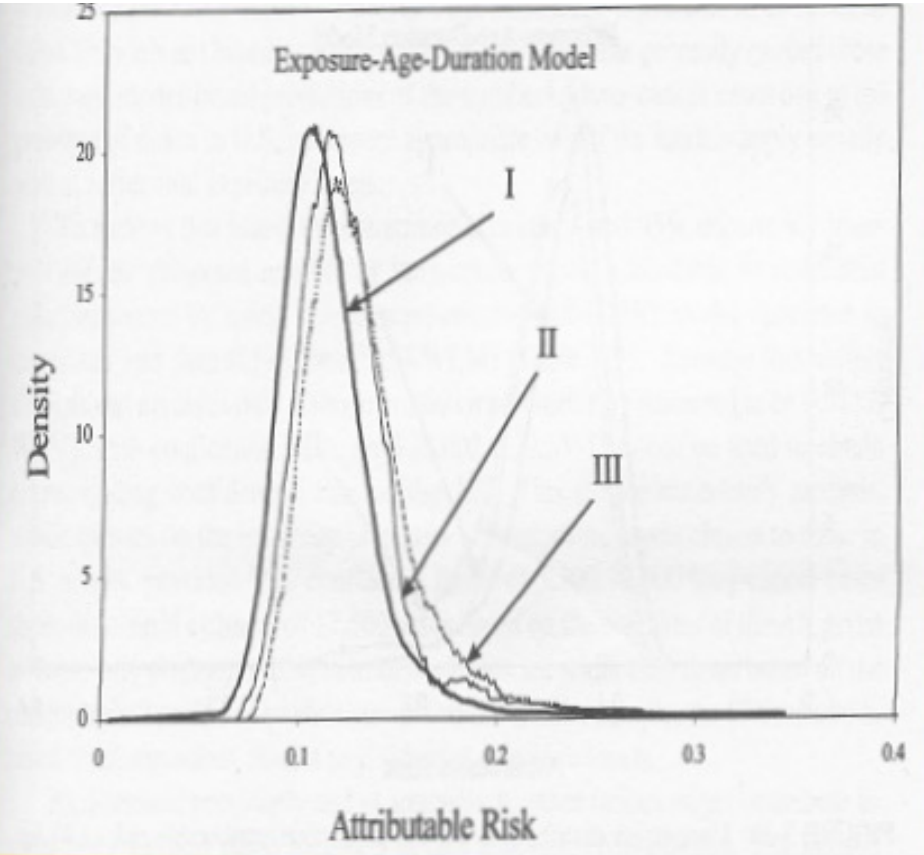


THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

IOM. 2007. *Improving the Presumptive Disability Decision-Making Process for Veterans.* Washington, DC: The National Academies Press.

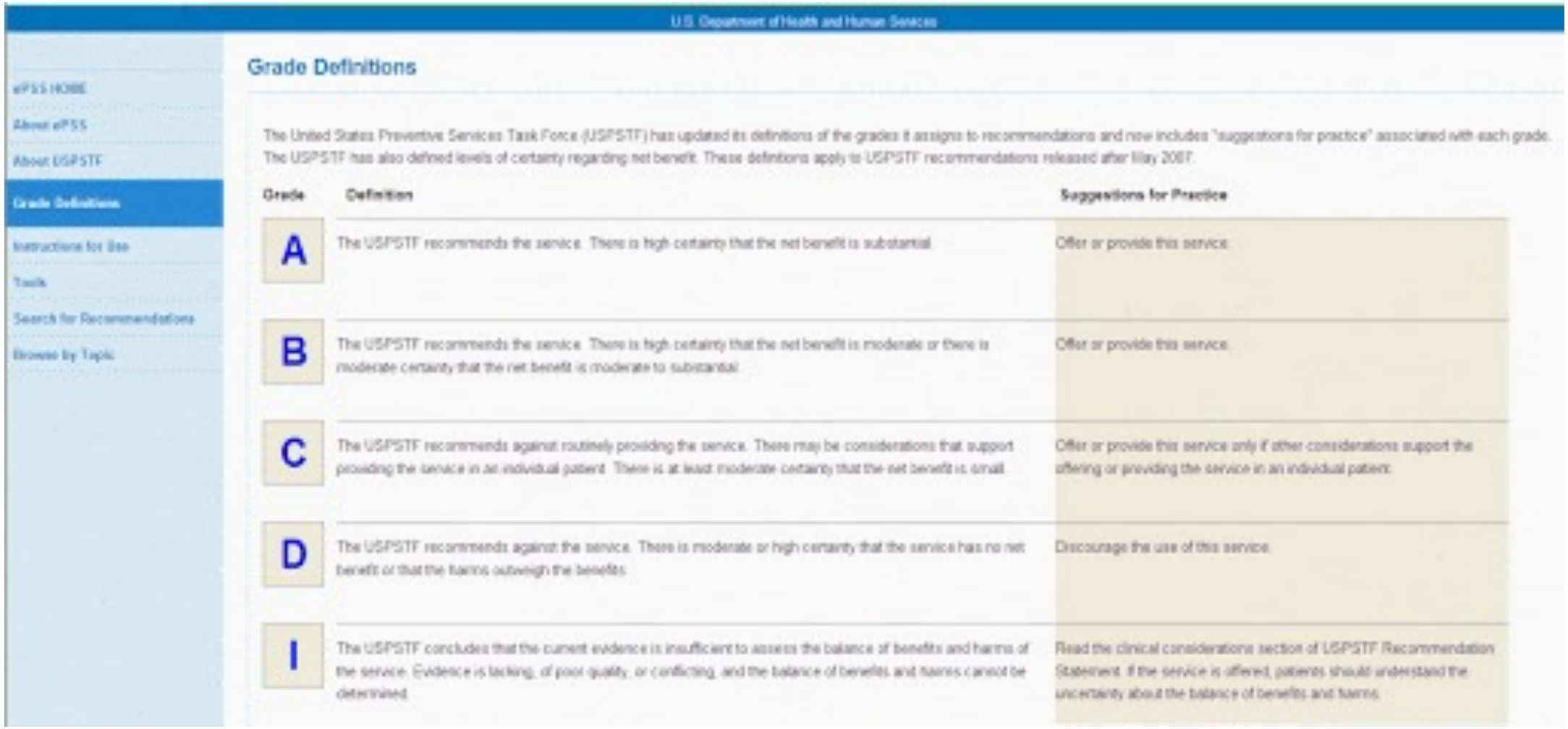
Radon: uncertainty distribution for attributable risk (BEIR VI)





Uncertainty distributions for the population attributable risk for males. I. Uncertainty in model parameters. II. Variability in K; variability in radon levels. III. Uncertainty/variability in K; variability in radon levels.

US PSTF levels: a benefits assessment



U.S. Department of Health and Human Services

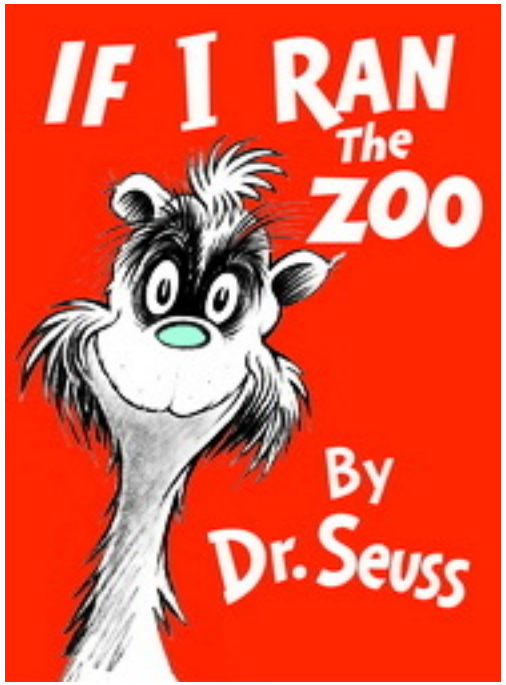
ePSS HOME
About ePSS
About USPSTF
Grade Definitions
Instructions for Use
Tools
Search for Recommendations
Browse by Topic

Grade Definitions

The United States Preventive Services Task Force (USPSTF) has updated its definitions of the grades it assigns to recommendations and now includes "suggestions for practice" associated with each grade. The USPSTF has also defined levels of certainty regarding net benefit. These definitions apply to USPSTF recommendations released after May 2007.

Grade	Definition	Suggestions for Practice
A	The USPSTF recommends the service. There is high certainty that the net benefit is substantial.	Offer or provide this service.
B	The USPSTF recommends the service. There is high certainty that the net benefit is moderate or there is moderate certainty that the net benefit is moderate to substantial.	Offer or provide this service.
C	The USPSTF recommends against routinely providing the service. There may be considerations that support providing the service in an individual patient. There is at least moderate certainty that the net benefit is small.	Offer or provide this service only if other considerations support the offering or providing the service in an individual patient.
D	The USPSTF recommends against the service. There is moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits.	Discourage the use of this service.
I	The USPSTF concludes that the current evidence is insufficient to assess the balance of benefits and harms of the service. Evidence is lacking, of poor quality, or conflicting, and the balance of benefits and harms cannot be determined.	Read the clinical considerations section of USPSTF Recommendation Statement. If the service is offered, patients should understand the uncertainty about the balance of benefits and harms.

Last thoughts



- I would—
  - Pick a defensible and in-use approach for evidence synthesis. Toxicology will challenge. Available reviews to start?
  - Use a transparent and replicable evidence integration approach.
  - Classify strength of evidence for decision-making.
  - For exposure, make transparent choices without systematic review.