A Brief Overview of Expert Judgment

John Evans

3/4 October 2013

Workshop on Methods for Research Synthesis
Cambridge, MA

What is Expert Judgment ?

- A method for quantitatively characterizing the state of knowledge about an uncertain quantity. The method seeks:
  - To elicit a subjective probability distribution for the quantity of interest from each of several experts; and
  - to summarize these distributions, in order to provide insight about:
    - the extent of uncertainty;
    - the key sources of uncertainty; and
    - the extent of, and reasons for, agreement / disagreement among the experts.

Some Questions to Consider

- Why might you want to do this ?
- What would be needed ?
- What decisions would you have to make ?
- What kind of results might you expect ?
- Does it make sense ? …. (never, always, sometimes)

Why Might You Want to Do This ?

- You are conducting a policy analysis, in which:
  - The value of the quantity of interest is crucial to policy decisions;

  AND

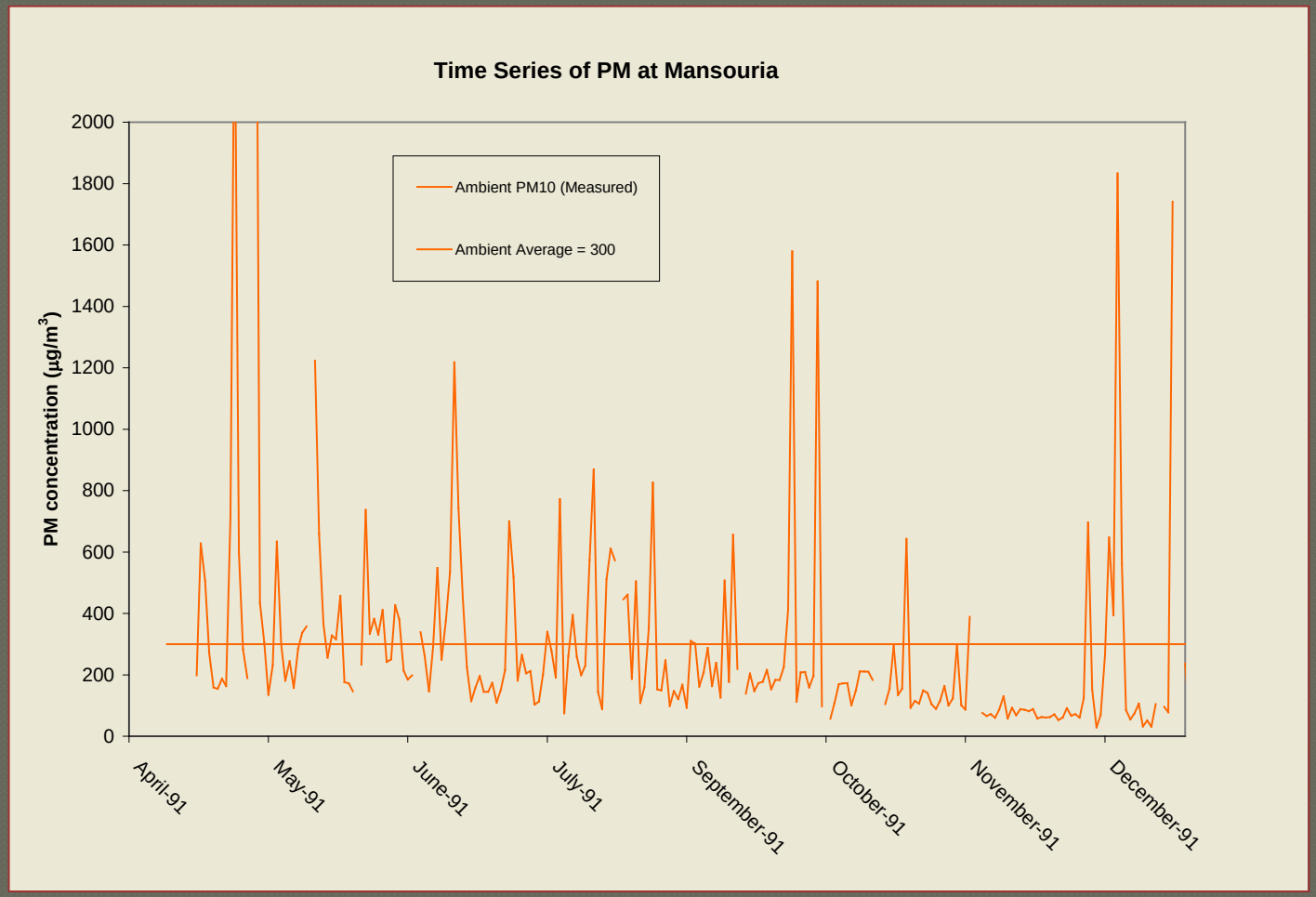
  - Ordinary statistical approaches (including meta-analysis) cannot provide an answer!

- Frequently this occurs when the question is one requiring extrapolation of findings from the setting in which they were obtained to another setting … and the issue involves judging the validity of the analogy which underlies the extrapolation.

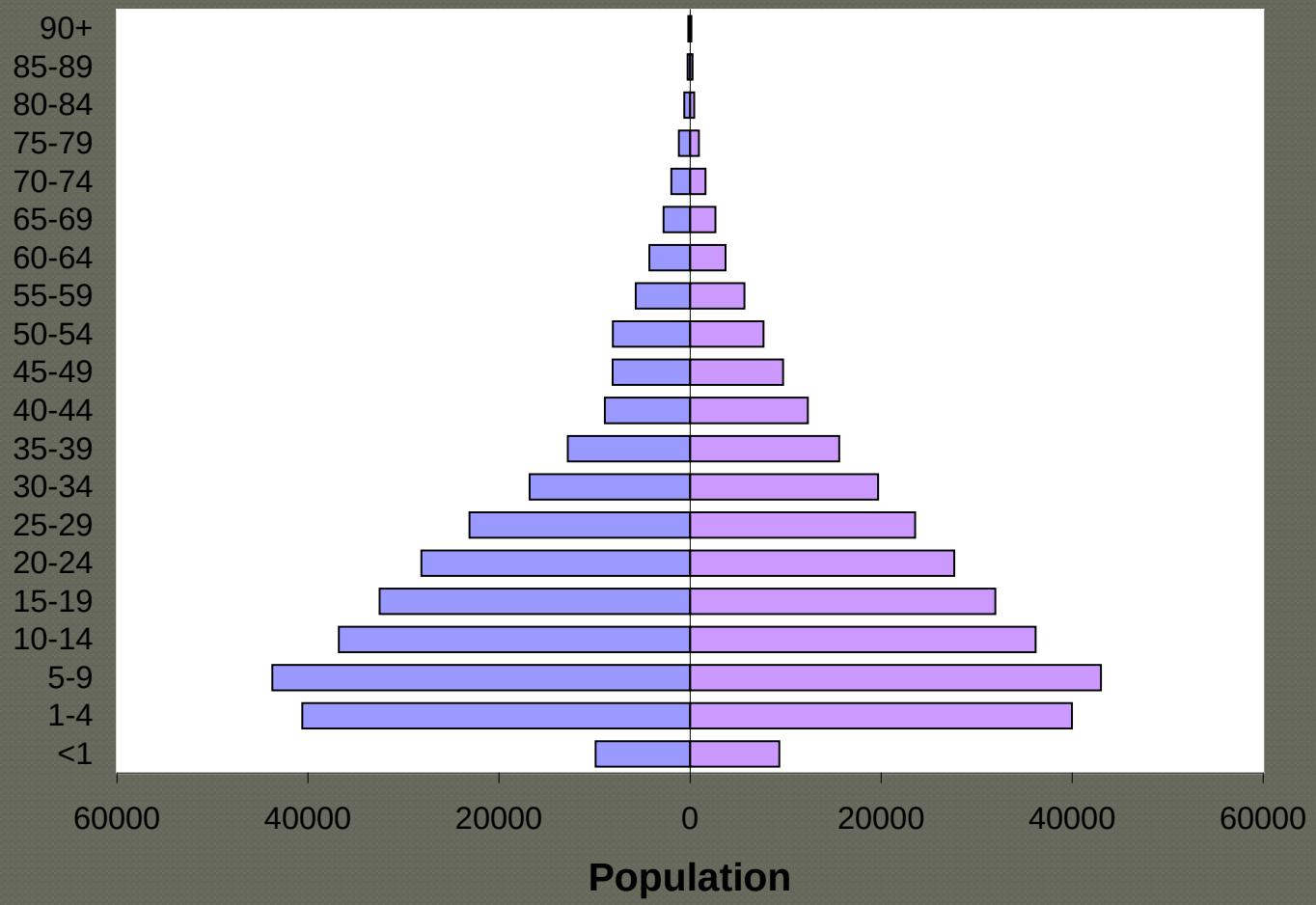
The Kuwait Oil Fires

Can Public Health Impacts be Estimated Using Studies of Lower Levels of Ambient $PM_{2.5}$ in the US ?

PM10 Levels in US Studies ~ 20 to 50 μg/m^3 Levels in Kuwait Fires Many Times This Large



US Studies Focus on People ≥ 25/30 Years Old
The Kuwaiti Population is Very Young



What Would You Need?

- A carefully framed question..(passes clairvoyance test)
- Some experts … (usually 4; 6-7; 9)
- A team of elicitors …(normative and subject area)
- Some time … (6; 12; 18+ months)
- Some money … (1/4; 1/2; 1+ million)

A Carefully Framed Question

The "Clairvoyance" Test

- A question without ambiguity about time, place, or values of other key parameters.

Question	*Setting*	*Exposure (Effect Interval)*	*Change*	*Pollutant*	*Composition*	*Baseline*
1	US	Long-term	1 μg/m^3	$PM_{2.5}$	Ambient	18 ug/m3

What is your estimate of the true, but unknown, percent change in the total annual, non-accidental mortality rate in the adult U.S. population resulting from a permanent 1 μg/m^3 reduction in long-term annual average $PM_{2.5}$ (from a population-weighted baseline concentration of 18 μg/m^3) throughout the U.S.? To express the uncertainty associated with the concentration-response relationship, please provide the 5th, 25th, 50th, 75th, and 95th percentiles of your estimate.

5% :_____ 25%:_____ 50% :_____ 75%:_____ 95%:_____

Some Experts

- Identifying experts
  - nomination by problem owner
  - publication / citation counts
  - nomination by peers / science policy panels
  - membership on NAS panels / editorial boards
- How many ? … (typically 4, 6-7, 9)
- Balance ? … (disciplinary / institutional / political)

Some Experts

Epidemiology of PM

- **H. Ross Anderson, M.D.**
  - Professor of Medicine, University of London, England
- **Bert Brunekreef, Ph.D.**
  - Professor of Environmental Epidemiology, University of Utrecht, the Netherlands
- **Ken Donaldson, M.D.**
  - Professor of Medicine, University of Edinburgh, Scotland
- **Nino Kuenzli, M.D., Ph.D.**
  - Assistant Professor, University of Basel, Switzerland (now at USC)
- **Juha Pekkanen, M.D., Ph.D.**
  - Head of Environmental Epidemiology, National Institute of Public Health (KTL), Finland
- **Annette Peters, M.D., Ph.D.**
  - Assistant Professor, GSF National Research Center for Environment and Health, Germany

What Decisions Would You Have to Make?

- Whether to hold an expert workshop.
- Whether to ask a single question
  … or a disaggregated set of questions.
- How to ask the questions
  … and whether to use elicitation aids.
- Whether to combine results across experts
  … and if so, how.

The Workshop

- Review Evidence
  (convince them that you're serious and dedicated)
- Develop Final Question(s) / Protocol
- Familiarize with Elicitation Procedure
- Discuss Biases & Heuristics
  (overconfidence, availability, anchoring & adjustment, …)
- Conduct Calibration Exercise

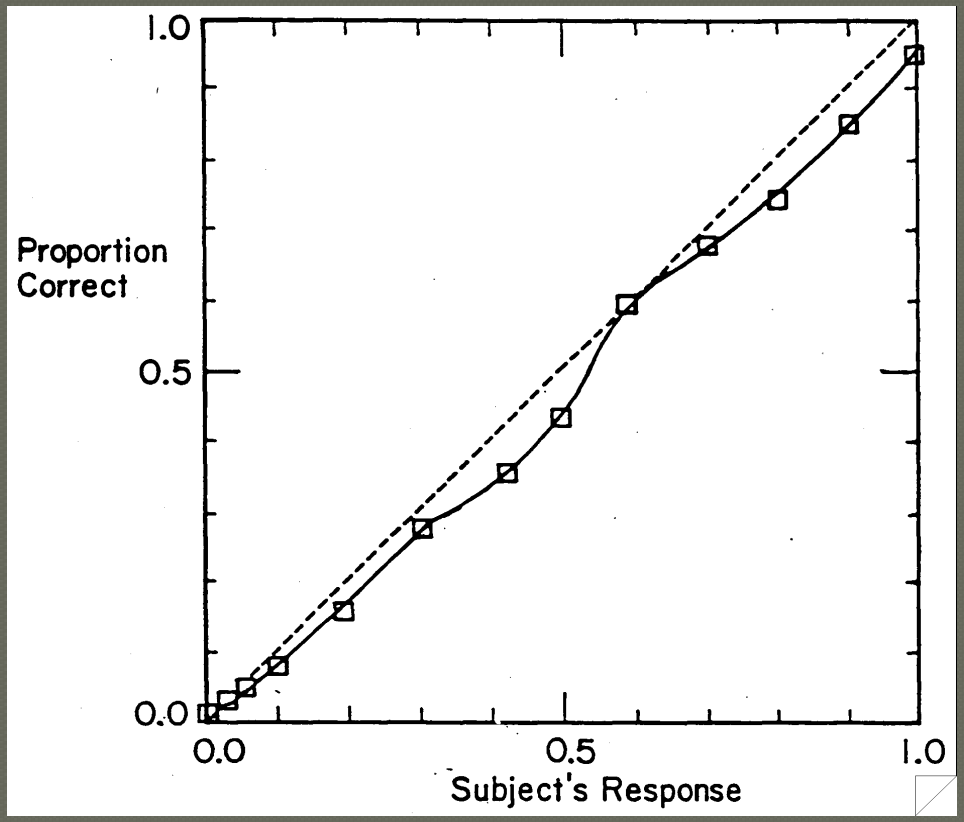
Calibration

- Calibration -- is a measure of the relationship between a subject's assessed probabilities and the actual frequency of occurrence of the associated events.

- If a "well calibrated assessor" tells you that there is an 80% probability that each of 100 discrete events will occur, then about 80 of them should actually occur.

Well-Calibrated Experts

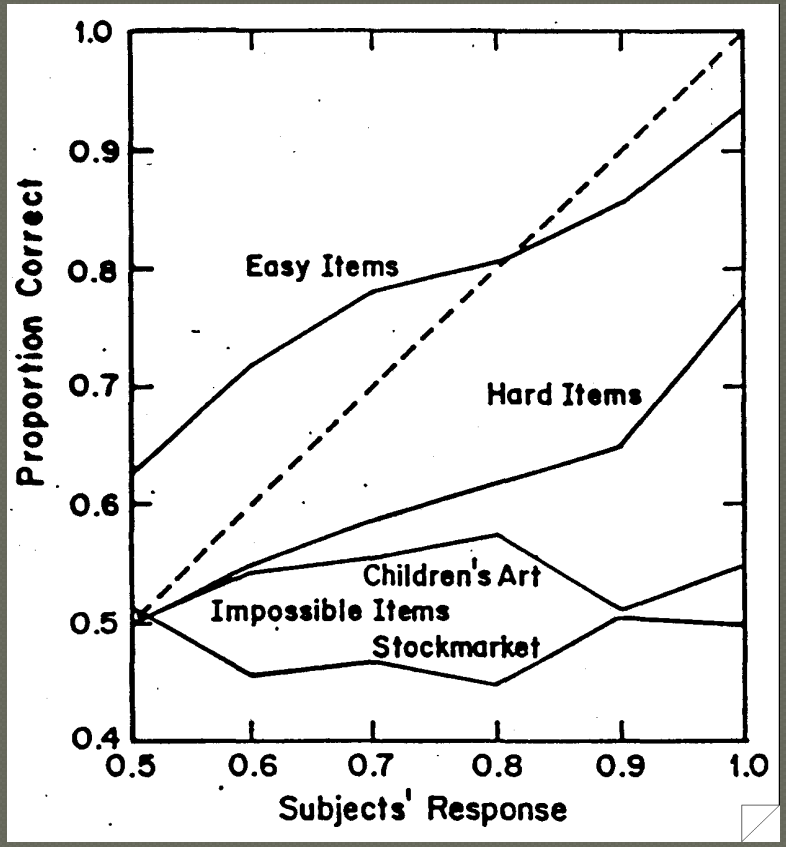
Weather Forecasters



Redrawn from results reported in A.H. Murphy and R.L Winkler "Reliability of subjective probability forecasts of precipitation and temperature", Applied Statistics 26, 41, 1977

Calibration of Lay Public

Easy, Hard & Impossible Items



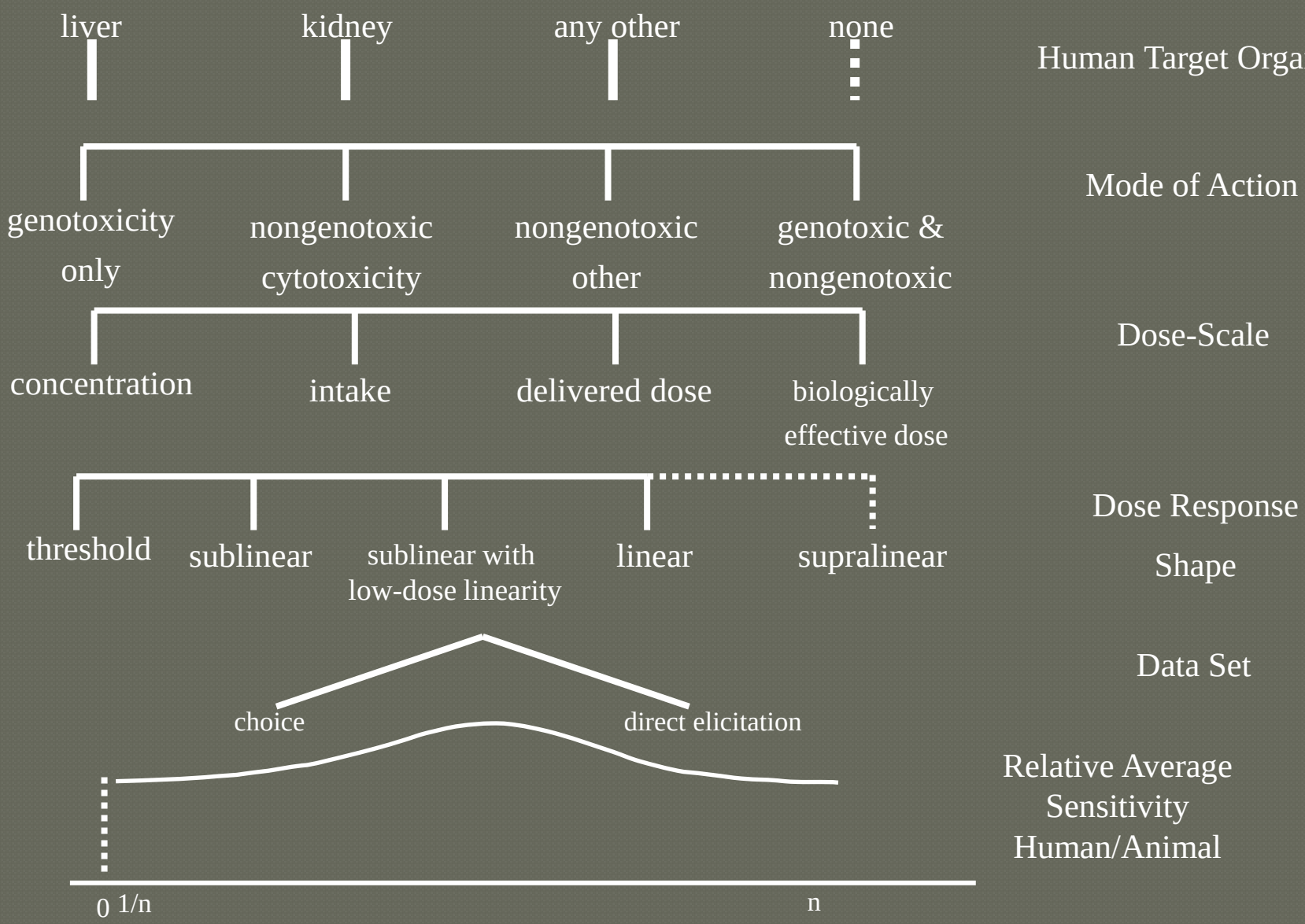
Note: These are half range items, e.g., which is longer the Suez Canal or the Panama Canal? How sure are you of your answer?

Redrawn from results reported in S. Lichtenstein and B. Fishhoff, "Do those who know more also know more about how much they know?", Organizational Behavior and Human Performance, 20, 159-183, 1977

Nature of Questions

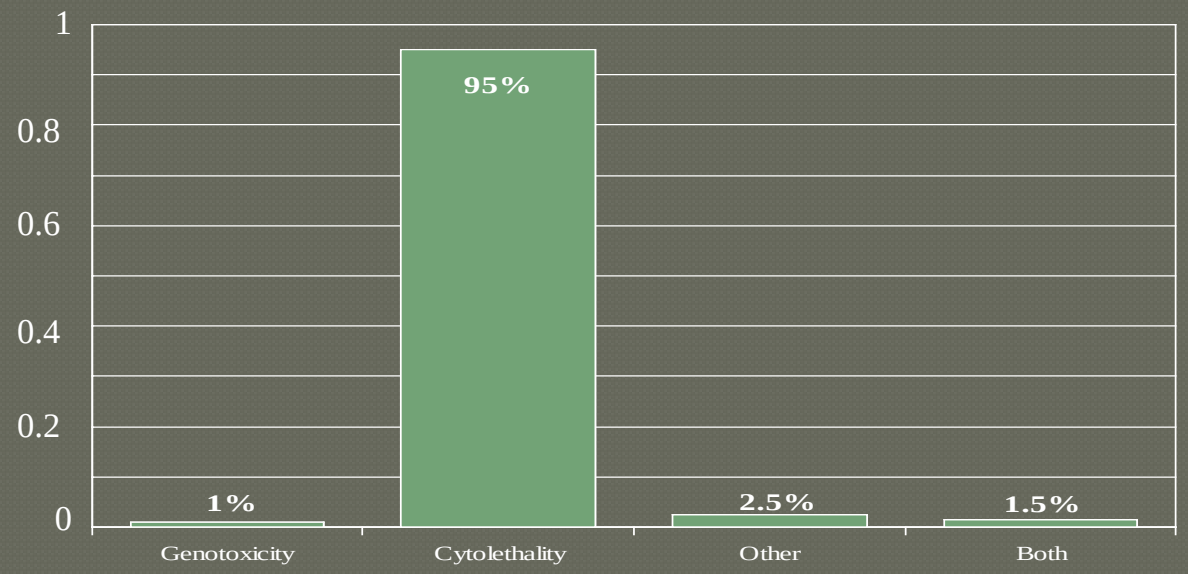
- Direct / Sequence – Oil Fires in Kuwait
  - Q1/2 – Long-Term 1 ug/3 Reduction in US/EU
  - Q3/4 – One Day 1 ug/m3 Reduction in US/EU
  - Q6/7 – Fraction of Effect Expressed in 1 Wk / 3 Mo Months
  - Q8/9 – Most/Least Toxic Constituents
  - Q10/11 – Deaths from the Oil Fires
- Disaggregate – Chloroform
  - Probability Tree

Probability Tree

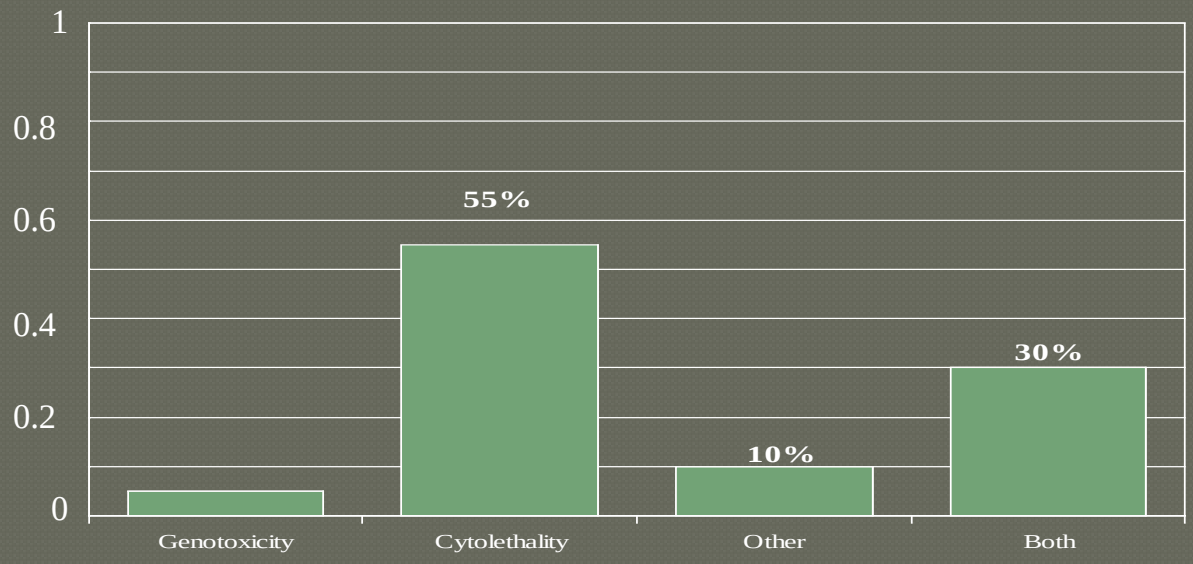


Liver Tumors: Mechanism?

Expert A

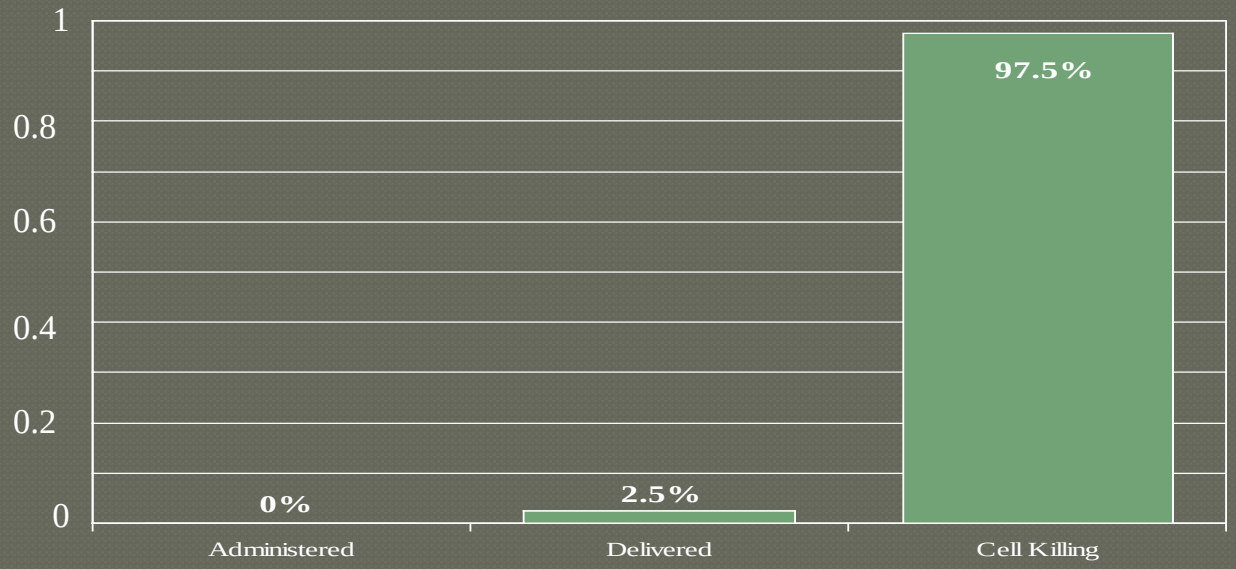


Expert E

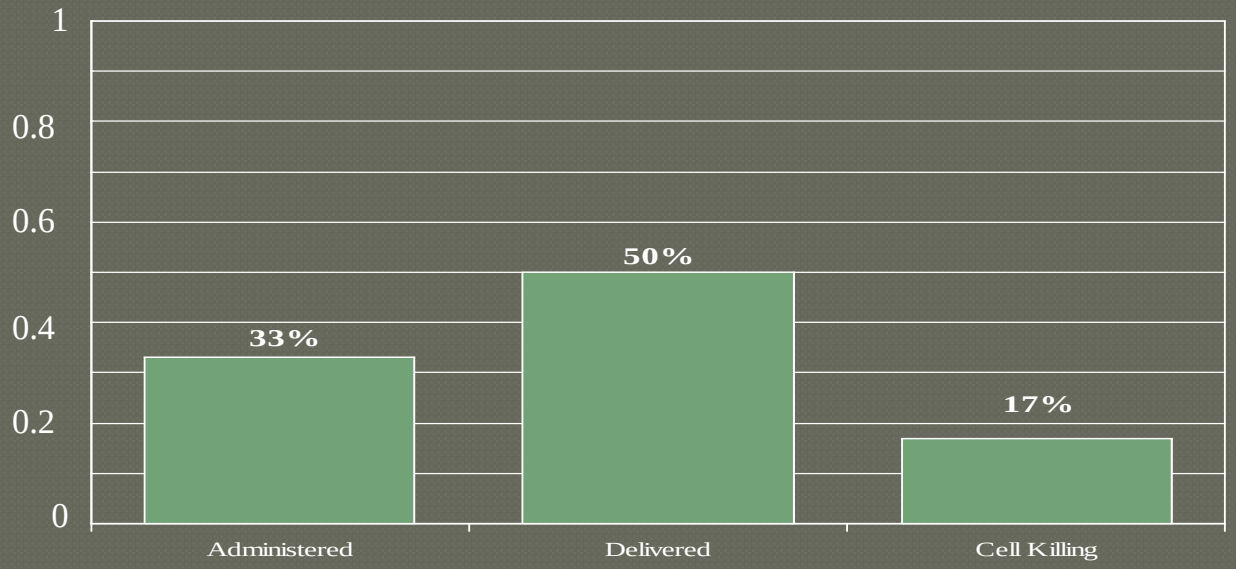


Liver Tumors / Mechanism Cytolethality / Dose Measure ?

Expert A



Expert E



The Interviews

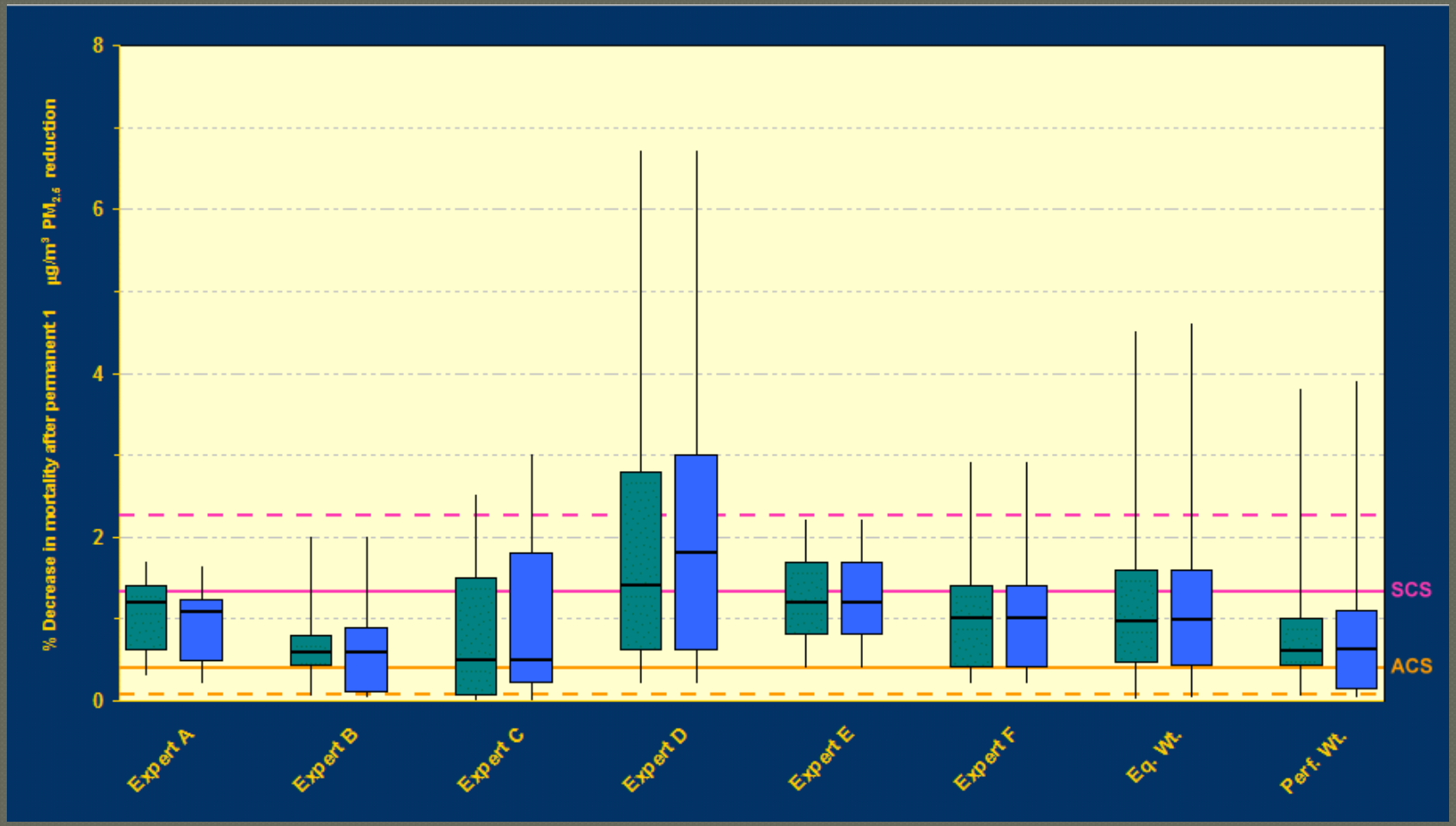
- Normally conducted in the expert's office… with full resources at hand.
- Typically begin with review of question of interest, review of interview protocol, informal discussion of available evidence.
- Then proceed to "warm up" questions:
  - What are the key properties of an ideal epidemiology study for measuring long-term mortality impacts of PM exposure?
  - Similarly, what are the key properties of an ideal epidemiological study of short-term mortality impacts of PM exposure?
  - What factors need to be considered to decide whether epidemiology results should be viewed as causal?

Elicitation
Basic Approach to Questions

- What evidence suggests large values for this relationship?
  - What is the highest plausible value?
  - Tell us a little about your reasoning, the evidence, and theories that lead you to this value.
  - Can you tell us of scenarios that would yield higher results?
- What evidence or theory suggests small values?
  - What is the lowest plausible value?
  - Tell us a little about your reasoning, the evidence, and theories that lead you to this value.
  - Can you tell us of scenarios that would yield lower results?

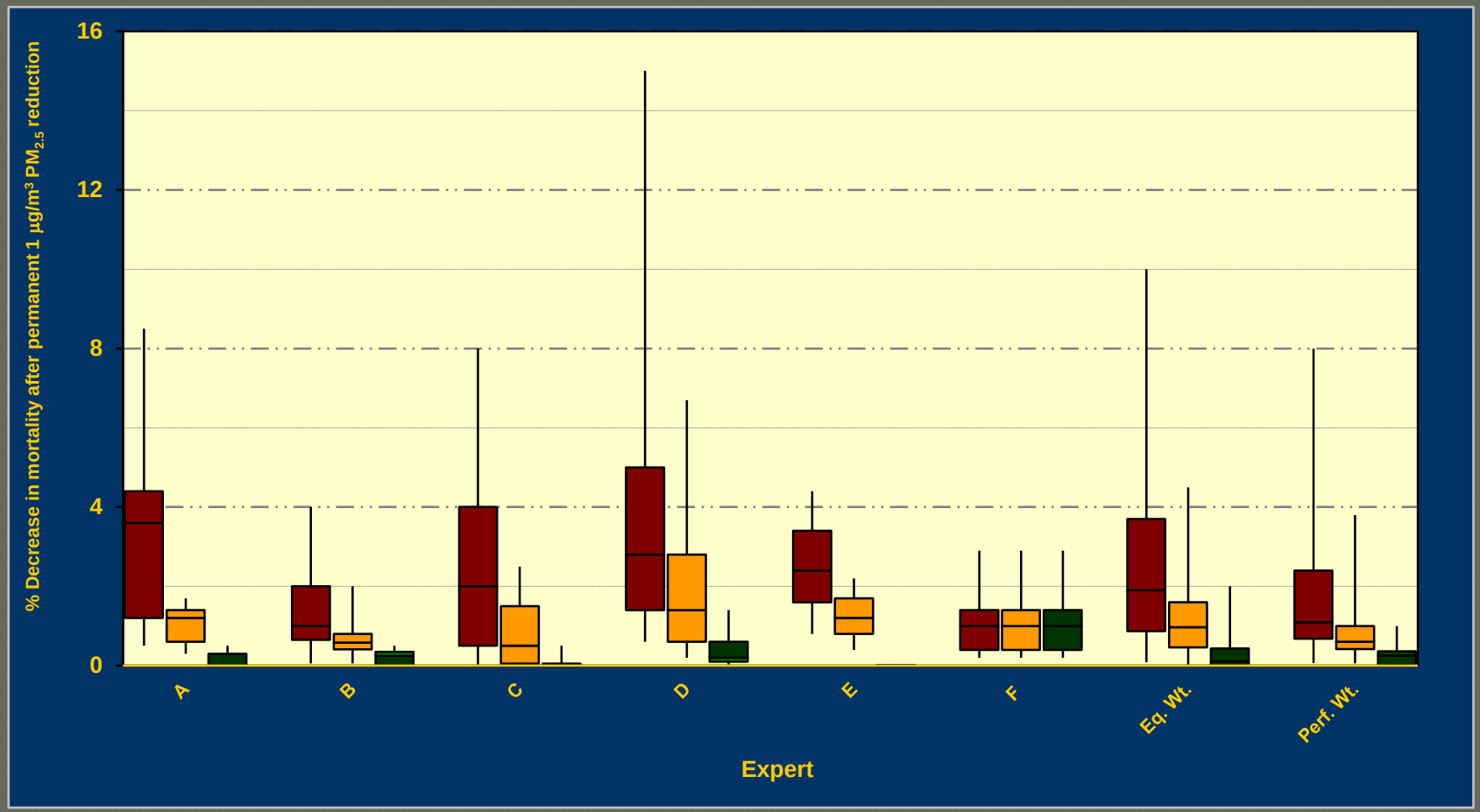
What Kind of Results Might You Get?

Long-Term Effect of Permanent 1 μg/m^3 PM2.5 Reduction in US



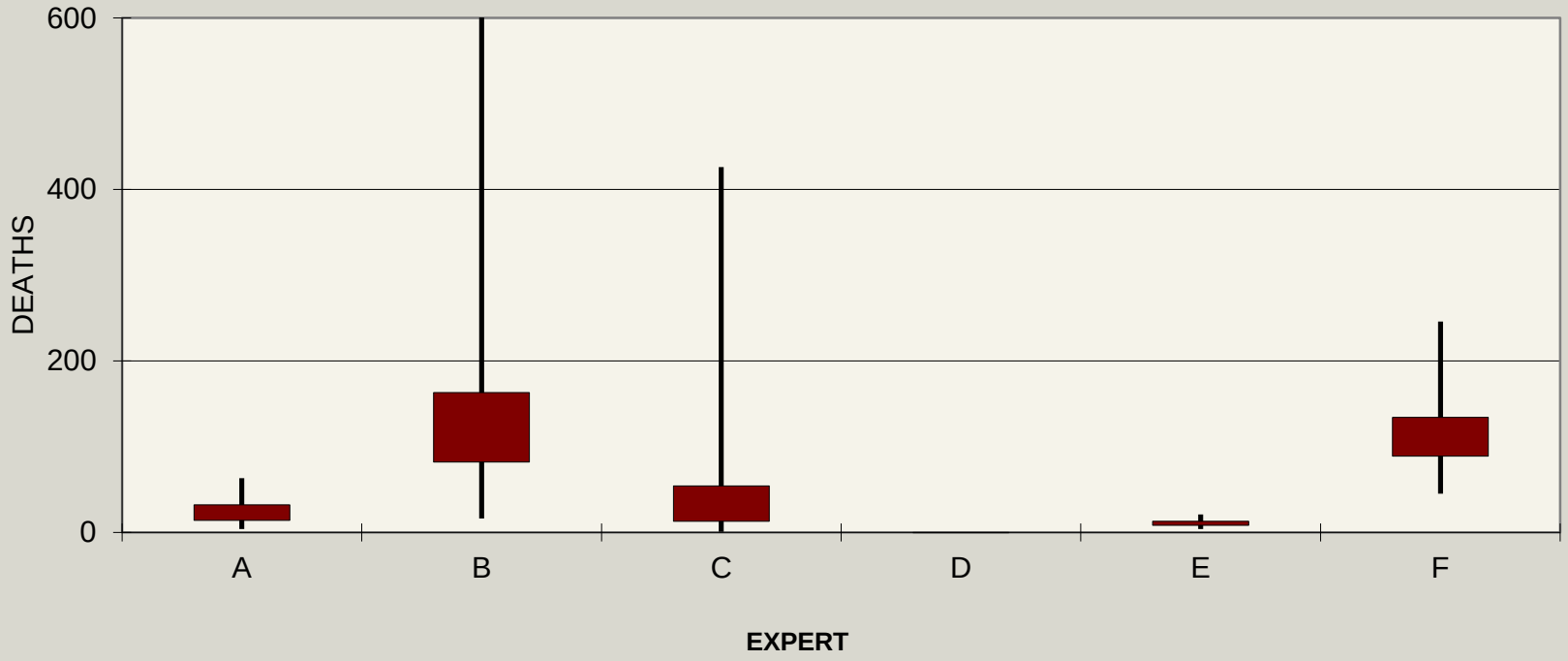
What Kind of Results Might You Get?

Long-Term Effect of 1 μg/m^3 Reduction of Most /Least Toxic Constituent



What Kind of Results Might You Get?

Deaths Attributable to Kuwait Oil Fires



The Answer

Deaths Attributable to Exposure to Smoke from the Kuwait Oil Fires

*Expert **	*5%*	*50%*	*95%*	*Approach*
E	*6*	*13*	*21*	*TS*
A	*4*	*32*	*63*	*TS*
C	*<1*	*54*	*426*	*C*
F	*37*	*110*	*210*	*C*
B	*16*	*164*	*872*	*C*
D	*575*	*2874*	*11496*	*C*

- Experts are listed in order of their median estimate of risk.
- Letter identifications are randomly assigned to experts.

Multiple Experts – Combining Judgments

Morgan & Cooke, RFF Workshop, March 2006

- Morgan – "When different experts hold different views it is often best not to combine the results, but rather to explore the implications of each expert's views so that decision makers have a clear understanding of whether and how much the differences matter in the context of the overall decision."

- Cooke – While it is important to provide the individual results, it is also important for the analyst to assist the problem owner create a meaningful synthesis of the results. Arguably performance weighting provides the most informative synthesis.

Some Closing Thoughts

Cooke, RFF Workshop, March 2006

- Experts CAN quantify uncertainty using subjective probability.
  - TU Delft EJ database (as of March 2006) 45 projects / 521 experts / 3688 variables
- EJ is NOT knowledge.
  - The scientific method, not EJ methods, produces agreement among experts.
  - EJ is for quantifying, not removing, uncertainty.
  - Ask experts only about uncertainty with respect to possible measurements.
  - *Not every problem is an EJ problem*.

The Last Word

Morgan, RFF Workshop, March 2006

- ”The elicitation of expert judgment, often in the form of subjective probability distributions, can be a useful way ***to combine the formal knowledge*** in a field as reflected in the literature ***with the informal knowledge and physical intuition of experts***.”

- “Elicitation is not a substitute for doing the needed science, but it can be a very useful tool in support of research planning … and in the formulation of public policy.”