

**Compilation of SAB Expert Elicitation Advisory Panel Members' Comments on  
Review of EPA's Draft Expert Elicitation Task Force White Paper (Draft 04/06/09).**

**Comments Received:**

<b>Comments from Dr. William Ascher .....</b>	<b>2</b>
<b>Comments from Dr. Mark Borsuk.....</b>	<b>3</b>
<b>Comments from Dr. Wändi Bruine de Bruin.....</b>	<b>5</b>
<b>Comments from Dr. Roger Cooke.....</b>	<b>8</b>
<b>Comments from Dr. John Evans .....</b>	<b>11</b>
<b>Comments from Dr. Scott Ferson.....</b>	<b>14</b>
<b>Comments from Dr. Christopher Frey .....</b>	<b>16</b>
<b>Comments from Dr. Alan Krupnick .....</b>	<b>19</b>
<b>Comments from Dr. Mitchell Small .....</b>	<b>20</b>
<b>Comments from Dr. Thomas Wallsten .....</b>	<b>21</b>
<b>Comments from Dr. Katherine Walker.....</b>	<b>22</b>

**Comments from Dr. William Ascher**

To: Angela Nugent  
From: Bill Ascher  
Date: April 10, 2009  
Re: Draft Review Panel Report on the Expert Elicitation White Paper

I have little to add to this well-done draft.

p. 6, line 31: "she must choose a single policy". This would be confusing to some readers: often policymakers will choose an ensemble of policies, or will have a contingency policy or policies as knowledge unfolds. This clause can simply be deleted.

There are some typos:

p. 7: line 30 path of Brownian motion to the mixing layer of the atmosphere, but it is not **know** how it

p. 8: line 5 when **thing** about "clairvoyance."

Regarding adding to the reference (p. 9, line14) we can cite:

Gilovich, Thomas, Dale Griffin, and Daniel Kahneman, eds. 2002. *Heuristics and biases: the psychology of intuitive judgment*. Cambridge: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky, eds. 2000. *Choices, values, and frames*. Cambridge: Cambridge University Press.

One substantive recommendation:

p. 9, line 37: More than just the generic statement that the white paper "should encourage research in the performance of alternative methods," the review panel should recommend that the white paper should engage in assessments of the performance of alternative expert elicitation methods. It would be worthwhile for the review panel to discuss how to do this feasibly, whether by doing a meta-analysis of studies of expert elicitation accuracy, or commissioning new assessments. If this is to be part of the panel's report, then it should be added to the otherwise fine summary letter to the Administrator.

**Comments from Dr. Mark Borsuk**

I'd like to commend Jim on doing a nice job of summarizing the panel's comments in a concise and readable way. The style, length, and content are all appropriate. My suggestions for improvement are along two main lines (in order of importance): (A) I felt that the cover letter did not bring to the forefront the points that were actually the most important suggestions. This is critical, as the letter serves as a kind of executive summary of the full report. (B) There are a few paragraphs of the report that could be written a bit more clearly. I'll offer my suggestions on each in turn.

**(A) Revisions to cover letter:**

I suggest replacing the bulleted points on page 1 (lines 33-44) and the first paragraph of page 2 (lines 1-11) with the following numbered suggestions:

1. Give greater attention to the extent to which EE is a complement rather than a substitute for other methods of quantifying uncertainty about a quantity or model parameter. We suggest that EE should be presented as a useful way to organize and synthesize what is already known about a quantity and not as a means for generating new primary data.
2. Address methods for evaluating and ensuring the quality of expert judgments, including tests of internal consistency, coherence and, when possible, performance.
3. Discuss the issue of transparency in the context of the time and costs involved. Some methods of elicitation would be extremely difficult to fully document, suggesting a tradeoff between full transparency and limited resources.
4. More fully address methods for aggregating experts' judgments. In most cases, aggregation is necessary for subsequent use of elicited quantities.
5. More carefully delineate the types of quantities suitable for EE. The panel urges that, when possible, the quantities being elicited be measurable (at least in principle, if not in practice). Only when experts agree on a common model that permits unambiguous translation from an unobservable parameter to a measurable quantity should values of parameters be elicited directly.
6. Give greater attention to the need to explicitly condition the quantities being elicited on other quantities of relevance. This is important because both the value and the uncertainty of most quantities will be dependent on the values of other quantities. Also, the specific nature of dependencies among multiple quantities being elicited may be required for subsequent use. The panel suggests that explicit conditioning may often be facilitated by the use of influence diagrams.
7. More fully review the literature on cognitive biases which may lead to incorrect elicitation of expert judgments.
8. Emphasize the need for flexibility in EE implementation. The panel suggests that the EPA be careful not to stifle innovation in EE methods by prescribing "checklist" or "cookbook" approaches. Rather, EE guidance should be in the form of goals and criteria for evaluating success that can be met by multiple approaches.

**(B) Suggested modifications of report text (specific modifications in red):**

1. Page 3, lines 40-42: Replace with: "...including tests for **internal consistency**, coherence in judgments over multiple factors, and, **when possible**, performance (measured **using quantities** the values of which **are known or will soon become known**)."
2. Page 4, lines 29-32: Add "definitions of the quantities being elicited" and "the relevant quantities upon which the quantity being elicited is being conditioned" to the list of items requiring transparency.
3. Page 4, lines 37-39: Replace with: "It should also discuss how deepening the interactions between elicitors and experts **makes documentation more difficult and expensive**."
4. Page 6, lines 24-27: Move point #1 to the end of the list as point #3. This makes for a more logical flow.
5. Page 7, lines 14-21: Throughout this paragraph, the word "parameter" should be replaced by "quantity" to be more general, especially since the preceding paragraphs discourages elicitation of parameters.
6. Page 7, line 18: After "...unspecified." Add the sentence, "The influence of unspecified factors on the quantity being elicited becomes part of the uncertainty in the value of that quantity."
7. Page 7, line 20: Insert the parenthetical "(ignoring correlation)", after "using marginal distributions".
8. Page 7, line 33-34: Add "conditional on others" after "potentially measurable quantity".
9. Page 8, lines 3-5: Replace sentence with, "The **need to explicitly maintain** a consistent conditionalization across multiple factors is easily neglected when **thinking** about 'clairvoyance.'"
10. Page 8, lines 7-8: Replace with, "...use of an influence diagram. **The panel notes that the example shown in Figure 6.1 of the white paper is not the type of influence diagram typically used in decision theory. We recommend a more relevant and clearly described example. (As a related note, the panel recommends that all figures...)**"
11. Page 9, lines 14-16: Replace last sentence of paragraph with "Strategies for overcoming these heuristics and biases to ensure accurate and honest assessments should be discussed in detail."

**Typos:**

12. Page 4, line 35: Replace "requiring" with "require".
13. Page 7, line 30: Replace "know" with "known".
- Page 9, line 37: Replace "is" with "it".

## **Comments from Dr. Wändi Bruine de Bruin**

Comments by Wändi Bruine de Bruin

Thank you for doing a great job at summarizing the discussions we had during the meeting. My comments focus on the transparency section, which seems to be missing:

- The importance of making sure that questions asked are similarly understood by participating experts from all relevant disciplines. The design of the questions should follow good practices in survey design, including:
  - the pilot testing of questions and recording qualitative explanations of quantitative responses, which should help with the interpretation and communication of the results, especially in situations where there is large disagreement between experts.
  - the use of question wording and response formats that reduce the use of “50%” as an expression of not knowing the answer, rather than a quantitative probability.
  
- The importance of making sure that public reports for audiences of non-experts are understood by its intended consumers (e.g., policy makers, others?). The design of these communications should follow good practices in risk communication, including:
  - a focus on decision-relevant information
  - presenting quantitative results with clear qualitative explanations (which should be elicited from experts, as described above)
  - the use of non-technical jargon

I am attaching my slides to my comments, to provide details on each, as well as a list of relevant references.

The letter doesn't seem to include suggestions related to transparency. Perhaps it should include the suggestion to be more explicit about procedures to improve transparency.

### **Transparency of EPA expert elicitation**

Wändi Bruine de Bruin, PhD  
Carnegie Mellon University  
Dept. of Social and Decision Sciences

### **Transparency to whom**

Experts participating in the elicitation  
Users of the expert elicitation  
Policy makers  
Stakeholders  
Members of the general public  
Review panel

## **Transparency to experts in the elicitation**

Good survey questions should be interpreted in the *same way* by question designers and respondents from different backgrounds  
allow respondents to express their full set of beliefs  
Failing to write good survey questions leads to missing, invalid, and protest responses  
Including saying “50%” in response to quantitative probability questions (Bruine de Bruin et al., 2000; 2002)

## **How to develop good survey questions**

Involve experts from all relevant disciplines in pilot tests *before* conducting elicitation  
Include read-aloud of protocol to ensure (shared) understanding  
Invite them to add questions to express relevant beliefs  
Provide clear instructions on how to answer questions  
Ask quantitative questions that  
Can be answered by experts in all relevant disciplines  
Are specific enough to have an answer (under “clairvoyance”)  
Avoid mental gymnastics as much as possible  
Write probability questions that avoid 50% responses  
Present linear probability scale rather than fill-in-the-blank  
Ask about “the percent of people” rather than “the probability that a person” will experience an event  
Allow for “don’t know” response, or ask what 50% meant  
Ask for explanations of quantitative responses

## **Transparency to users of expert elicitation**

Effective risk communications  
Should help policy makers to make more informed decisions  
Should be understood by all of its potential users  
Existing communications are often not effective, because they  
are not written with users in mind  
do not provide decision-relevant information  
use expert jargon  
Users may only read summaries or press releases  
Press releases (of medical studies) often overstate results, failing to mention study limitations and industry funding (Woloshin & Schwartz, 2002)

## **How to develop effective communications**

Involve users *before* developing communications, even *before* conducting the elicitation  
Find out what they need to know to make more informed decisions  
Ensure that they understand and trust every step of the elicitation

### Present communications in simple terms

6<sup>th</sup> grade reading level is recommended for public health pamphlets, can be used to present complex information, and benefits all readers (*i.e. 18 is too high*)

Use simple graphs and explain them in the accompanying text

Pilot-test communications before releasing them

Conduct read-aloud protocols with users to ensure understanding

Fact-check with experts to ensure accuracy

Use systematic presentation format for overall report, executive summaries *and* press releases

## Examples of topics

### to systematically cover in reports

#### Research question

What is the main research question and what policy question will it inform?

Why was expert elicitation needed to answer the main research question?

#### Methods

How was the elicitation conducted?

Who were the experts, how were they selected, and did they represent all relevant views and disciplines?

#### Results

What is the degree of consensus?

Why did the experts disagree if/when they did?

#### Conclusions

How do these results inform the policy question?

What are the main limitations?

## Review panel

Charge questions for the review panel should cover

The expert elicitation

The communication of results

The review panel should include experts from relevant disciplines

Substantive experts from relevant disciplines

Technical experts

Survey design experts

Risk communication experts!

Intended users?

Bruine de Bruin, W., Fischbeck, P.S., Stiber, N.A. & Fischhoff, B. (2002). What number is “fifty-fifty”? Redistributing excess 50% responses in risk perception studies. *Risk Analysis*, 22, 725-735.

Bruine de Bruin, W., Fischhoff, B., Brilliant, L., & Caruso, D. (2006). Expert judgments of pandemic influenza risks. *Global Public Health*, 1, 178-193.

Bruine de Bruin, W., Fischhoff, B., Millstein, S.G. & Halpern-Felsher, B.L. (2000). Verbal and numerical expressions of probability: “It’s a fifty-fifty chance.” *Organizational Behavior and Human Decision Processes*, 81, 115-131.

Bruine de Bruin, W., Parker, A.M., & Fischhoff, B. (2007). Individual differences in Adult Decision-Making Competence. *Journal of Personality and Social Psychology*, 92, 938-956.

Fischhoff, B. & Bruine de Bruin, W. (1999). Fifty-fifty=50%? *Journal of Behavioral Decision Making*, 12, 149-163.

Fischhoff, B. (1994). What forecasts (seem to) mean. *International Journal of Forecasting*, 10, 387-403.

Morgan, M.G., Fischhoff, B., Bostrom, A., & Atman, C. (2001). *Risk communication: The mental models approach*. New York: Cambridge University Press.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods and the logic of conversation*. Hillsdale, NJ: Erlbaum.

Woloshin, S., & Schwartz, L.M. (2002). Press releases: Translating research into news. *Journal of the American Medical Association (JAMA)*, 287, 2856-2858.

## Comments from Dr. Roger Cooke

Comments on Draft SAB EJ report  
Roger Cooke  
04-15-09

Excellent summary by Jim Hammitt. I have a few comments which might be incorporated.

**Comment 1:** The policy with regard to the use of expert names, attributed to Cooke 91 is garbled. The EU procedures guide reflects the policy that has been followed through all of the EU-USNRC work and many other applications. It is described and motivated in The EU Procedures Guide, which itself is based on the extensive experience gained in the EU – US Nuclear Regulatory Joint Uncertainty Studies of Accident Consequence models. The following excerpt describes the method for identifying and selecting experts, and the policy regarding expert names, used in that suite of studies.

From R.M. Cooke, L.H.J. Goossens, B.C.P. Kraan  
**Probabilistic Accident Consequence Uncertainty Assessment  
Procedures Guide Using Expert Judgement**  
EUR 18820EN European Commission. Luxembourg 2000, Euratom.

### **“ 2.6 Selection of experts**

*After the set of experts has been identified, a choice is made which experts to use in the study. In general, the largest number of experts consistent with the level of resources should be used. In any event, at least four experts for a given subject should be chosen. A panel of eight experts is to be recommended as a rule of thumb. The choice should be made so as to diversify the knowledge bases and institutions of employment. At least two experts should be from outside the institution performing the study,.... The following general selection criteria are used:*

- *reputation in the field of interest*
- *experimental experience in the field of interest*
- *number and quality of publications in the field of interest*
- *diversity in background*
- *awards received*
- *balance of views*
- *interest in and availability for the project*

*The nature and the broadness of a panel may require experts with very broad experience for which only a few are available (generalists). Panels may need a diversity of in-depth expertise; a mix of generalists and specialists. The requirement for the specialists is that they cover the whole panel's field sufficiently. For instance, in a panel on health effects, specialists may be required on the various organs of a human body, whereby only a few generalists have experience in the whole field of health effects. The following selection procedure for experts is recommended:*

1. *All potential experts named during the expert identification phase will be contacted (by mail and later by personal contact) to find out whether they are interested and whether they consider themselves a potential expert for that particular panel. During personal contacts potential experts are also asked to name other potential experts.*
2. *Potential experts send in a CV (curriculum vitae) indicating their expertise and availability for that specified panel.*
3. *All CV's will be reviewed by a nomination committee consisting of one to three persons from the project staff and one or two additional persons with thorough expertise in the field of interest not being involved as a potential expert themselves.*

*When the list of candidate experts is determined, these are contacted and invited to participate in the study. It is essential to clarify the conditions of participation, including:*

1. *Type of assessment task*
2. *Remuneration*
3. *Distribution of study results*
4. *Use of the experts name*
5. *Feedback of expert judgment data*

#### *Use of Experts' Names*

*Every expert is very jealous of his/her name and professional reputation. It is essential to clarify how the names will be used. The following procedure, developed over a number of years, seeks to satisfy the demands of openness and objectivity in science, as well as demands of freedom from conflict of interest, harassment and legal liability which may legitimately be raised by the experts themselves. If indeed expert judgment is scientific data, then it must be open to peer review. On the other hand, the expert's affiliation and professional activities may create a conflict of interest if his/her name is associated with the actual assessments. If told that his name would be published with his assessments, an expert in toxicology working at a pharmaceutical company might well say, "if you want the company viewpoint, ask the president of the company".*

*The proposed procedure is the following:*

1. *Expert names and affiliations are published in the study.*
2. *All information, including expert names and assessments, is available for competent peer review, but is not available for unrestricted distribution.*
3. *Individual assessments are available for unrestricted distribution, assessments are not associated with names but identified as "expert 1, 2,3,..." etc.*
4. *Expert rationales are available for unrestricted distribution.*
5. *Each individual expert receives feedback on his/her own performance assessment.*
6. *Any further published use of the expert's name requires the expert's approval. " "*

**Comment 2:** The White Paper is a step in the right direction. A Procedures Guide or Best Practice document should draw on a wide experiential base. It is recommended that EPA conduct several EE studies on problems that are NOT hot-button issues, employing different methods, and evaluating results. A bench test study could also be useful, where different teams are invited to deploy different methods on a common problem. In this way we can we learn which methods are best for which problems. Considerable experience with structured expert judgment exists in other fields, including nuclear, aerospace, volcanology, health, environmental transport and finance. The challenge is to bring this experience to bear on the specific problem areas within EPAs mandate.

## Comments from Dr. John Evans

### Charge Question A – Background and Definition of Expert Elicitation

- o Page 2, paragraph beginning at line 26

§ I believe that all of this is true – but I’m not sure that it is relevant. We were asked whether the white paper provides a comprehensive accounting of the potential strengths, limitations and uses of EE. I believe that we concluded that it does. While it is true that the white paper does not discuss the strengths and weaknesses of alternative approaches I do not think that this is a legitimate criticism of the report. Perhaps this section could be reworded – to say that the strengths and weaknesses of EE could best be understood if presented in comparison with the strengths and weaknesses of other methods. I would agree with such a framing.

- o Page 2, line 38 – wording “extrapolate from” is ok, but not great

- o Page 3 – I love the entire page

### · Charge Question B – Transparency

- o Page 4 – lines 14-32 are great

§ But on line 25 there needs to be a “)” after the word “obsolete” before the “;”

- o Page 4 – lines 34 to end of page are very good

§ But in lines 39/40 – the wording “the report should assess each method in terms of how much of the process is intrinsically a black box” could be improved – the idea is a very good one.

- o Page 5 – lines 1 through 13 – ideas are good, but text seems verbose.

### · Charge Question C 1 – Selecting Experts

- o Page 5 – This section is generally quite good...

§ But, it should begin with a statement to the effect that the problem of expert selection is COMMON to any effort to use expert opinion in support of the development of regulatory policy – whether informal, formal, structured or unstructured – and therefore that the guidance offered below applies uniformly – and is not intended to be a critique of formal elicitation of expert opinion.

§ That said, the 1st paragraph (lines 26 – 31) is GREAT!

1. It could perhaps be improved by explicitly saying that these are conflicting goals – requiring the decision maker/analyst to find some compromise in any practical application.

§ The 3rd paragraph (lines 40-43) seems unclear to me and perhaps could be dropped.

§ The 4th paragraph seems sloppy and unfocused to me. An attempt should be made to rewrite this paragraph.

· Charge Question C 2 – Multi-expert Aggregation

o I like this entire section as is.

· Charge Question C 3 – Problem Structure

o The 1st (page 7, lines 7-12), 2nd (page 7, lines 14-21), and last (page 8, lines 12-21) paragraphs of this section are great.

§ The 2nd paragraph could cite our work on chloroform, which carefully considered conditionality

o The 3rd (page 7, lines 23-36) and 4th (page 7, line 38-45 and page 8, lines 1-5) paragraphs are ok – but are verbose

§ They should be better focused and tightened.

§ In some ways, I think that they go beyond the scope of the charge questions.

§ But, I believe that the issue they raise (clairvoyance – definition and limitations) is important enough to require two or three carefully written sentences.

§ Also, two small details should be corrected:

1. Page 7, line 31 “know” should be “known”

2. Page 8, line 5, “thing” should be “thinking”

· Charge Question C4 and C5 – Findings and Recommendations

o This section seems weak to me.

o The first paragraph is ok.

o The third paragraph is fine.

o The second paragraph seems out of place, and possibly goes beyond the scope of the charge.

§ Perhaps these comments could be moved to one of the other sections – they are closely related in some ways to the dependency discussion covered on page 7 of the response to charge question 3.

- Charge Question D – Development of Future Guidance

- o This section is fine.

**Comments from Dr. Scott Ferson**

Cover letter (second email)

>  
> Page 1, line 26: The word 'comprehensive' should be perhaps  
> replaced by 'reasonable'. Important things like numerical examples  
were left out.

>  
> Page 2, line 2: Insert the phrase 'EE in contrast with those of'  
> after 'weaknesses of'. We specifically asked for a discussion of  
> the drawbacks of EE itself which are conspicuously absent, and  
> which they acknowledged they had written but omitted in their  
> zeal.

>  
> Line 9: Change the phrase 'experiment with' to 'consider', which  
> does not change the intended meaning, but avoids the suggestion  
> that EPA tinkers capriciously with policy.

>  
> Main draft (first email)

>  
> Page 2, line 10: The word 'comprehensive' should be perhaps  
> replaced by 'reasonable'.

>  
> Bullet 1: It is not merely a discussion the weaknesses of other  
> approaches that is missing from the white paper. This panel report  
> seems also to avoid mentioning the phrase 'weaknesses of EE' or  
> 'disadvantages of EE' or 'drawbacks of EE'. This just seems odd.

>  
> Page 3, bullet 2: This brief sentence does not mention the needed  
> balance.

>  
> Line 25: The point of EE is not merely 'characterizing their  
> knowledge' but also aggregating and integrating it into a coherent  
> expression.

>  
> Page 6, lines 24-27: Consider appending the sentence 'The only way  
> to escape this combinatorial explosion without selecting an  
> aggregation method is to employ bounding and enveloping  
> strategies.'

>  
> Line 39: The panel report does not give the references for Cooke  
> (1991) or Jouini and Clemen (1996), so it should mention they are  
> cited in the white paper. It is embarrassing to use the adjective  
> 'recent' to refer to a paper that is thirteen years old.

>  
> Page 7, lines 18-21: I am disappointed with the synopsis of the  
> panel's discussion about the importance of correlations and  
> dependencies. I kind of think their short shrift in the white  
> paper is a serious problem. Yes, it's hard to elicit them; that's  
> why the guidance needs to address the issue.

>  
> Page 8, line 5: The sentence is not grammatical.

>  
> Line 45: The word 'among' should be 'in' or something.

>  
> Page 9, lines 6-8: But if it results in possible distributions  
> rather than impossible ones, then the interference is certainly  
> warranted and essential. Ideally, it is the only interference  
> there should be.

>  
> Line 14: The biases are listed, along with basic references at  
> Wikipedia ([http://en.wikipedia.org/wiki/List\\_of\\_cognitive\\_biases](http://en.wikipedia.org/wiki/List_of_cognitive_biases),

> and [http://en.wikipedia.org/wiki/Cognitive\\_bias](http://en.wikipedia.org/wiki/Cognitive_bias)). Tucker et al.  
> (2008) reviews the impact of some of them on risk perception.  
>  
>

**Comments from Dr. Christopher Frey**

Comments on Draft Report and Draft Transmittal Letter  
SAB Expert Elicitation Advisory Panel  
H. Christopher Frey  
April 20, 2009

**Comments on Draft Transmittal Letter**

The letter is more than one page and perhaps would have more impact if it were shorter.

Suggested deletions are:

- Page 1, sentence on lines 13-16
- Page 1, sentences on lines 21-26
- Page 1, delete sentence on line 30-31, start 2<sup>nd</sup> sentence with "The panel recommends that..."
- Page 1, lines 42-44. Is this a major recommendation? It is in the report and no doubt will be addressed, but does not seem to be at the level of significant that the Administrator needs to see.
- Page 2, lines 4-7. Will the "Hierarchy of Methods" document contain the comparison that is implied here? If not, then citing it will not further elucidate. This seems like a minor recommendation that does not need to be conveyed to the Administrator.

**Comments on the Draft Report**

Overall, the content of the draft report seems appropriate except where noted below. Specific comments:

- Charge Question A, page 3, line 39-40. The bigger issue to ensuring quality is the choice of the experts to avoid selecting persons with motivational biases. I am somewhat skeptical about performance assessment of experts on test questions about quantities for which they are not experts. While performance assessment provides information, is it of such unquestioned value that it should be prescriptively required?
- Under Charge Question A, request that the panel give consideration to adding another recommendation, which is for EPA to include a glossary of key terms with their practical definitions. A suggested list of such terms is attached. For many of these terms, the authors could simply adopt definitions from other sources, such as the World Health Organisation or prior EPA reports. The challenge for the reader is that many of these terms are not defined in the document, and some are used inconsistently.
- Charge Question B, page 5, line 45. Replace "certain" with "particular" to avoid confusion regarding the use of the term "certain" as possibly opposed to being "uncertain."
- Bottom of page 4 and top of page 5. The draft report seems to take a very different tone than I recall from the first day of the panel discussion (I was not able to attend on the second day) or that I support. There is lengthy text to the effect that there are challenges to documenting the "basis of judgment." However, my recollection of discussion by the panel is that there was concern expressed by several (myself included) that the document did not place adequate emphasis on the need to explain the basis of the judgments. In contrast, the draft panel report

seems to absolve EPA of making any such recommendation – see especially page 5, line 12. I simply do not agree with this. It is hardly resource intensive to take some notes during the elicitation upon which to write a paragraph or so explanation of the rationale provided by the expert for the lower bound, upper bound, and central tendency of a judgment. It is difficult to imagine any rationale for the panel not strongly recommending to EPA that this should be endorsed in its report. While there may be limitations, it is also the case that there are limitations with nearly any method or data set, so making practical use of methods that have limitations is not unique.

- There is a statement on the top of Page 5 to the effect that Delphi methods are much easier to describe with “full transparency.” Perhaps the *process* of the Delphi method is easier to document. However, the outcomes and their basis may be no clearer than from other methods. In particular, a key part of documenting EE is to document the structuring and conditioning steps, so it is clear as to the groundrules upon which the judgment is based. It is not obvious that the Delphi method fully ensures this. If it does not, then it cannot be “fully” transparent.
- My recommendation is to delete the text from page 4, line 37 to page 5, line 13. A focus of our teleconference should be whether, and if so how, to replace this text.
- Charge Question C.1, page 5, line 41, for clarity, replace “it” with “the pool.”
- Page 6, lines 1-3. The idea that a contractor is somehow more objective than EPA seems questionable. This *could* be true, but it is not generally true. Whether this is true or not depends on the contractor and how the contract is written. If the contractor is answerable to EPA, it is not clear how motivational biases on the part of the contractor are avoided without a bit more explanation.
- Charge Question C.3, page 7, lines 20-21. Insert “independent” before marginal distributions just to emphasize the point.
- The panel discussed and appeared to agree that correlations should not be directly elicited. A statement to this effect should be added after line 21 (p. 7).
- Page 7, line 28. Insert “of aerosols” after “(ddv). Spell out “EU-USNRC” before providing the abbreviations. Cite a reference for the study.
- Page 7, line 23 to page 8 line 5 seems like a lot of text and detail. Recommend that this be shortened. EPA will also receive individual comments that can be more detailed.
- Page 8, line 8. Figure 6.1 was a bad example of an influence diagram. Text here should be revised to say. “Problem structure and consistent conditionalization may be facilitated by use of an influence diagram. The influence diagram illustrated in Figure 6.1 of the white paper should be replaced with an improved example that is adequately labeled so that it can be understood without reference to additional text.”
- Charge Question D, page 9. The text in the paragraph on lines 10-16 and lines 29 to 34 are repetitive. Suggest consolidating in one place and cross-referencing.

### **Terms to add to the glossary and to use consistently throughout the document**

Accurate

Aggregation

Assumption  
Assumptions  
Availability  
Averaging  
Bias  
Conditional Probability  
Data gap  
Data quality  
Decision options  
Dependence  
Domain expert  
Elicitation  
Elicitor  
Encoding  
Estimates  
Event  
Extrapolation  
Heuristics  
Input  
Model  
Model choice  
Objective  
Overconfidence  
Paradigm  
Parameter  
Precision  
Quality  
Quantity  
Relationship  
Representativeness  
Robust  
Subjective  
Subjective Judgment (?)  
Subjective Probability  
Weighting

**Comments from Dr. Alan Krupnick**

Hi, Angela. I have been away on a project out of the country. Here is a para I think would be important to add to the transmittal letter. I don't know how others would feel about this, but I think it is the central issue with EE.

In particular the Panel believes it is important to set clear boundaries for when EE is more appropriate than other methods and when it is not. The Panel also recognizes that these issues are not settled in the literature or among experts in the field. Yet, EPA is encouraged to set such limits given the needs and resources of the Agency and the information contained in its own EE report and the SAB report.

### Comments from Dr. Mitchell Small

I think the draft report looks great. I have two small suggestions... no comments on the letter, it looks fine.

Suggestions for SAB Expert Elicitation Panel Report: 04/10/09 Draft

M. Small  
April 13, 2009

1. I think more can be said regarding the use of EE for ongoing evaluation of research options, VOI, etc. This point is raised in the bottom paragraph of page 2, so I would expand in a new paragraph that follows after that one (page 2, line 45+):

To integrate EE studies into ongoing scientific learning, research planning, study implementation, and the interpretation of results, expert elicitations should address uncertainty in both the current state of knowledge (including model parameters and relationships) and the outcomes of studies proposed to reduce these uncertainties. For example, experts could be queried for their probability distributions of relationships given alternative outcomes of a study. This method of elicitation has been employed by statisticians in the elicitation of “predictive distributions” (Kadane and Wolfson, 1998), which combined with the expert’s priors, allow derivation of their likelihood function for the experimental outcomes. Alternatively, direct elicitation of the likelihood function for a proposed experiment can be made, e.g., asking experts to estimate the sensitivity and selectivity of a proposed bioassay study (Small, 2008). With this assessment, the EE results can be used as part of value-of-information (VOI) studies to prioritize research, and subsequently updated in an adaptive manner as new research results are obtained.

Kadane, J.B. and L.J. Wolfson. 1998. Experiences in elicitation (with discussion). *The Statistician*, 47: 1-20.

Small, M.J. 2008. Methods for assessing uncertainty in fundamental assumptions and associated models for cancer risk assessment. *Risk Analysis*, 28(5): 1289-1307.

2. If think the requirement for elicitation of “directly measureables” on page 3, paragraph beginning on line 12, is a bit too strong. Sometime experts are so used to working with model parameters (e.g., having seen many estimates in the literature), that they may be most comfortable with these, even though they may not be directly measureable. So long as a (or the expert’s) model is available to relate the experts’ uncertainty in parameter values to uncertainty in modeled outcomes that are measureable (e.g., ambient air pollution concentrations), we can show the expert the implications of their parameter choices for the measurable outcome, compare their result to observed data if and when they become available, etc. I thus suggest adding a new sentence or two in the middle of line 18:

In some cases experts may be most familiar with model parameter values, especially when these have been derived and reported by multiple researchers in the literature. In this case elicitation of the parameter value may be appropriate (even if it is not directly measureable), so long as a model or models can then be used to illustrate the implications of the expert’s parameter choices for the measureable output of interest.

**Comments from Dr. Thomas Wallsten**

Wallsten comments on the panel report

Both the report and the cover letter are very well done. My only comment of note concerns point 5 on page 3 (lines 39-42) of the report, which now reads:

“The white paper should address methods for evaluating and ensuring the quality of expert judgments, including tests for coherence of judgments over multiple factors and performance (tested using judgments for seed quantities, the values of which will become known).”

Two separable issues are being raised here. One, focusing on coherence, concerns the degree to which the judgments conform to the probability axioms. The other, testing judgments for seed quantities, concerns accuracy in a calibration sense. They should be separated. Suggested text for point 5 is:

The white paper should address methods for evaluating and ensuring the quality of expert judgments, including tests for coherence and consistency of judgments over multiple factors and performance. In addition, some panel members recommend accuracy be tested by obtaining judgments for seed quantities, the values of which will become known.”

I qualified the second point (“some panel members recommend...”) because there is not agreement on that point. I’m not aware of evidence suggesting that good calibration in one domain implies good calibration in all others.

**Comments from Dr. Katherine Walker**

KWalker

Comments on SAB Expert Elicitation Advisory Panel 04/06/09 Draft

I thought the summary did a generally nice job of capturing a far ranging and complex conversation at the panel meeting. Some thoughts on clarifications follow.

P2 line 34-5. Missing text.

In characterizing the use of EE and other methods, attention should be given to the extent to which EE is a complement to rather than a potential substitute for the ideal method? or source of data?.

P3 third paragraph beginning with 'Recognition...

I think the second sentence is not quite accurate, and not entirely logically connected to the sentence that follows. EPA is interested in the probability of some outcome, but not just any probability. We would like that probability, whether estimated using a research study, some other model, or an expert's judgment, to be is 'correct' -- or in EE terms, at least well-calibrated and, ideally, informative. Thus, we are interested in what the 'correct' model is for predicting an outcome -- it matters whether a pollutant's effects demonstrate a threshold or not, to use the example in the paragraph. We do sometimes ask experts to weight or place probabilities on different types of models (see Evans et al. 1994 chloroform study) according to how well we think they represent or can predict reality.

In order for this sentence to connect more logically to the next sentence, perhaps what we want to say here is:

EPA is generally interested in the probabilities of specific (etc...) outcomes. While the Agency recognizes the importance of understanding the particular scientific mechanism or model (e.g. linear no-threshold dose-response function...) for explaining those outcomes, it recognizes that they may never be known exactly or at least not within the time frame for making a decision. Hence, the objective when using EE should be to elicit...

Line 18-19. Drop the last sentence. Or at least drop the "can" and insert "could" and continue after truth "if the right experiment were designed." Or something to that effect.

P 5. Top paragraph.

First, I don't agree with the blanket generalization that a remotely conducted Delphi is easier to describe with full transparency. Having sat with an expert who also had in hand a request to participate in a Delphi exercise for which he really didn't understand the questions, I think there are all kinds of scenarios in which Delphi could be less transparent. Who knows what the experts were thinking or how they came to their judgments? Second, I don't think we should be comparing favorably or unfavorably any other method since it wasn't our charge to assess them.

I do think that EPA can, in its guidance, encourage people to be creative as well as systematic in how they might capture some of the elements that they think really need to be recorded. If one is interested in how the judgments evolved over the course of the interview, one could imagine ways to improve recordkeeping to do that. Granger Morgan and colleagues have developed protocols that have used creative ways to capture judgments and inputs to judgments. The key thing is to set priorities for what is critical to understand about an expert's 'basis for judgment' and then figure out how to capture and communicate it.

Charge Question C.1 Selection of experts –

P5 line 26-27. In the same spirit of evaluating or agreeing upon the methodology used in the elicitation, not the outcome, the general criteria and process for selecting experts, not (ideally) the set of experts should be what is subject to scrutiny by stakeholders. The criteria and process should be the standard against which the set of experts is judged to avoid individual interference by stakeholders.

Line 29 Second sentence. Drop end of sentence reading “that are well-calibrated”. We can't a priori know for selection purposes whether experts are well calibrated. Similarly, drop the last sentence in the paragraph “To enhance transparency.... it may be helpful” up to “capable” and combine with previous sentence:

“To enhance the quality of the results, experts should have the ability to characterize their beliefs in terms of probability distributions and be capable of articulating the basis for their judgments.”

**Deleted:** that are well-calibrated. To enhance the transparency and credibility of the study, it may be helpful if the experts are

In general, I think we if we're talking about general criteria for selecting experts we should be more comprehensive, not only pick out certain attributes. For example, as adapted from Graham, Wolfe and Roberts 1988:

**Deleted:** of

- Relevant expertise
- Balance (disciplinary, institutional, political)
- Willingness to participate

Or we should just specify that EPA's guidance should state that the EE should lay out clearly its criteria for selection. If there are some criteria that, from the panel's and EPA's perspective must be included in any EE, they should be stated clearly and distinctly from a broad range of criteria that might be more project specific.

The point of the third paragraph beginning on line 40 is unclear. Again our advice to EPA here should be that they should establish a transparent process for meeting the objectives or criteria they set for the panel – it's not necessarily important to have two pools or one pool, or “ invit[ing] stakeholders to nominate experts”. I think if we start singling out approaches like this we give them too much emphasis. It is the job of the guidance ultimately to suggest many different alternatives or allow for many different approaches to meeting the objectives for the panel. In particular, in light of the last paragraph's support for lack of sponsor control over

expert selection, I'm not sure I'd even agree that inviting stakeholder nominations is any more appropriate.

#### Charge Question C.2 Multi –Expert aggregation

Did we not also discuss and agree that the decision whether or not to combine judgments should be explicitly discussed and planned for in the design of the EE both to make sure that the correct data are collected in the elicitation (e.g. performance weights, peer or self weights) and to make sure that combination is not done “after the fact” to manipulate the outcomes. I thought we did, and if so, should note it here.

#### Regarding references

I've attached a list I prepared for a recent lecture which I think has some references that were missing from the original EPA white paper. However it wasn't intended to be comprehensive.

I think that like any review, the white paper should specify its search strategy and criteria for inclusion. This would be a more transparent and even-handed approach.

#### **Basic Background**

Cooke, Roger M. 1991 Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press.

Morgan and Henrion 1990 Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge University Press, New York

O'Hagan, A, Buck, C, Daneshkhah, A, Eiser, JR, Garthwaite, PH, Jenkinson, DJ, Oakley, JE, Rakow, T 2006. Uncertain Judgements: Eliciting Experts' Probabilities. John Wiley & Sons Ltd. Chichester, England.

**Winkler, R.L. and RT Clemen. Multiple Experts vs. Multiple Methods: Combining Correlation Assessments Decision Analysis Vol 1(3) September 2004 pp 167-176.**

#### **Applications of Expert Judgment Elicitation**

##### Nuclear Reactor Safety Studies

Bonano et al, 1989. Elicitation and use of expert judgment in performance assessment for high-level radioactive waste repositories. NUREG/CR-5411, U.S. Nuclear Regulatory Commission, Washington, DC

Camp et al., 1990;  
Merkhofer and Ruchal, 1989

#### Expert Judgments Sulfur Emission and Oxidation Rates

Morgan et al., 1984 Technical Uncertainty in Quantitative Policy Analysis: A Sulfur Air Pollution Example. *Risk Analysis* 4:201-216.

#### Lead (Pb) NAAQS Risk Assessment

Whitfield and Wallsten 1989. A risk assessment for selected lead-induced health effects: an example of a general methodology. *Risk Analysis* Jun;9(2):197-207.

#### Ozone Chronic Lung Injury Assessment

Whitfield, R.G., T.S. Wallsten, R. L. Winkler, H.M. Richmond, and S.R. Hayes. 1991. *Assessing the Risks of Chronic Lung Injury Attributable to Long-Term Ozone Exposure*. Argonne National Laboratory Report ANL/EAIS-2. NTIS/DE91016814. Argonne, IL. July.

#### Cancer Risk

Evans et al. 1994 Use of probabilistic expert judgment in uncertainty analysis of carcinogenic potency. *Regul Toxicol Pharmacol.* 20:15-36.

Evans J.S., Graham J.D., Gray G.M. and Sielken R.L. 1994a "A distributional approach to characterizing low-dose cancer risk." *Risk Analysis* 1994: 14(1): 25-34.

#### Climate Change Impacts on Forest Ecosystems

Morgan, M.G.; Pitelka, L.F.; Shevliakova, E. Elicitation of expert judgments of climate change impacts on forest ecosystems. *Climatic Change* **2001**, 49, 279-307.

#### Benzene Concentrations in Ambient, Indoor, Personal Air: A calibration study conducted as part of NHEXAS

Walker K.D.; Evans, J.S.; MacIntosh, D. **2001**, Use of expert judgment in exposure assessment. Part I. Characterization of personal exposure to benzene. *J Expo Anal Environ Epidemiol.* 11, 308-322.

Walker, K.D.; Catalano, P.; Hammitt, J.K.; Evans, J.S. **2003** Use of expert judgment in exposure assessment: Part 2. Calibration of expert judgments about personal exposures to benzene. *J Expo Anal Environ Epidemiol.*, 13, 1-16.

#### Estimation of the PM<sub>2.5</sub>-Mortality Concentration-Response Relationship

Industrial Economics, Inc. **2004**. An expert judgment assessment of the concentration-response relationship between PM<sub>2.5</sub> exposure and mortality. U.S. Environmental Protection Agency, Research Triangle Park, NC.

Roman, H. A., et al., 2008. Expert Judgment Assessment of the Mortality Impact of Changes in Ambient Fine Particulate Matter in the U.S. *Environmental Science and Technology.* 42(7): 2268 – 2274.

Cooke, RM, Wilson, A, Tuomisto, J, Morales, O, Tainio, M, Evans JS. , 2007 A Probabilistic Characterization of the Relationship between Fine Particulate Matter and Mortality: Elicitation of European Experts. *Environmental Science and Technology*; 41: 6598-6605.

Tuomisto, J, Wilson, A, Cooke, RM, Tainio, M Evans, J. , 2008; Uncertainty in Mortality Response to Airborne Fine Particulate Matter: Combining European Air Pollution Experts. *Reliability Engineering and System Safety* 93(5): 732-744.

Cooke, RM, Goosens LHJ. TU Delft expert judgment database. *Reliability Engineering and System Safety*, 2008; 93(5): 657-674.

Clemen, RT, (2008). A Comment on Cooke's Classical Method. *Reliability Engineering and System Safety*, 2008; 93 (5): 760-765.

### **Measuring Performance**

Christensen-Szalanski and Bushyhead, 1981. Physicists' Use of Probabilistic Information in a Real Clinical Setting. *Journal of Experimental Psychology: Human Perception and Performance* 7(4).

Cooke, RM 1991 (cited above)

Lin and Bier, 2008. *Reliability Engineering and System Safety*,

Lichtenstein, Fischhoff, and Phillips, 1982. Calibration of probabilities: The state of the art to 1980. In: Kahneman, Slovic, and Tversky, 1982.

Matheson and Winkler, 1974 Scoring Rules for Continuous Probability Distributions. *Management Science*, 22.

Murphy and Winkler, 1977 Reliability of Subjective Probability Forecasts of Precipitation and Temperature, *Applied Statistics*, 26.

### **Heuristics and Biases**

Ariely, D. 2008. Predictably Irrational: The Hidden Forces that Shape our Decisions Harper Collins Publishers, NY

Kahneman, D, Tversky, A. 1974. Judgment under uncertainty: Heuristics and biases. *Science*; 185: 1124-1131.

Kahneman, Slovic and Tversky eds. 1982. Judgment Under Uncertainty: Heuristics and Biases, Cambridge University Press, New York.

Gilovich T, Griffin D, Kahneman D., eds. 2002. The Psychology of Intuitive Judgment. Cambridge University Press. Cambridge UK

### **Choosing experts/numbers of Experts?**

Graham, Hawkins, Roberts, 1988. Expert Scientific Judgement in Quantitative Risk Assessment. Publication missing.

Hora, S. 2004. Probability Judgments for Continuous Quantities: Linear Combinations and Calibration. *Management Science* 50(5).

### **Methods**

Staël Von Holstein and Matheson, 1979. [A Manual for Encoding Probability Distributions](#), SRI International, Palo Alto, California

### **PM2.5 Elicitation – Additional citations**

National Research Council. Estimating the Public Health Benefits of Proposed Air Pollution Regulations. Washington, DC: National Academy Press, 2002. Available at: <http://www.nap.edu/catalog/10511.html>. Accessed on October 6, 2008.

U.S. Environmental Protection Agency. Regulatory Impact Analysis Associated with the Clean Air Interstate Rule (CAIR). Washington, DC: U.S. Environmental Protection Agency, EPA-452/R-05-002, March 2005.

### **Combination of Expert Judgments**

Ariely, D et al. 2000 The effects of averaging subjective probability estimates between and within judges. *J. Experiment. Psych.: Appl.* 6:130-147

Clemen, R and Winkler, 1999. Combining probability distributions from experts in risk analysis. *Risk Analysis* 19:187-203.

Genest and Zidek, 1986. Combining Probability Distributions: A critique and annotated bibliography. *Statist. Sci.* 1: 114-148.

### **Some Software you might explore...but let your problem be your guide.**

#### **Elicitation software: *Constructor***

e.g., <http://www.ramas.com/constructor.htm> (freeware, beta test)

#### **Calibration Software: *Excalibur* (Cooke/ Univ of Delft)**

“EXCALIBUR (acronym for EXpert CALIBRation) is a Windows program that allows parametric and quantile input from experts for continuous uncertain quantities and combines these according to the methods described in [R. M. Cooke "Experts in Uncertainty"](#), Oxford University Press 1991. In particular user-weights, equal weights and performance based weights are supported. Robustness analysis shows how sensitive the results are to choice of expert and choice of calibration variables. Discrepancy analysis shows how the experts differ from a Decision Maker. Output is compatible with modern text processors and spreadsheets.”

[http://dutiosc.twi.tudelft.nl/~risk/index.php?view=article&catid=3%3Asoftware-info&id=6%3Aexcalibur&option=com\\_content&Itemid=5](http://dutiosc.twi.tudelft.nl/~risk/index.php?view=article&catid=3%3Asoftware-info&id=6%3Aexcalibur&option=com_content&Itemid=5)

### **Decision modeling: *Analytica* (Lumina systems)**

“Analytica is a visual tool for creating, analyzing, and communicating decision models.

- Its intuitive [influence diagrams](#) let you create a model the way you think, and communicate clearly with colleagues and clients.
- Its Intelligent Arrays™ let you create and manage multidimensional tables with an ease and reliability unknown in spreadsheets.

Its efficient Monte Carlo lets you quickly evaluate risk and uncertainty, and find out what variables really matter and why.”

<http://www.lumina.com/>