Compilation of Member Comments on the EEAP Draft Advisory
On EPA's Expert Elicitation White Paper
SAB Teleconference of August 28, 2009

A. Lead Reviewers:

1. Dr. Granger Morgan:

I have read both the Task Force White Paper and the draft SAB review. The first half of this review comments on the SAB sub-committee's review. However, because this is a topic on which I have considerable direct experience, in the second half I have also added a few comments on the draft Task Force White Paper.

Comments on the SAB review

Both in the letter and in the body of the report the SAB call for a "critical analysis of the strengths and weaknesses of EE in comparison with other approaches." Having read the White
Paper before reading the SAB comments, I was surprised by this. I thought the White Paper did a pretty good job of discussing the pros and cons of expert elicitation. If the sub-committee wants to make this recommendation, it would certainly be helpful to the EPA author team if they would list some of the strengths and weaknesses they have in mind that they think the Task Force did not discuss.

A second major theme of the review is the importance of comparing expert elicitation to "other methods for quantifying uncertainty." Given that the report contains some comparison in
Section 4.3 (pp. 51-52), again it would be helpful to be explicit about the other methods the subcommittee has in mind. On those two pages, there is discussion of five topics, none of which strike me as substitutes for expert elicitation. Perhaps what the sub-committee is asking for is simply to make this more explicit. In this regard, here are a few thoughts:

- Statistical/frequentist approaches: In my view, if the data exist to adopt a classical approach to fully characterizing the uncertainty about some coefficient value, there should be no need to employ expert elicitation. Thus, this is not an "alternative approach."

- Judgment/subjectivist: In my view, this is the same thing as what one obtains in a formal way via expert elicitation. The topics listed are more in the vein of analytical methods to be used when one has obtained subjective probability density functions of coefficients of interest, or probabilities to attach to the potential veracity of alternative model functional forms.

- Scenario analysis: While widely used, scenario analysis is typically invoked as a way to stretch the thinking of decision makers, not to characterize uncertainties about the value of a specific coefficient value such as the slope of a damage function. One could view the work of Evans *et al.* (1994a,1994b)

as an exercise in eliciting probabilities that different experts attach to the veracity of different scenarios (i.e., models of what biological processes are important), but that hardly makes scenario analysis an alternative to expert elicitation. Further, as outlined in Morgan and Keith (2008), scenario analysis is subject to, and/or may lead to, some potentially very serious cognitive biases. Indeed, as that paper notes, some practitioners actually use it to persuade rather than to inform!

• Other methods: This section mentions explicitly only interval methods, fuzzy methods and meta-analysis. The first two are part of a larger set of alternatives (including also belief functions, certainty factors, second order probabilities, etc.) to conventional probability. There is a literature on these (Henrion, 1999; Smithson, 1988). I don't believe it would be productive for the EPA to go off and spend more time exploring alternatives to probability. The last item "meta analysis" is something that should be done whenever there are data, as an *input* to, not an alternative to, expert elicitation.

• Sensitivity analysis: This is clearly important. If running a coefficient value through its plausible range can flip the answer to a policy question, that says decision makers need to think hard before deciding what to do. But, again it is not an alternative. The same is true if the subjective probabilistic judgments of different experts lead to different policy conclusions (more on that below when I turn to the discussion on combining experts).

Bottom line – what other "alternatives" does the subcommittee want the Task Force to address, and what issues do they think should be discussed?

I am very much in agreement with the implicit suggestion in point 2 in the letter – which I read in part as a request to make explicit the fact that most of the cognitive biases, etc. that arise in expert elicitation also arise in conventional consensus panel (or peer review) processes, but these are masked because these processes are less transparent.

The main body of the review begins on pages 1 and 2 with a list of 11 items. I see no problems with items 2, 4, 8, and 10. I have the following observations about the others:

1. The SAB review says the draft White Paper "reads too much like an advocacy document" and should "adopt a more neutral, analytic tone." Can the subcommittee cite some examples? If I were the author, without some better guidance, I'd not know what the SAB wants me to change, or where you think I've strayed over the line into advocacy.

3. The issue of "…other methods of quantifying uncertainty…" is discussed above. As to expert elicitation not being a means to generating new primary

data, the report says this quite explicitly in several places, including section 3.3.3 titled "Do expert elicitation results represent new data or knowledge."

5. When the subcommittee subsequently elaborates on this they might explicitly suggest that while the protocols used, the experts employed, etc. should all be open and public, in many cases, it will be desirable to *not* link specific answers to specific experts.

6. Aggregation – I have a fundamentally different view than the SAB sub-committee and will discuss the issues in a few paragraphs below.

7. I don't understand the recommendation that "model-dependent parameters should only be elicited when they can be unambiguously translated or inferred from measurable quantities." Suppose for example I want to explore how much difference it makes if I use several alternative plausible damaged functions for a carcinogen (Morris, 1990). Is the subcommittee saying I should not elicit probability distributions on the coefficients that define the shape of those alternative functions from different experts who believe that these different functional forms might obtain in the real world? If that is what the sub-committee is saying, then it needs to justify its position. If that is not what it is saying, it needs to clarify its position.

9. "Fully review the literature on cognitive biases." The basic review that is provided is not bad. Rather than replicate what a number of others have done, why not simply ask the Task Force to add a bibliography of additional literature that potential practitioners should read?

11. I have discussed this above in the context of the letter. As I say there, meta-analysis should not be viewed as an alternative but rather as an input to expert elicitation. As noted above, and at least implied in the White Paper, unstructured expert committees and peer review are subject to the same biases as expert elicitation but are typically not made explicit. Note too that the White Paper argues that expert consensus committees may be more subject to overconfidence (distributions that are too narrow) than the results of a set of separate expert elicitations (we have empirical evidence that supports this assertion in the context of IPCC consensus summaries – see note below).

Turning now to the body of the report, note that after item 1 on the bottom of page 2 the next numbered item is numbered 3.

On page 3, I strongly disagree with the assertion:

> "…EPA is generally interested in the probabilities of specific…outcomes, not in whether a particular scientific model (e.g., linear no-threshold dose-response function) is 'correct.'"

Perhaps in the past EPA has found it expedient to ignore the possibility that there is uncertainty about the functional form of damage functions, but that is not scientifically justifiable. This sentence must simply be removed.

I agree with the stress in item 4 in the list on page 1 (and in the discussion in the middle of page 4 and in line 32 on page 9) on the importance of checking for internal consistency and coherence in experts' responses. Indeed, the sub-committee might suggest that in designing elicitation protocols it is important to try to build in opportunities for such checks.

On pages 6 and 7 with respect to the assertions that "some form of aggregation is usually required…[and that]…the white paper devotes inadequate attention to methods of aggregating experts' judgments." I have the following comments:

- Asking a set of experts for their judgment about the value of an uncertain quantity is *not* like polling a bunch of voters about how they will vote. Science is not a democratic process in which we vote on what value we should choose for π or Plank's constant. One of the set of experts interviewed, perhaps even one of the outliers, may be right, and all the others may be wrong. Thus, it seems to me that a policy maker should first ask, "given the range of opinions in the scientific community, does the policy I choose depend on which expert I listen to?" If the answer is no, that's great. If the answer is yes, clearly one needs to explore whether there is any other way to gain insight, before making a decision.
- Often elicited distributions are not the final answer but rather are an input to some more complex model and decision processes. Suppose, for example, that I have elicited PDFs for the value of a chemical rate constant or loss rate that will then get plugged into a nonlinear model that will be used to assess environmental or health impacts. If I really want to combine, because the model is not linear I should run the model separately for each expert and combine the results rather than combine all the experts and then run the model.
- There is implicit reference on page 7 to what is explicit in the attachment to the white paper – namely that by "training" experts one can both infer how to correctly weight them and that such training may allow one to dramatically reduce cognitive bias. There has not been much empirical investigation of these claims, and what I know about leaves me pretty un-persuaded – but the SAB might wish to ask Baruch Fischhoff for his view on this.

I recognize that there is a considerable literature on combining experts (some of which I consider misguided) but the key point is that rather than sending the Task Force off to plow through the intricacies of all the various methods that have been proposed, I'd much prefer to see the subcommittee suggest they spend more time thinking about when it is and is not appropriate to be combining.

On page 9, during my tenure on the SAB I have repeatedly reminded my economist colleagues that the value of "willingness to pay" is not a "scientific question." Yes, there may be a measurement issue when one tries to run a contingent valuation study. But, it is not the case that the Administrator is bound to use that value in making any decision. Choosing the investment rate to prevent mortality or morbidity (or to avoid ecological damage) is a normative social question, not a scientific question. Consistent with any specific guidance from the Congress, the Administrator is empowered to make that social judgment on behalf of U.S. society. I agree strongly with the sub-committee's warning on page 10 that high quality expert elicitation cannot be reduced to a "cookbook" formula, and that attempts to do so may seriously impede both its adequate use and future refinement.

Finally, if the sub-committee is going to suggest additional references, they might also include CCSP Product 5.2, "Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Climate Decision Making" available at; www.climatescience.gov/Library/sap/sap5-2/final-report/default.htm.

Comments on the White Paper

Page 20. The CCSP has not done formal expert elicitation of the sort discussed in the White Paper. Indeed, at the request of a lead author in the chapter on aerosol forcing in the 4th assessment, we performed an expert elicitation on that topic, and got it peer reviewed and published before the deadline (Morgan et al., 2006), only to have the Chair of WG I rule out its use on the grounds that expert elicitation is not science. There was discussion earlier this summer of making use of such methods in the 5th assessment. If this happens, it will more likely be in WG II or WG III. Incidentally, the range of uncertainties expressed by the 21 experts in our study was significantly wider than that of the consensus summary reported by WG I. We see the same result in a more recent study now in review that looked at value of climate sensitivity.

Page 26, Section 3.3.4 and page 30. See my comments above on this topic. Some elaboration might be good.

Pages 28, 74, and 128. See my comments above on combining experts.

Page 29. The most carefully conducted group process I know is that described by Budnitz et al.
(1995).

Section 3.5.2 there is rather too much dependence on just Schackle. Consider expanding the set of references (e.g., see Henrion, 1999).

Page 47. Future oil price is a poor example since it is unclear there is such a thing as an expert on future energy use or prices. See the two historical reconstructions in Morgan and Keith (2008).

Page 59. The analytical hierarchy process is not axiomatically derived and can lead to inconsistent outcomes in some circumstances (Dyer, 1990).

Pages 66 and 74. It is essential that one pre-test any elicitation protocol. In work we have done, we have typically used Post-docs working in the same field as the experts.

Page 68. I don't think "generalist" is appropriate. One really needs a domain expert who knows the literature very well and can push the expert with counter examples, etc.

Page 76. Whenever we've done elicitations with scientists (i.e., numerate respondents) they have had no interest in using a probability wheel.

Page 79. We have done one expert elicitation by mail and learned that there are a real number of pitfalls. For example, many experts set the 5 and 95 percentile points on the box plots they constructed much too close to the central box. Before I'd be comfortable supporting the general use of mail-based elicitation, I'd want to see more experimental study of how to avoid obvious pit-falls. My colleague David Keith at U. Calgary has developed an "on line" elicitor that may overcome some of these problems. It is being tested now.

References cited:

Budnitz, R.J., G. Apostolakis, D.M. Boore, L.S. Cluff, K.J. Coppersmith, C.A. Cornell, and P.A.
Morris, 1995: *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on uncertainty and the use of experts*, Lawrence Livermore National Laboratory, Livermore CA, UCRL-ID 122160, 170pp.

Dyer, J. S., 1990: "Remarks on the Analytic Hierarchy Process," *Management Science*, *36*, 249-
258.

Evans, J.S., G.M. Gray, R.L. Sielken, Jr., A.E. Smith, C. Valdez-Flores, and J.D. Graham,
1994a: "Using of probabilistic expert judgment in uncertainty analysis of carcinogenic potency," *Regulatory Toxicology and Pharmacology*, *20*, 15-36.

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

Henrion, M., 1999: "Uncertainty," In: *MIT Encyclopedia of the Cognitive Sciences* [Wilson, R.A. and F. Keil (eds.)], The MIT Press, Cambridge, MA.

Morgan, M. G. and D. Keith, 2008: "Improving the way we think about projecting future energy use and emissions of carbon dioxide," *Climatic Change*, *90*(3), 189-
215.

Morgan, M.G., P.J. Adams, and D. Keith, 2006: "Elicitation of expert judgments of aerosol forcing," *Climatic Change*, *75*, 195-214.

Morris, S.C., 1990: *Cancer Risk Assessment: A quantitative approach*, Marcel Dekker, New York, 408pp.

Smithson, M., 1988: *Ignorance and Uncertainty: Emerging paradigms*, Springer-Verlag, New York, 393pp.

2. **Dr. Catherine Kling:**

a) Does the draft report adequately address the original charge questions to the SAB Panel? Yes, this is a thorough and well conceived report. Very nicely done.

There is no executive summary which I mention in case this was an omission rather than intentional (such a short report hardly seems to need an executive summary).

b) Is the draft report clear and logical? Yes, a couple of minor suggestions in this regard.

i) Both in the letter (recommendation #1) and in the advisory, the panel recommends that the writers of the white paper adopt a more "neutral, analystic" tone. I suspect that this message would be clearer to the white paper authors if an example or two were provided.
ii) The seventh recommendation is that the "panel urges that the quantities being elicited be measurable..." How can a quantity be non-measurable? Since I have not read the white paper, perhaps this is self-evident from the context, if not, perhaps an example or rewording would be useful.
iii) Extremely minor: there is no "2" in the list of issues identified under Charge question (there is a "1" and then a "3" on page 3).
iv) For additional clarity regarding the point made on page 9, lines 9-14, the panel might consider adding a sentence like "Thus, we urge the EPA to clearly articulate that economic values such as WTP are legitimate quantitites for EE to address."
v) On page 9, lines 31-43, the panel discusses elicitation protocols in the context of multivariate distributions. If there are references to point the white paper authors to concerning the points raised here, it would be useful to include them.

c) Are the conclusions drawn, and/or recommendations made, supported by information in the body of the draft SAB report? Yes, very clearly.

3. **Dr. John Balbus:**

Overall this review is thoughtful and constructive in tone. It succinctly addresses a reasonable number of problems with the guidance document and makes cogent

suggestions for improvement, including appendices with specific references and terms for a glossary.

I have a few relatively minor comments.

The letter provides 3 recommendations, the report itself highlights 11 recommendations. #2 and #3 are the same in both lists. The letter's #1 is #11 of 11 in the report. The first recommendation in the report has a somewhat chiding tone about the guidance document having an advocacy tone that is not further discussed in the report. I would recommend combining recommendations 1 and 11 as the first recommendation in the report, eliminating the chiding phrase, and then mentioning in the letter that there are a total of ten detailed recommendations, of which the top three are presented in the letter. If the committee is concerned with an advocacy bias beyond a simpler lack of critical analysis, this should be expanded upon in the report.

Recommendation #5 (tradeoff between transparency and resources) is not very well developed in the report. There is a mention of greater resources associated with enhanced elicitor interaction in the discussion of transparency, but there are probably other aspects of transparency that are resource intensive as well (e.g, stakeholder outreach and solicitation, etc.) This recommendation could use a little more discussion in the report.

5 line 16-17: Not clear why problem framing and definitions of quantities elicited are part of the results rather than the process?

Charge question C1a: Not sure the charge question of whether the different criteria and strategies for selecting experts is adequately addressed. The text discusses different criteria for different study purposes, but it is not clear whether the EPA guidance addresses this adequately or not. The text for this question also suggests that studies are improved by having the experts' beliefs well-calibrated but then notes that it is nearly impossible to calibrate judgments for the quantities being assessed. The SAB report should clarify whether calibrating on surrogate data is adequate to address this issue, and whether this is already covered in the EPA report or a fresh recommendation from the committee.

B. Other SAB Members:

4. <u>Dr. Rogene Henderson</u>

The letter is clear and concise, offering three rather general recommendations to improve the white paper.

Much of the actual review is written in a language (jargon?) unfamiliar to me. Because I am interested in EE and have participated in one EE, I went back to the original white paper to see if the jargon used in our SAB review was similar to that used in the white paper. I thought perhaps the language was appropriate and informative for the audience to whom it was addressed. Much to my pleasure, I found the white paper was quite transparent and easily followed. .I had no problem with the language in the white paper. I truly think the review of the white paper could be written with much less jargon and be made transparent and understandable to the authors, as well as to those outside the field..

Page 1, starting on line 23: Here we go from the three recommendations in the letter to 11 recommendations. Why the big difference? Can the 11 recommendations be condensed and clarified.? For example, under #2 (line 27), the last sentence (lines 30-31) is redundant and could be deleted. Rec. # 8 (page 2, line5) could be written more clearly and concisely. The whole list of recommendations needs to be edited for clarity.

Page 3, lines 8-13: I do not agree that there are no formal methods for making some of the extrapolations mentioned. For example, the EPA has standard methods for extrapolating from animal data to human estimates.

Page 7, line 8: I am curious what a "seed" variable is. This is an example of jargon that is not known to those outside the field. I note it is not even listed in Appendix B. There are many other examples, but I mention only one.

Page 8: I appreciated the use of an example here. Examples might be used to clarify other parts of the document.

5. <u>Dr. Jerald Schnoor</u>:

I have read the Draft EEAP report on expert elicitation in its entirety, and I am quite impressed with its tone, analysis, and content. It is a very well written report. The cover letter summarizes the report's contents very well, and the report is clean and concise. This SAB Committee is well qualified and distinguished in this area. I am pleased to have this report go to EPA under the SAB banner.

6. <u>Dr. Duncan Patten</u>:

I think the panel did an excellent job of reviewing and commenting on the EE Task Force White Paper. Some recommendations that reappear in several forms or are applied to different components of the EE White Paper are right on target. These relate mostly to comparing EE to other forms of expert opinion, how use of EE might

enhance EPA decision making, strengths and weaknesses of EE, etc. The comment that the authors of the EE White Paper appear to be promoting EE as the approach EPA should take is important as this type of White Paper should be a review and commentary (and as pointed out, a comparison of approaches) and not an advocacy paper.

I'm sure there are details that could be improved upon in the review but the panel has developed comments that, if followed, will greatly improve the original white paper. This is not an area I'm familiar with, so my review and comments are very general.

7. **Dr. Judith Meyer:**
The charge questions were addressed.
The draft report is clear and logical, although I have a couple suggestions for improvements indicated below.
The conclusions and recommendations are supported by the material in the report.

Suggested improvements:

1. The letter to the Administrator is very clear with straightforward recommendations that strike me as being right on target. It is also very short, so I don't hesitate to suggest adding one thing. Item 11 in the Introduction lists alternative approaches to EE, and I think the letter to the administrator would be improved if those alternative approaches were also listed in recommendation 1 of the letter. When I read the letter, I did not know what those alternatives were, and including them makes the recommendation sound less vague.

2. Why is there not an executive summary? The introduction contains elements of an executive summary, but someone wanting a brief summary of the report's recommendations would not know to look in an introduction for that. I would recommend taking executive summary material out of the introduction and having a separate section called Executive Summary.

3. Are there situations in which EE is particularly appropriate and ones in which it is not appropriate? Is that clearly addressed in the report? Should it be?

4. I commend the committee for providing the recent literature references in Appendix A, which helps the Agency be more responsive to the committee's recommendations. This appendix should also be mentioned on p. 10 lines 26-28, assuming that it contains literature relevant to the point being made here.

5. p. 6, line 7: The recommendation that the experts are acceptable to stakeholders implies that all stakeholders are acting rationally and without agendas (be they hidden or obvious). In many regulatory decisions that is not the case. Did the committee consider that when making that statement?

8. **Dr. Thomas Theis:**
I have read the expert elicitation report and the letter to the administrator. In my view it satisfies the three criteria (a, b, and c). I have no substantive comments.

9. **Dr. Meryl Karol:**
Thank you for the opportunity to review and comment on the EEAP report. In response to the questions:
a) The report adequately addresses the charge questions.
b) The report is logical, but clarity should be improved. Several words need explanation and many sentences appear unnecessarily complex. For example, what is meant by "conditionalization" (p.8)? The sentence, p. 9, lines 37-41, is overly complex and unclear. By comparison, the last paragraph on p.10 is crisp and clear.
c) Recommendations are supported in the text, but recommendation #2 in the cover letter is awkwardly worded and could be made clearer.

10. **Dr. Bernd Kahn:**
The SAB Advisory for the EPA EE White Paper is very well done carefully balanced. I have two comments on page 3:
a) There is a number 1 on p.2, l.39 and a number 3 on p.3, l.25, but I see no number 2 between them; should it be at the beginning of l.4 or l.16?

b) Concerning the paragraph on EE studies (lines 16 - 23), should the panel even suggest 'EE studies', a very secondary matter compared to the primary EPA studies that are important but woefully underfunded? A reasonable 'study' achieved without diverting funds would be comparing measured outcomes with the earlier EE findings as often as possible, but even those would only raise the question if one can distinguish between outcomes affected by good/bad Experts and good/bad EE design.

11. **Dr. Otto Doering:**
The white paper - It does address the charge, it is clear and logical, the conclusions made are supported by the information.

One editorial suggestion; on page 6, the second paragraph, the sentence in parentheses at the end of the paragraph does not seem to relate fully to the body of the text of the paragraph. The paragraph asks for characterization and then the following sentence in parentheses mentions the impossibility of calibration. Is there a little more that could be said to make these distinctions clearer to the reader.

12. **Dr. David Dzombak:**
a) Does the draft report adequately address the original charge questions to the SAB Panel?
Charge Question B. The report does not address directly the portion of the charge question that says " Please comment on whether the white paper presents adequate mechanisms for ensuring transparency when 1) considering the use of

EE (chapter 4), 2) selecting experts (chapter 5); and 3) and presenting and using EE results (chapter 6)". The response discusses opportunities for enhancing transparency, but does not address directly the question of the adequacy of mechanisms for ensuring transparency.

b) Is the draft report clear and logical?

The Introduction is written like an Abstract, with a brief overview of the issue and then presentation of a summary of the findings and recommendations. I suggest that the section be retitled "Abstract" or "Executive Summary". A conventional Introduction can be provided (introduction of the issue, background information, presentation of objectives), or it can be omitted and the reader can go straight to the responses to the specific charge questions.

c) Are the conclusions drawn, and/or recommendations made, supported by information in the body of the draft SAB report?

The findings and recommendations are adequately supported by information given in the body of the draft SAB report.

13. <u>Dr. Valerie Thomas</u>:

The Draft Expert Elicitation Review is clear, well-written, and informative.

a) Does the draft report adequately address the original charge questions to the SAB Panel? Yes.

b) Is the draft report clear and logical? Yes.

c) Are the conclusions drawn, and/or recommendations made, supported by information in the body of the draft SAB report? Yes.

14. <u>Dr. James Sanders</u>:

Please review the draft EEAP report and provide your written response to the charge questions for an SAB quality review:

a) Does the draft report adequately address the original charge questions to the SAB Panel? Yes. The report is brief and to the point and address each charge question in a coherent manner.

b) Is the draft report clear and logical? Yes. I do have two minor comments, which are below.

c) Are the conclusions drawn, and/or recommendations made, supported by information in the body of the draft SAB report. Yes, the panel has provided the responses requested, their answers are well justified and logical.

Two minor comments:

a) This is a very brief report, 10 pages. I was a little surprised by the summary recommendations on page 1 and 2, placed under Introduction (which they clearly are not). It would seem to make better sense to set them apart as a summary at the beginning of the report--however, with the short overall length, I guess that would appear to be too much? In any event, I recommend that some way be found to set these recommendations (which are all appropriate) away from the introductory paragraphs, or perhaps turn the Introduction section into a summary?

b) Charge question A, pp. 2-3. The panel's responses are numbered in this response, but number 2 is missing. It goes from 1., on p. 2 to 3., on page 3.

15.