The Honorable Lisa P. Jackson
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Review of EPA's Draft Expert Elicitation Task Force White Paper.

Dear Administrator Jackson:

EPA's Office of the Science Advisor requested that the Science Advisory Board (SAB) review a white paper on expert elicitation (EE) prepared by a task force of the Agency's Science Policy Council. EPA's draft white paper defined expert elicitation as "a formal process by which expert judgment is obtained to quantify or probabilistically encode uncertainty about some uncertain quantity, relationship, parameter, or event of decision relevance" (p. 5). In response to the Agency's request, an SAB panel conducted a peer review of the draft white paper. The enclosed advisory report provides the advice and recommendations of the panel.

The panel commends the task force for preparing a broad and thoughtful white paper on the potential use of expert elicitation at the Agency. The white paper was commissioned by EPA's Science Policy Council "to initiate a dialogue within the Agency about the conduct and use of EE and then to facilitate future development and appropriate use of EE methods" (p. 2). The panel judges that the white paper succeeds in providing much of the information needed for the proposed dialogue and to facilitate future development and appropriate use of EE. The white paper provides a good introduction to EE for readers who may be unfamiliar with it and careful discussion of many of the issues that must be faced if the Agency is to use EE in the future.

The panel recommends that the white paper:

1. Provide a critical analysis of the strengths and weaknesses of EE in comparison with those of other approaches.
2. Distinguish issues particular to EE from issues that arise in any analysis of environmental intervention or attempts to incorporate expert judgment.
3. Give greater attention to the extent to which EE is a complement to rather than a substitute for other methods of quantifying uncertainty.

Finally, the panel encourages EPA to continue to explore the use of EE, to support

research on the performance of EE and alternative approaches, and to conduct additional EE studies to gain experience and understanding of the advantages and disadvantages of EE and other methods in diverse applications.

Thank you for the opportunity to provide advice on this important and timely topic. The SAB looks forward to receiving your response to this advisory.

Sincerely yours,

Dr. Deborah L. Swackhamer, Chair Science Advisory Board	Dr. James K. Hammitt, Chair Science Advisory Board Expert Elicitation Advisory Panel

U.S. Environmental Protection Agency
Science Advisory Board
Expert Elicitation Advisory Panel

CHAIR

Dr. James K. Hammitt, Professor, Center for Risk Analysis, Harvard University, Boston, MA

MEMBERS

Dr. William Louis Ascher, Donald C. McKenna Professor of Government and Economics, Claremont McKenna College, Claremont, CA

Dr John Bailar, Scholar in Residence, The National Academies, Washington, DC

Dr. Mark Borsuk, Assistant Professor, Engineering Sciences, Thayer School of Engineering, Dartmouth College, Hanover, NH

Dr. Wändi Bruine de Bruin, Research Faculty, Department of Social & Decision Sciences, Carnegie Mellon University, Pittsburgh, PA

Dr Roger Cooke, Professor of Mathematics at Delft University of Technology and Chauncey Starr Senior Fellow for Risk Analysis at Resources for the Future, Resources for the Future, Washington, DC

Dr. John Evans, Senior Lecturer on Environmental Science, Harvard University, Portsmouth, NH

Dr. Scott Ferson, Senior Scientist, Applied Biomathematics, Setauket , NY

Dr. Paul Fischbeck, Professor, Engineering and Public Policy and Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA

Dr. H. Christopher Frey, Professor, Department of Civil, Construction and Environmental Engineering, College of Engineering, North Carolina State University, Raleigh, NC

Dr. Max Henrion, CEO and Associate Professor, Lumina Decision Systems, Inc., Los Gatos, CA

Dr. Alan J. Krupnick, Senior Fellow and Director, Quality of the Environment Division, Resources for the Future, Washington, DC

Dr. Mitchell J. Small, The H. John Heinz III Professor of Environmental Engineering, Department of Civil and Public Policy, Carnegie Mellon University, Pittsburgh, PA

Dr.Katherine Walker, Senior Staff Scientist, Health Effects Institute, Boston, MA

Dr. Thomas S. Wallsten, Professor and Chair, Department of Psychology, University of Maryland, College Park, MD

SCIENCE ADVISORY BOARD STAFF

Dr. Angela Nugent, Designated Federal Officer, 1200 Pennsylvania Avenue, NW 1400F, Washington, DC, Phone: 202-343-9981, Fax: 202-233-0643, (nugent.angela@epa.gov)

INTRODUCTION

EPA's Office of the Science Advisor requested that the Science Advisory Board (SAB) review a draft white paper on expert elicitation (EE) prepared by a task force of the Agency's Science Policy Council. The panel held a public meeting on February 25-26, 2009 and a public teleconference on April 22, 2009 to discuss its review of the white paper. This report addresses the charge questions requested by the Agency. In this introduction, the panel provides some overarching comments on the white paper and highlights its major recommendations for EPA's developing use of EE.

The panel commends EPA for preparing a broad and thoughtful white paper on the potential use of expert elicitation at the Agency. The white paper was written by a task force charged by the EPA Science Policy Council "to initiate a dialogue within the Agency about the conduct and use of EE and then to facilitate future development and appropriate use of EE methods" (p. 2). The panel judges that the white paper succeeds in providing much of the information needed for the proposed dialogue and to facilitate future development and appropriate use of EE. The white paper provides a comprehensive introduction to EE for readers who may be unfamiliar with it and careful discussion of many of the issues that must be faced if the Agency is to use expert elicitation (EE) in the future. This report offers some comments on the white paper and suggestions for improvement.

The panel recommends that the white paper:

1. Adopt a more neutral, analytic tone. In parts, it reads too much like an advocacy document for EE.
2. Distinguish issues particular to EE from issues that arise in any analysis of environmental intervention (e.g., problem structuring) and those that arise in any attempt to incorporate expert judgment (e.g., selection of experts to an advisory committee). Because EE is a comparatively transparent process, its use highlights many issues that are critical to other processes as well.
3. Give greater attention to the extent to which EE is a complement to rather than a substitute for other methods of quantifying uncertainty about a quantity or model parameter. 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.
4. Address methods for evaluating and ensuring the quality of expert judgments, including tests of internal consistency, coherence and, when possible, performance.
5. Discuss the tradeoffs between greater transparency (of both process and results) and resources required for a study.
6. More fully address methods for aggregating experts' judgments. Aggregation is often necessary for subsequent use of elicited quantities.

7. More carefully delineate the types of quantities suitable for EE. The panel urges that the quantities being elicited be measurable (at least in principle, if not in practice). Model-dependent parameters should be elicited only when they can be unambiguously translated into or inferred from measurable quantities.

8. Give greater attention to the need to explicitly condition the quantities being elicited on other relevant quantities. This is important because the value and uncertainty of most quantities will depend on the values of other quantities. Also, dependencies among multiple quantities being elicited may be required for subsequent use. Influence diagrams can be helpful for maintaining consistent conditioning among quantities.

9. More fully review the literature on cognitive biases that can lead to inaccurate elicitation of expert judgments.

10. 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.

11. Provide a critical analysis of the strengths and weaknesses of expert elicitation in comparison with other approaches that might be alternatives to EE such as meta-analysis, peer review, unstructured expert committees, and collecting primary data.

In addition, the panel encourages EPA to continue to explore the use of EE and to support research on EE and alternatives to gain experience and understanding of the advantages and disadvantages of EE and other methods in diverse applications.

RESPONSE TO AGENCY CHARGE QUESTIONS

Charge question A - background and definition of expert elicitation

Does the white paper provide a comprehensive accounting of the potential strengths, limitations, and uses of EE? Please provide comments that would help to further elucidate these potential strengths, limitations, and uses. Please identify others (especially EPA uses), that merit discussion.

The white paper provides a good overview of EE and issues relevant to its use by EPA. We offer some suggestions for improvement.

1. The white paper could be enhanced by adopting a more balanced, analytic tone. In particular, it could provide a critical analysis of the strengths and weaknesses of EE and compare these with the strengths and weaknesses of other approaches that might be alternatives to EE in particular cases, including meta-analysis, peer review, unstructured

expert committees (e.g., SAB, National Research Council committees), and primary data collection.

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, other approaches that can be used to characterize information. EE does not create primary data. It is a structured and rigorous process for characterizing experts' understanding of the implications of existing data and models. When predicting the consequences of alternative policies, it is typically necessary to extrapolate from the findings of empirical studies (e.g., animal to human, epidemiological cohort to general population or sensitive subgroup, past to future). EE (and other methods for incorporating expert judgment) can be used to address this extrapolation, whereas other formal methods (e.g., meta-analysis) generally cannot. In short, EE should be presented as a useful way to organize and understand what is known about a matter and to identify what remains to be studied.

EE studies can be integrated into research planning if they elicit information on how an expert's judgments would be influenced by possible outcomes of a research study. For example, experts can be queried about their probability distributions of relationships given alternative outcomes of a study (Kadane and Wolfson, 1998) or direct elicitation of the likelihood function for a proposed experiment can be made (Small, 2008). With these assessments, the EE results can be used as part of value-of-information studies to prioritize research and subsequently updated in an adaptive manner as new research results are obtained.

3. The white paper should include a fuller discussion contrasting subjective (Bayesian) and objective (frequentist) probabilities. Frequentist probabilities describe the chance of various outcomes conditional on a hypothesis (e.g., that data follow a standard normal distribution); subjective probabilities characterize an individual's degree of belief that a particular event will occur (e.g., that a specified exposure will result in cancer). For regulatory purposes, EPA is generally interested in predicting environmental and other outcomes conditional on alternative policies; hence the subjectivist interpretation is often more relevant.

Recognition of the relevance of subjective probabilities has several implications. First, EPA is generally interested in the probabilities of specific (e.g., environmental, health, economic) outcomes, not in whether a particular scientific model (e.g., linear no-threshold dose-response function) is "correct." Hence, the objective when using EE should be to elicit judgments about quantities about which people could know the truth, if the appropriate research were conducted. 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. Elicitation of such a parameter value may be appropriate (even if it is not directly measurable), as long as it can be unambiguously translated into a measurable quantity.

Second, since subjective probabilities measure an individual's degree of belief, different experts may legitimately attach different probabilities to the same event. There may be no "correct" probability and, in general, no unique or well-accepted method for choosing among probabilities held by equally well-qualified experts. EE is a method for eliciting and integrating individual experts' judgments about a matter into a coherent expression and characterizing their knowledge using probability.

4. Perhaps because it is a relatively transparent process, EE highlights several issues that are common to many methods that can be used to obtain judgments from domain experts or other individuals. The white paper would benefit from greater acknowledgment of this fact, distinguishing between issues that are common to any method of eliciting judgments from individuals and those that are specific to EE. For example, selection of experts is likely to be critical to any process for eliciting expert judgments, whether it is a survey, an expert committee (e.g., SAB, National Research Council), Delphi method, or others. Similarly, structuring the analysis and defining the parameters for which probabilities are specified are critical even when parameter values will be based on literature review, measurement, or other sources. Judgments are inherent in many decisions made by analysts regarding choice and interpretation of data, models, metrics, and results.

5. 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. Accuracy and calibration can be tested by obtaining judgments for seed quantities in the expert's field, the values of which will become known after the expert provides his distribution.

6. The white paper should be informed by and reference more recent literature. A list of suggested references appears in Appendix A.

Charge question B – transparency

Transparency is important for analyses that support Agency scientific assessments and for characterization of uncertainties that inform Agency decision making. 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). Please identify any additional strategies that could improve transparency.

In general, EE is at least as transparent as most alternative methods for obtaining expert judgments. Unlike committee processes, each expert provides a set of judgments about the quantities that are elicited and so the degree of overlap or disagreement among experts can be made readily apparent. It can be argued that transparency would be further enhanced by associating each distribution with the expert who provided it, but the panel

concludes that the disadvantages of identification (e.g., implicit pressure to provide a distribution consistent with an institutional position) more than offset the advantages.

To enhance transparency, it is important to characterize the range of expertise and identify the experts' rationales for the quantitative judgments (for credibility and to decide when new understanding renders the results obsolete). Some of the benefits of enhanced transparency include the ability to: 1) evaluate strengths and weaknesses of the study in the future; 2) evaluate and enhance credibility by demonstrating that the approach was applied rigorously; and 3) withstand litigation.

In determining what should be transparent, it is useful to distinguish between process and results. Aspects of the process that should be transparent include the methods used to select experts, their identities and relevant characteristics (e.g., scientific discipline), the questions used to elicit judgments and the methods used to ensure that the questions are clear to the experts and elicitors, and the interactions between experts and elicitors. Aspects of the results that should be transparent include the problem framing, definitions of the quantities elicited and characterization of other quantities on which the quantities that are elicited are conditioned, the experts' judgments, and their rationales for their judgments (e.g., key empirical studies, suspected biases of existing data).

The white paper should provide further discussion about how to capture each expert's assumptions and basis for his or her judgments. It should also discuss the tradeoffs associated with deepening the interactions between elicitor and expert. The extended interaction between expert and elicitor that is often employed is intended to produce a more carefully considered judgment, i.e., one that better reflects the expert's understanding of a topic. However, this interaction can influence the results as compared with a more restricted interaction, e.g., in a remotely-conducted Delphi or survey. The extent of interaction also has implications for the resources required to conduct and document a study. The interaction between expert and elicitor and the rationale for the expert's judgment may be documented through an interview transcript, a written description of the rationale that the expert drafts or approves, a brief note, or other means. It may be useful to create a table that lists the aspects that can be easily conveyed transparently and those that cannot.

Charge question C.1 – selecting experts

Section 5.2 considers the process of selecting of experts.
a) Although it is agreed that this process should seek a balanced group of experts who possess all appropriate expertise, there are multiple criteria that can be used to achieve these objectives. Does this white paper adequately address the different criteria and strategies that may be used for nominating and selecting experts?
b) Are there additional technical aspects about this topic that should be included?

The problem of expert selection is common to any effort to use expert opinion in support of the development of regulatory policy – whether informal or formal, structured or unstructured. Hence the guidance offered below applies uniformly and is not intended to be a critique of formal elicitation of expert opinion.

The panel notes that for an EE study to succeed, the experts selected must be credible, the set of experts must be acceptable to stakeholders, and the process for selection should be clearly documented and replicable. To enhance the transparency and credibility of the study, experts should articulate the basis for their judgments. When quantitative judgments are to be obtained, whether through EE or alternative methods, the study will be improved if experts have the ability to characterize their beliefs in terms of probability distributions that are well-calibrated. (Note that it is typically impossible to assess calibration of experts' judgments for the quantities that are the subject of the study, because the true values will not become known in a relevant time period. Calibration on other quantities in the expert's field, the values of which become known, can be assessed).

It can be argued that expert selection should depend on the intended purpose of the study, e.g., to elicit the range of reasonable judgments for sensitivity analysis or to represent the frequency with which different views are held in a scientific community (p. 69). The panel cautions that it is difficult to evaluate satisfaction of these criteria in advance (e.g., to determine whether an outlying perspective is "reasonable"). In addition, in some domains the set of reasonable perspectives may not be adequately represented without including more than nine experts (hence requiring approval from OMB).

Charge question C.2 – multi-expert aggregation

Sections 5.4 and 6.7 present multi-expert aggregation.

a) Among prominent EE practitioners there are varied opinions on the validity and approaches to aggregating the judgments obtained from multiple experts. Does this white paper capture sufficiently the range of important views on this topic?

b) Are there additional technical aspects about this topic that should be included?

The panel recognizes that there is disagreement among EE scholars about the extent to which multi-expert aggregation is desirable and about the most appropriate methods for aggregation when it is conducted. It offers the following remarks.

1. Some form of aggregation is usually required, whether explicit or implicit. For example, a policy maker cannot adopt different policies conditional on which expert is most accurate. When expert judgments are obtained about multiple parameters in a model, it is neither feasible nor useful to report model outputs for each combination of judgments (e.g., Expert A's judgment on parameter 1, Expert B's judgment on parameter

2, etc.) because the number of combinations grows geometrically with the number of parameters and experts.

2. The white paper devotes inadequate attention to methods of aggregating experts' judgments. It should provide a fuller discussion of performance-based methods (Cooke, 1991) and of other significant work (e.g., Jouini and Clemen 1996). Note that some methods for aggregating judgments require that particular information be collected as part of the elicitation (e.g., judgments on seed variables, peer or self weights reflecting expert quality).

3. Whether experts' judgments are combined or not, each judgment should be reported individually. This allows readers to see the individual judgments, to evaluate their similarities and differences, and potentially to aggregate them using alternative approaches. When the effects on model outputs of differences among experts' judgments about input values are not obvious, it may be useful to also report how model outputs depend on differences among the experts' judgments.

Charge question C.3 – problem structure

Section 5.2.2 discusses how the problem of an EE assessment is structured and decomposed using an "aggregated" or "disaggregated" approach.
a) The preferred approach may be influenced by the experts available and the analyst's judgment. Does this discussion address the appropriate factors to consider when developing the structure for questions to be used in an EE assessment?
b) Are there additional technical aspects about this topic that should be included?

The panel agrees that the problem structure must be acceptable to the experts, specifically that it accords with their knowledge. It urges that the quantities for which judgments are elicited be quantities that are measurable (at least in principle, if not necessarily in practice). To the extent that experts use a common model that permits unambiguous translation between a model parameter and a quantity that is measurable (in principle), elicitation of judgments about the parameter may be more convenient.

The white paper should give more attention to dependence among quantities. Dependence is important for at least two reasons. First, for experts to provide judgments about the value of some quantity, they must be told the variables on which that quantity is being conditioned. Second, when experts are asked to provide judgments about multiple quantities, dependencies among these quantities may be relevant.

Regarding the first point, if the quantity being elicited is dependent on other variables, then the expert must be told which of those variables should be considered known (or held constant) and which should be considered unknown (or left unspecified). For the variables considered to be known, the values must be specified so that the expert

can take into account their influence on the elicited quantity. The influence of variables left unspecified must be folded into the expert's uncertainty distribution.

Regarding the second point, when experts are asked to provide judgments about multiple quantities, dependencies among these quantities may be a serious concern. For example, using independent marginal distributions (ignoring correlation) for multiple uncertain parameters in a model can produce misleading outputs. Elicitation of mutually dependent quantities is complex and there is as yet no accepted best method. Evans et al. (1994) illustrate one approach, in which dependencies among multiple factors relating to the toxicity of chloroform were illustrated as a detailed tree and judgments about each factor were conditioned on the values of other factors in the tree. Jones et al. (2001) initially elicited marginal distributions for continuous variables, then characterized dependence by asking experts to report the probability that one variable would exceed its subjective median conditional on another variable exceeding its subjective median. Clemen et al. (2000) report experimental tests of different methods; more recent methods are discussed by Kurowicka and Cooke (2006).

The "clairvoyance test," which requires "that an omniscient being with complete knowledge of the past, present, and future could definitively answer the question" (p. 12, fn. 4) attempts to capture the first issue of dependence but is inadequately articulated. A better approach is to describe the measurement that one would make to determine the value of the parameter, including which of the other factors would be controlled. To illustrate, consider the elicitation of an expert's judgment about the maximum hourly ozone concentration in Los Angeles next summer. Maximum hourly ozone depends on temperature, wind speed and direction, precipitation, motor-vehicle emissions, and other factors. Depending on the purpose of the elicitation, the distribution of some of these may be specified. A clairvoyant would know the actual values of all these factors, but the expert cannot. Uncertainty about the values of the factors that are not specified must be folded into the expert's distribution. If experts are also asked their judgment about PM concentrations, the conditionalization on factors affecting PM concentrations should be consistent with that for the ozone question.

Maintaining a consistent conditionalization across a large study is critical. Problem structure and consistent conditionalization can be facilitated by use of an influence diagram that depicts the variables of interest and causal relationships or dependencies among these variables. The panel recommends replacing the diagram in Figure 6.1 with one formatted as an influence diagram showing relationships among variables.

The white paper identifies four categories of uncertainty (parameter, model, scenario, and decision-rule) and suggests that EE may be used to address each of them (pp. 50-51). The panel suggests that scenario and decision-rule uncertainty are not suitable objects for EE. Scenario uncertainty involves questions of designing the analysis (e.g., selection of receptor populations and exposure sources to include). Scenario design may affect experts' judgments about a quantity (because the quantity may be conditional

on factors that are specified by the scenario), but EE is not an appropriate tool for obtaining expert judgment about how best to design scenarios (although expert judgments about the relative importance of multiple factors to the value of an endpoint of interest can be a relevant input to scenario design). Decision-rule uncertainty concerns the principles that will be used to make a policy decision. This choice is one to be made by policy makers subject to statute, guidance, and other applicable criteria, not by expert judgment about what principles will (or should) be applied.

The white paper distinguishes scientific information from social values or preferences and suggests that EE should not be used to provide values and preferences (pp. 11, 110). The panel acknowledges the distinction between consequences and values or preferences but notes that description of public preferences used as inputs to economic evaluation (e.g., willingness to pay for a specific reduction in health risk) is a scientific question that may be legitimately addressed using EE.

Charge question C.4 & 5 – findings and recommendations

4) Sections 7.1 and 7.2, presents the Task Force's findings and recommendations regarding: 1) selecting EE as a method of analysis, 2) planning and conducting EE, and 3) presenting and using results of an EE assessment. Are these findings and recommendations supported by the document?

5) Please identify any additional findings and recommendations that should be considered.

Overall, the findings and recommendations are supported by the white paper. The panel suggests that these sections should include a more balanced discussion of the strengths and weaknesses of EE and compare its use with other tools.

An important topic that receives little attention in the white paper is that of the coherence of judgments from an expert. When an expert provides probability distributions to characterize personal knowledge about each of several quantities, the expert is providing information about a multivariate probability distribution. When there are dependencies among variables, it can be very easy to report distributions that do not satisfy basic properties of multivariate distributions (e.g., that the covariance matrix is positive semidefinite). Elicitation protocols should be structured to help an expert provide a coherent multivariate distribution that is consistent with his or her knowledge, for example by eliciting distributions of one variable conditional on several alternative levels of another variable on which it is dependent, rather than eliciting a correlation coefficient between the two variables. Elicitation protocols can also include consistency checks, both to test for coherence of probability distributions and to confirm that the judgments are consistent with the expert's information.

The literature on cognitive biases is richer than is indicated in the white paper. In

addition to estimation biases such as anchoring and availability heuristics that are discussed, there are biases relating to uncertainty perception such as probability misperception, the conjunction fallacy, pseudocertainty, overconfidence, base-rate fallacy, and neglect of probability, all of which may distort perceptions of experts (Tucker et al., 2008). Strategies for overcoming these cognitive illusions and biases to ensure accurate and honest assessments should be discussed.

The panel suggests that the white paper could be made more accessible to the wide audience for which it is intended by including in the white paper glossary additional key terms with practical definitions. A list of some suggested terms is attached (Appendix B).

Charge question D – development of future guidance

As EPA considers the future development of guidance beyond this white paper, what additional specific technical areas should be addressed? What potential implications of having such guidance should be considered? Do the topics and suggestions covered in the white paper regarding selection, conduct, and use of this technique provide a constructive foundation for developing "best practices" for EE methods?

The topics and suggestions covered in the white paper regarding selection, conduct, and use of EE provide a constructive foundation for developing a description of "best practices" for EE, but some parts of the white paper should be revised to incorporate newer literature than is currently included (e.g., cognitive biases and elicitation of quantities, methods for assessing performance of experts, and aggregation of judgments across experts).

In considering moving to guidance, the panel counsels EPA to be careful not to stifle innovation in EE methods and to encourage research on the performance of EE and alternative methods for characterizing uncertainty. As noted in the white paper, considerable experience with structured expert judgment exists in other fields, including nuclear, aerospace, volcanology, health, and finance. The challenge is to bring this experience to bear on the specific problem areas within EPA's mandate. It may be useful for EPA to conduct several EE studies on issues that are not critical to current policy decisions, employing different methods and evaluating results. Different teams could employ different methods to a common quantity to facilitate comparison of results. The panel encourages the development of guidance characterized as a set of goals and criteria for evaluating success that can be met by multiple approaches rather than something that will be used as a checklist or "cookbook."

References

Clemen, R. T., Fischer, G. W., and Winkler, R. L. 2000. Assessing dependence: Some experimental results. *Management Science* 46, 1100-1115.

Cooke, R.M., and L.J.H. Goossens, *Procedures Guide for Structured Expert Judgment*, European Commission Directorate-General for Research, EUR 18820, 2000.

Cooke, R.M., 1991. *Experts in Uncertainty*, Oxford University Press..

Jones, J.A. et al., 2001. Probabilistic Accident Consequence Uncertainty Assessment using COSYMA: Methodology and Processing Techniques, EUR 18827, European Communities.

Jouini, M.N., and R.T. Clemen. 1996. Copula Models for Aggregating Expert Opinions. *Operations Research* 44: 444-457.

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

Kurowicka, D., and R.M. Cooke. 2006. *Uncertainty Analysis with High Dimensional Dependence Modeling*, Wiley.

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

Tucker, W.T., S. Ferson, A. Finkel, and D. Slavin (eds.) 2008. Strategies for Risk Communication: Evolution, Evidence, Experience. *Annals of the New York Academy of Sciences*, Volume 1128, Blackwell Publishing, Boston.

Appendix A
Suggested additional references for inclusion in a revised White Paper

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

Ariely, D., Au, W-T, Bender, R. H., Budescu, D. U., Dietz, C. B., Gu, H., Wallsten, T.S., and Zauberman, G. 2000. The effects of averaging probability estimates between and within judges. *Journal of Experimental Psychology: Applied* 6, 130-147.

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.

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

Converse, J.M., & Presser, S. 1986. Quantitative Applications in the Social Sciences. Survey questions: Handcrafting the standardized questionnaire. Newbury Park: Sage Publications."

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

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.

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

Glimcher, P.W. 2003. *Decisions, Uncertainty, and the Brain: The Science of Neuroeconomics*. MIT Press/Bradford Press.

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

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

Karlin, S. and W. J. Studden. 1966. *Tchebyshev Systems: With Applications in Analysis and Statistics*. Interscience, New York.

Krieglera, E. et al. 2009. *Imprecise probability assessment of tipping points in the climate system*. PNAS 106,, 5041–5046.

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

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.

Tuomisto, J.T., A. Wilson, J.S. Evans, M. Tainio. 2008. Uncertainty in mortality response to airborne fine particulate matter: Combining European air pollution experts, *Reliability Engineering and System Safety* 93(5): 732-744.

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

Smith, J.E. 1990. Moment Methods for Decision Analysis. Ph.D. Dissertation, Stanford University, Stanford, California.

Wallsten, T.S.,& Diederich, A 2001. Understanding Pooled Subjective Probability Estimates. *Mathematical Social Science*, 41, 1-18.

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

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

In addition, many useful documents are available at the following websites:

NUREG EU Probabilistic accident consequence uncertainty analysis
http://www.osti.gov/bridge/basicsearch.jsp

EU Probabilistic accident consequence uncertainty assessment using COSYMA
http://cordis.europa.eu/fp5-euratom/src/lib_docs.htm

RFF workshop expert judgment
http://www.rff.org/rff/Events/Expert-Judgment-Workshop.cfm

Radiation Protection Dosimetry 90 (2000)
http://rpd.oxfordjournals.org/content/vol90/issue3/index.dtl

TU Delft Website
http://dutiosc.twi.tudelft.nl/~risk/

Appendix B

Suggested terms to add to the glossary in the White Paper and to use consistently throughout the document

Accurate
Aggregation
Assumption
Assumptions
Availability
Averaging
Bias
Cognitive Illusion
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 Probability
Weighting