



**UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460**

**OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD**

January 21, 2010

EPA-SAB-10-003

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 defines 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." In response to the Agency's request, an SAB panel conducted a peer review of the draft white paper. The enclosed advisory report responds to the charge questions posed by the Agency.

The SAB commends the task force for preparing a comprehensive 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." The SAB judges that the white paper succeeds in providing much 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 SAB offers some recommendations to improve the white paper:

1. Describe the strengths and weaknesses of EE in comparison with those of other approaches for aggregating information and quantifying uncertainty from multiple sources. Other methods for aggregating information include meta-analysis and expert committees. This discussion should consider when EE should be used as a complement or substitute for other methods.
2. Maintain and emphasize the distinction between issues that are particular to EE and issues that arise in any analysis of environmental policy or in any method to incorporate expert judgment. Because EE is a transparent method, it can highlight issues such as selection of experts, cognitive biases, and problem structuring that are also important for other approaches.
3. Address methods for evaluating and ensuring the quality of the elicited judgments, including tests of coherence (e.g., consistency among judgments of mutually dependent quantities) and performance (e.g., calibration, defined as consistency of elicited probability distributions with true values of quantities, which can only be evaluated for quantities whose values become known).
4. Expand the discussion about combining judgments across experts to consider: (a) how the decision about whether and how to combine depends on the objective of the study; (b) the level of the analysis at which to combine (e.g., combine judgments about a model input or combine model outputs derived by running a model using each expert's judgment about the input); and (c) performance-based methods for combination.
5. More carefully delineate the types of quantities suitable for EE. The SAB recommends that the quantities being elicited be measurable (at least in principle, if not in practice). Models used in environmental assessment are, of course, simplifications of the real world and often include parameters that do not correspond to any measurable feature of the real world (e.g., transfer coefficients in a compartmental fate-and-transfer model; dispersion coefficients in an atmospheric model). Model-dependent parameters should be elicited only when they can be unambiguously translated into or inferred from measurable quantities.
6. Give greater attention to the need to be explicit about the values of other quantities that are relevant to the quantity being elicited. This is important for two reasons. First, an expert's judgment about the value of a quantity will depend on whether other quantities are fixed, and if so at what values. (If not fixed, the expert must incorporate uncertainty about the values of these other quantities and their effects on the value of the elicited quantity into his judgment.) Second, when multiple quantities are elicited, the values of some of them may be mutually dependent (e.g., the value of one quantity may depend on the value of another or some common factor may influence the values of both quantities). If the quantities are used as inputs to a model, it may be important to incorporate the dependence among them in order to accurately characterize uncertainty about the model output. Influence diagrams can be helpful for maintaining consistency about the values at which quantities are fixed.

7. Emphasize the need for flexibility in EE implementation. The SAB 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.

Finally, the SAB 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,

/Signed/

Dr. Deborah L. Swackhamer
Chair
Science Advisory Board

/Signed/

Dr. James K. Hammit
Chair
Science Advisory Board Expert
Elicitation Advisory Panel

Enclosures

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board (SAB), a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The SAB is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names of commercial products constitute a recommendation for use. Reports of the SAB are posted on the EPA website at <http://www.epa.gov/sab>.

Enclosure A
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 & Environmental Engineering and Engineering & 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, Washington, DC

Enclosure B
U.S. Environmental Protection Agency
Science Advisory Board
Fiscal Year 2009

CHAIR

Dr. Deborah L. Swackhamer, Professor of Environmental Health Sciences and Co-Director Water Resources Center, Water Resources Center, University of Minnesota, St. Paul, MN

SAB MEMBERS

Dr. David T. Allen, Professor, Department of Chemical Engineering, University of Texas, Austin, TX

Dr. John Balbus, Adjunct Associate Professor, George Washington University, School of Public Health and Health Services, Washington, DC

Dr. Gregory Biddinger, Coordinator, Natural Land Management Programs, Toxicology and Environmental Sciences, ExxonMobil Biomedical Sciences, Inc., Houston, TX

Dr. Timothy Buckley, Associate Professor and Chair, Division of Environmental Health Sciences, School of Public Health, The Ohio State University, Columbus, OH

Dr. Thomas Burke, Professor, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

Dr. James Bus, Director of External Technology, Toxicology and Environmental Research and Consulting, The Dow Chemical Company, Midland, MI

Dr. Deborah Cory-Slechta, Professor, Department of Environmental Medicine, School of Medicine and Dentistry, University of Rochester, Rochester, NY

Dr. Terry Daniel, Professor of Psychology and Natural Resources, Department of Psychology, School of Natural Resources, University of Arizona, Tucson, AZ

Dr. Otto C. Doering III, Professor, Department of Agricultural Economics, Purdue University, W. Lafayette, IN

Dr. David A. Dzombak, Walter J. Blenko Sr. Professor of Environmental Engineering, Department of Civil and Environmental Engineering, College of Engineering, Carnegie Mellon University, Pittsburgh, PA

Dr. T. Taylor Eighmy, Interim Vice President for Research, Office of the Vice President for Research, University of New Hampshire, Durham, NH

Dr. Baruch Fischhoff, Howard Heinz University Professor, Department of Social and Decision Sciences, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA

Dr. James Galloway, Professor, Department of Environmental Sciences, University of Virginia, Charlottesville, VA

Dr. John P. Giesy, Professor, Department of Zoology, Michigan State University, East Lansing, MI

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

Dr. Rogene Henderson, Senior Scientist Emeritus, Lovelace Respiratory Research Institute, Albuquerque, NM

Dr. James H. Johnson, Professor and Dean, College of Engineering, Architecture & Computer Sciences, Howard University, Washington, DC

Dr. Bernd Kahn, Professor Emeritus and Director, Environmental Radiation Center, Nuclear and Radiological Engineering Program, Georgia Institute of Technology, Atlanta, GA

Dr. Agnes Kane, Professor and Chair, Department of Pathology and Laboratory Medicine, Brown University, Providence, RI

Dr. Meryl Karol, Professor Emerita, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA

Dr. Catherine Kling, Professor, Department of Economics, Iowa State University, Ames, IA

Dr. George Lambert, Associate Professor of Pediatrics, Director, Center for Childhood Neurotoxicology, Robert Wood Johnson Medical School-UMDNJ, Belle Mead, NJ

Dr. Jill Lipoti, Director, Division of Environmental Safety and Health, New Jersey Department of Environmental Protection, Trenton, NJ

Dr. Lee D. McMullen, Water Resources Practice Leader, Snyder & Associates, Inc., Ankeny, IA

Dr. Judith L. Meyer, Distinguished Research Professor Emeritus, Odum School of Ecology, University of Georgia, Athens, GA

Dr. Jana Milford, Professor, Department of Mechanical Engineering, University of Colorado, Boulder, CO

Dr. M. Granger Morgan, Lord Chair Professor in Engineering, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA

Dr. Christine Moe, Eugene J. Gangarosa Professor, Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA

Dr. Duncan Patten, Research Professor , Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT, USA

Mr. David Rejeski, Director, Foresight and Governance Project , Woodrow Wilson International Center for Scholars, Washington, DC

Dr. Stephen M. Roberts, Professor, Department of Physiological Sciences, Director, Center for Environmental and Human Toxicology, University of Florida, Gainesville, FL

Dr. Joan B. Rose, Professor and Homer Nowlin Chair for Water Research, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI

Dr. Jonathan M. Samet, Professor and Chair , Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

Dr. James Sanders, Director and Professor, Skidaway Institute of Oceanography, Savannah, GA

Dr. Jerald Schnoor, Allen S. Henry Chair Professor, Department of Civil and Environmental Engineering, Co-Director, Center for Global and Regional Environmental Research, University of Iowa, Iowa City, IA

Dr. Kathleen Segerson, Professor, Department of Economics, University of Connecticut, Storrs, CT

Dr. Kristin Shrader-Frechette, O'Neil Professor of Philosophy, Department of Biological Sciences and Philosophy Department, University of Notre Dame, Notre Dame, IN

Dr. V. Kerry Smith, W.P. Carey Professor of Economics , Department of Economics , W.P Carey School of Business , Arizona State University, Tempe, AZ

Dr. Thomas L. Theis, Director, Institute for Environmental Science and Policy, University of Illinois at Chicago, Chicago, IL

Dr. Valerie Thomas, Anderson Interface Associate Professor, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA

Dr. Barton H. (Buzz) Thompson, Jr., Robert E. Paradise Professor of Natural

Resources Law at the Stanford Law School and Perry L. McCarty Director, Woods Institute for the Environment Director, Stanford University, Stanford, CA

Dr. Robert Twiss, Professor Emeritus, University of California-Berkeley, Ross, CA

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

Dr. Lauren Zeise, Chief, Reproductive and Cancer Hazard Assessment Branch, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Oakland, CA

SCIENCE ADVISORY BOARD STAFF

Mr. Thomas Miller, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

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

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. As described in the white paper, EE "is a formal, systematic process of obtaining and quantifying expert judgment on the probabilities of events relationships, or parameters...It can enable quantitative estimation of uncertain values and can provide uncertainty distributions where data are unavailable or inadequate. In addition, EE may be valuable for questions that are not necessarily quantitative such as model conceptualization or design of observational systems." The white paper describes EPA's experience with EE in the Office of Air and Radiation, Office of Air Quality Planning and Standards and the use or recommendation of the approach for such different EPA applications as assessing the magnitude of sea level risk associated with climate change and ecological model development.

The draft white paper was intended "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." The white paper discussed the potential utility of using expert elicitation to support EPA regulatory and non-regulatory analyses and decision-making. It provided recommendations for expert elicitation "good practices" and described steps for a broader application across EPA.

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 comprehensive overview of EE, its strengths and limitations, and issues relevant to its use by EPA. We offer some suggestions for possible improvement.

1. Include a more focused discussion of when to use EE that compares it with other approaches that might be used as alternatives, or complements, in particular cases.

EE is a method to characterize what is known about the value of a quantity of interest. For example, EPA may be concerned about the shape and slope of an exposure-response function (e.g., when analyzing the consequences of policies to control exposure to a pollutant). EE is a structured method for synthesizing existing data, models, and understanding by eliciting subjective probability distributions from subject-matter experts. Other methods for characterizing such quantities by synthesizing existing information include, inter alia, (a) unstructured expert judgment of EPA or other analysts, perhaps complemented by literature review, (b) meta-analysis of empirical studies, (c) unstructured expert committees (e.g., SAB, National Research Council), and (d) structured group processes (e.g., Delphi). Another method

to estimate a quantity of interest is to collect additional primary data. Primary data collection can provide more data, and data that are more relevant to the problem that motivates EPA's interest.

EE can be employed as a substitute or complement to other approaches. In some cases, results of a single empirical study or a meta-analysis of multiple studies may provide an appropriate characterization of what is known about a quantity. In others, it may be appropriate to conduct a meta-analysis as input to EE. In still other cases, it may be appropriate to conduct EE without any meta-analysis. Even when additional primary data are collected, it may still be appropriate to conduct an EE to interpret the implications of these data for the problem of interest to EPA.

EE may be particularly useful in cases where it is necessary to extrapolate some distance from available data (e.g., from data on laboratory animals to humans, or from epidemiological data on an occupationally exposed human population to an environmentally exposed population).

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 identify research priorities and may be updated in an adaptive manner as new research results are obtained.

In summary, EE is a useful way to organize and understand what is known about a quantity and to identify what remains to be studied.

2. Include a fuller discussion contrasting subjective (Bayesian) and objective (frequentist) probabilities. Frequentist probabilities describe the (objective) chance of an outcome conditional on a hypothesis (e.g., the probability an individual with specified exposure will develop cancer conditional on a linear no-threshold dose-response model with specified slope); subjective probabilities characterize an individual's degree of belief that a particular event will occur (e.g., that an individual with specified exposure will develop cancer).

Recognition of the relevance of subjective probabilities has several implications. First, EPA is generally interested in the probabilities of specific environmental, health, and economic outcomes, not in whether a particular scientific model is "correct." In an oft-quoted remark of George Box, "all models are wrong, but some are useful." In evaluating the outcomes of alternative policies, EPA should (and sometimes does) incorporate uncertainty about which of several models provides the best approximation.

Second, 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. The white paper describes the goal of EE as characterization of experts' beliefs "about relationships, quantities, events, or parameters of interest" (p. 22). Quantities and events, if potentially measurable, are appropriate objects for elicitation. In contrast, elicitation of relationships or parameters that cannot be measured, even in principle, can be dangerous. Experts who do not work with the

specific model in which a parameter is defined may have little knowledge about the value of the parameter. Moreover, the relationship between the parameter value and outcomes that are potentially measurable may depend on the choice among several alternative models, some or all of which the expert may reject.

Consider an example from Jones et al. (2001). The spread of a radioactive plume from a power plant is often modeled as a power-law function of distance, i.e.,

$$\sigma(x) = P x^Q,$$

where σ is the lateral plume spread and x is downwind distance from the source. P and Q are parameters whose values depend on atmospheric stability at the time of release. This model is not derived from physical laws but provides a useful description when parameters are estimated using results of tracer experiments. Experts have experience with values of $\sigma(x)$ measured in tracer experiments and values of lateral spread at multiple distances from the source can be elicited. However, the problem of “probabilistic inversion,” i.e., identifying probability distributions on P and Q that, when propagated through the model, produce the elicited distributions for lateral spread is difficult; indeed, there may not be any solution or the solution may not be unique (Jones et al., 2001; Cooke and Kraan, 2000). It is unreasonable to expect an expert to be able to perform this probabilistic inversion in the context of an EE. (Note that the problem of probabilistic inversion also exists when the distributions of lateral spread are obtained from measurements rather than from EE.) Other examples of model parameters that may not be suitable quantities for elicitation abound. These include the transfer coefficients in compartmental models describing environmental fate and transport or pharmacokinetics in the human body and the parameters of the multistage dose-response model often used for carcinogenic chemicals.

Third, 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 well-qualified experts. EE is a method for eliciting and integrating an expert’s judgments about a quantity into a coherent expression and characterizing the expert’s knowledge using probability.

3. Distinguish issues that are specific to EE from those that are common to any method of eliciting judgments or those common to any method for assessing consequences of environmental policies. Perhaps because it is a relatively transparent process, EE highlights many issues that are common to other methods that can be used to obtain judgments from domain experts or other individuals (as recognized in the white paper). For example, selection of experts is likely to be critical to EE, expert committees (e.g., SAB, National Research Council), Delphi methods, surveys, and peer review. Structuring the analysis and defining the quantities of interest are critical even when values will be obtained by literature review, measurement, or other methods that do not require explicit participation by experts. Judgments are inherent in many decisions made by analysts regarding choice and interpretation of data, models, metrics, and results.

4. The white paper could 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.

Overall, the white paper is sensitive to issues of transparency. However, the extent to which “mechanisms for ensuring transparency” are described varies by topic. The white paper does present adequate mechanisms for ensuring transparency with regard to selecting experts and presenting and using EE results, but does not present such mechanisms when considering the use of EE.

Although chapter 4 discusses a wide range of factors that should be considered when determining whether to conduct an EE study, it does not appear to describe any mechanisms for ensuring transparency about this decision. The question of whether to use EE in a particular instance should be viewed as part of the larger question of which analytic methods to use, and any mechanisms for ensuring transparency about choice of methods should be applicable to consideration of whether to use EE. Transparency regarding choice of methods is perhaps best ensured through including a discussion of methods whenever results of an analysis are presented. This discussion can include description of the rationale for the particular methods chosen and discussion of the comparative strengths and weaknesses of alternative methods that were not adopted.

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. Although it can be argued that transparency would be further enhanced by associating each distribution with the expert who provided it, 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 in most cases.

To enhance transparency, it is important to characterize expertise of the experts (individually and jointly) and to identify the experts’ rationales for their 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 and other challenges.

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 could provide further discussion about how to capture each expert's assumptions and basis for his or her judgments, acknowledging 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 each 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 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.

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?

Section 5.2 provides a good description of criteria and strategies for selecting experts. As noted, the problem of expert selection is common to any effort to use expert judgment in support of the development of regulatory policy – whether informal or formal, structured or unstructured. Hence the guidance offered below applies to other methods of including expert judgment as well.

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 better if experts have the ability to characterize their beliefs in terms of probability distributions that are well-calibrated and informative (i.e., relatively sharp). Typically, it is 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. There are exceptions, however: Hawkins and Evans (1989) and Walker et al. (2003) evaluated individual experts' judgments about subsequently measured human exposure to hazardous air pollutants. Calibration on seed variables (i.e., other quantities in the

expert's field, the values of which become known in a timely manner) can be assessed. A test for whether assessing calibration on seed variables is useful is to ask whether the perceived quality of the experts' judgments on the quantities of interest is affected by their performance (collectively or individually) on the seed variables. Assessing experts' calibration on almanac questions (e.g., the length of the Nile River) is not useful when such questions are not within their domain of expertise and not relevant to quantities that are of interest.

The white paper suggests that expert selection may depend on whether the purpose of the study is to elicit the range of reasonable judgments or to provide a central estimate of the scientific community (pp. 69, 72). The panel offers two cautions: First, it may be difficult to select experts to represent the range of reasonable judgments because their judgments may not be known before the elicitation and it may be difficult to determine what judgments are "reasonable." Second, scientific truth is not determined by majority vote, and so the frequency with which a view is held is not necessarily a good indicator of its validity. Moreover, estimates of any central tendency from an EE study may be sensitive to the exact set of experts selected, because of the small number of experts included. Moreover, it is difficult to recruit a valid probability sample of experts because of difficulties in (a) defining the universe from which a sample should be drawn and (b) overcoming selection biases associated with experts' availability and willingness to participate in what can be a time-consuming and challenging process.

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?*

As noted in the white paper, 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. The extent to which aggregation may be appropriate may depend on the purpose of the study (e.g., to estimate consequences of a policy change or to characterize current understanding of some relationship). Aggregation of experts' judgments can be considered part of a more general question about when to aggregate across sources of information.

One aspect of this question is: how much should analysts aggregate across information sources when presenting estimates of policy consequences to a policy maker (and to other interested parties)? Information sources can include not only individual experts but also alternative models (e.g., dose-response models with or without a threshold), data used to estimate model parameters (e.g., different epidemiological cohorts), and others. One possibility is to aggregate as many relevant information sources as possible and to present the results in the form of a probability distribution or other summary of the likely magnitude of effects for relevant endpoints. Another possibility is to present multiple estimates of the magnitude of effects based on alternative information sources so that the policy maker (and others) can aggregate these multiple estimates judgmentally or using some other approach. Clearly, some

aggregation is virtually always required to yield a manageable number of alternative estimates for the decision maker to consider (e.g., even if there are only three parameters and three information sources for each, there are 27 alternative estimates). However, some indication of how the estimates depend on critical choices among information sources is also useful.

A second aspect of the question is: at what stage of analysis to aggregate? With a non-linear model, the output when running the model using parameters based on an aggregation of information sources will generally differ from an aggregation of the outputs obtained when running the model using parameters based on each information source alone.

The white paper would be improved by including a fuller discussion of performance-based combination methods (Cooke, 1991). Note that it is possible to empirically evaluate the quality of alternative methods for combining distributions when the values of the quantities that are elicited become known. For example, Cooke and Goossens (2008) compared the performance of alternative methods of combining experts' distributions for seed variables (see Clemen, 2008, and Cooke, 2008 for discussion), and one could evaluate the quality of alternative combinations of expert judgments in cases where the values of the target quantities become known (e.g., Walker et al., 2003; Hawkins and Evans, 1989).

Whether experts' judgments are combined or not, the panel agrees with the recommendation that each judgment be reported individually (p. 83). 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 (see related discussion and examples in response to charge question A).

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 values of other quantities on which that quantity is being conditioned. Second, when experts are asked to provide judgments about multiple quantities, it may be important to elicit their judgments about dependencies among these quantities as well.

Regarding the first point, if the quantity being elicited is dependent on the values of other quantities, then the expert must be told which of those quantities should be considered known (or held constant) and which should be considered unknown (or left unspecified). For the quantities 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 quantities left unspecified must be folded into the expert's uncertainty distribution.

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 on other quantities) but is inadequately articulated. A better approach is to describe the measurement that one would make to determine the value of the quantity, 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.

Regarding the second point, when experts are asked to provide judgments about multiple quantities, dependencies among these quantities may be important. 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) 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).

Maintaining a consistent “conditionalization” (i.e., a set of assumptions about which quantities are fixed at what levels or following what probability distribution) 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 scenarios that provide useful information about how the outputs of a model depend on various assumptions about input values. This question is distinct from about the magnitude of a potentially measurable quantity, such as a model input. Hence EE is not an appropriate tool for obtaining expert judgment about how best to design scenarios (although expert judgments about the values of input quantities, the relative importance of multiple factors to the value of an endpoint, or other issues can be a relevant input to scenario design). Decision-rule uncertainty concerns the principles that will be used to make a policy decision. The choice of principles 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 value judgments and preferences and suggests that EE should not be used to provide values and preferences (pp. 11, 110). The panel acknowledges the distinctions between consequences, values, and preferences but notes that characterization of public preferences that may be used as inputs to economic evaluation (such willingness to pay for a specified reduction in health risk) is a scientific question that may be legitimately addressed using EE. Description of public preferences is distinct from the question of the role of these preferences in policy making. Analogously, whether the dose-response function for a toxicant has a threshold and the level of the threshold are scientific questions that are distinct from the questions of whether and how these quantities should be used in policy making.

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 discussion of the strengths and weaknesses of EE as compared with other approaches (e.g., meta-analysis, expert committees).

An important topic that receives little attention in the white paper is that of the coherence of judgments of 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, base-rate fallacy, and neglect of probability, all of which may distort experts' perceptions (Tucker et al., 2008). Strategies for overcoming these cognitive illusions and biases to ensure accurate and honest assessments should be discussed.

The white paper reports, accurately, that EEs conducted in the manner it describes require substantial resources – they are neither quick nor inexpensive. The quantity of resources needed for an EE depend on the complexity of the question, including the need to structure the problem so that the quantities are sufficiently well-defined that they are appropriate for elicitation, the number of experts, the need for pre- or post-elicitation workshops, the extent to which the elicitation interview and the rationale for specific judgments are documented, and other factors. Some studies have been conducted at lower cost, e.g., of the 45 studies conducted by the group at Technical University Delft, most required between one and three person-months (Cooke and Goossens, 2008) although others have required one person-year and up to a week of time from each expert (Goossens et al, 2008). It would be useful to clarify the tradeoffs between cost and quality of the results of an EE study and to understand how it varies with study design.

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. Some suggested terms are listed in 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 the development of 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. 2008. Comment on Cooke's Classical Method. *Reliability Engineering and System Safety* 93: 760-765.
- Clemen, R. T., Fischer, G. W., and Winkler, R. L. (2000). Assessing dependence: Some experimental results. *Management Science* 46, 1100-1115.
- Cooke, R.M. 1991. *Experts in Uncertainty: Opinion and Subjective Probability in Science*, Oxford.
- Cooke, R.M. 2008. Response to Comments. *Reliability Engineering and System Safety* 93: 775-777.
- Cooke, R.M., and B. Kraan, 2000. Uncertainty in Compartmental Models for Hazardous Materials – A Case Study. *Journal of Hazardous Materials* 71: 253-268.
- Cooke, R.M., and L.J.H. Goossens. 2000. *Procedures Guide for Structured Expert Judgment*, European Commission Directorate-General for Research, EUR 18820.
- Cooke, R.M., and L.J.H. Goossens. 2008. TU Delft Expert Judgment Data Base. *Reliability Engineering and System Safety* 93: 657-674.
- Goossens, L.J.H., R.M. Cooke, AR. Hale, and Lj. Rodic-Wiersma. 2008. Fifteen Years of Expert Judgment at TU Delft. *Safety Science* 46: 234-244.
- Hawkins, N.C., and J.S. Evans. 1989. Subjective Estimation of Toluene Exposures: A Calibration Study of Industrial Hygienists. *Applied Industrial Hygiene*, 4: 61-68.
- Jones, J.A. et al., 2001. Probabilistic Accident Consequence Uncertainty Assessment using COSYMA: Methodology and Processing Techniques, EUR 18827, European Communities.
- 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.
- Cooke, RM, Goossens L.H.J. 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.
- Morgan, M.G., Fischhoff, B., Bostrom, A., & Atman, C. 2001. *Risk communication: The mental models approach*. New York: Cambridge University Press.
- Morgan, M.G., H. Dowlatabadi, M. Henrion, D. Keith, R. Lempert, S. McBrid, M. Small, T. Wilbanks (eds.), *Best Practice Approaches for Characterizing, Communicating, and Incorporating Scientific Uncertainty in Decisionmaking*, Final Report, Synthesis and Assessment Product 5.2, CCSP, National Oceanic and Atmospheric Administration, Washington D.C., 2009. available at <http://www.climate-science.gov/Library/sap/sap5-2/final-report/default.htm>.

- 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
Conditionalization
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
Seed variable
Subjective
Subjective probability
Weighting