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 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" (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 broad 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. 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.

Deleted: comprehensive

Deleted: reasonable

Deleted: report offers some comments on the white paper and suggestions for improvement. In particular, the panel suggests

Formatted: bibliography, No bullets or numbering

Formatted: Bullets and Numbering

3. Give greater attention to the extent to which EE is a complement rather than a substitute for other methods of quantifying uncertainty about a quantity or model parameter. We suggest that EE should be presented as a useful way to organize and synthesize what is already known about a quantity and not as a means for generating new primary data.

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 issue of transparency in the context of the time and costs involved. Some methods of elicitation would be extremely difficult to fully document, suggesting a tradeoff between full transparency and limited resources.

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). Only when experts agree on a common model that permits unambiguous translation from an unobservable parameter to a measurable quantity should values of parameters be elicited directly.

8. Give greater attention to the need to explicitly condition the quantities being elicited on other relevant quantities. This is important because both the value and the uncertainty of most quantities will be dependent on the values of other quantities. Also, the specific nature of dependencies among multiple quantities being elicited may be required for subsequent use. The panel suggests that explicit conditioning may often be facilitated by the use of influence diagrams.

9. More fully review the literature on cognitive biases which may lead to incorrect 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.

The panel believes it is important for EPA to provide a critical analysis of the strengths and weaknesses of expert elicitation in comparison with those of other approaches that might be alternatives to EE in particular cases, including meta-analysis, peer review, unstructured expert committees, and additional original research to characterize or reduce the uncertainty of concern. It understands that EPA is preparing another white paper on the "Hierarchy of Methods for Characterizing Uncertainty" that will discuss the choice among alternative methods and recommends that the EE white paper reference this forthcoming document.

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

Deleted: <#>Adopt a more neutral, analytic tone. In parts, it reads too much like an advocacy document for EE.¶
¶
<#>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.¶

Deleted: <#>Incorporate the findings of more recent scholarly literature, in particular on cognitive biases and elicitation of quantities and on methods for assessing the performance of experts and aggregation of judgments across experts.¶
¶

Deleted: contrast

Deleted: experiment with

research on EE and alternatives 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

Introduction

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.

Deleted: comprehensive

Deleted: reasonable

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 does not provide a critical analysis of the strengths and weaknesses of expert elicitation in contrast with those 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 additional original research (e.g., primary data collection). The strengths and weaknesses of EE could best be understood if presented in comparison with the strengths and weaknesses of other methods. The panel understands that EPA is preparing another white paper on the "Hierarchy of Methods for Characterizing Uncertainty" that will discuss the choice among alternative methods and recommends that the EE white paper explain that a comparative analysis of approaches will be presented in this forthcoming document.

Deleted: reference

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 to characterizing information. EE does not create new 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 to sensitive subgroup, past to future). EE (and other methods for incorporating expert judgment) can be used to address this extrapolation, whereas other methods, such as

Deleted: (WORD MISSING?)

Deleted: ,

Deleted: but

meta-analysis, generally cannot. In short, EE should be presented as a useful way to organize and understand what is already known about a matter and to identify what remains to be studied.

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

2. The white paper could be improved by adopting a more balanced, neutral, analytic tone. In parts, it reads too much like an advocacy document for EE.

3. The white paper should include a 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 certain event will occur. 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 (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. In this case elicitation of the parameter value may be appropriate (even if it is not directly measurable), so long as a model or models can then be used to illustrate the implications of the expert's parameter choices for the measurable output of interest..

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 method for choosing among probabilities held by equally well-qualified experts. EE is a method for eliciting and

Deleted: ever

Deleted:

Deleted: . While the Agency recognizes the importance of understanding the particular scientific mechanism or model (e.g. linear no-threshold dose-response function…) for explaining those outcomes, it recognizes that they may never be known exactly or at least not within the time frame for making a decision.

Deleted: , 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 probabilities of events that are measurable in principle, if not necessarily in practice (e.g., the number of deaths that will occur in a specific year from airborne PM conditional on particular regulations). In other words, EE should be used to elicit

Deleted: can

Deleted: e

Deleted: e

Deleted: individuals (including

Deleted:)

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 could be used to obtain judgments from domain experts or other individuals. The white paper would benefit from greater acknowledgment of this fact, perhaps 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 elicited is critical even when parameter values will be based on literature review and the analyst's own judgment.

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. In addition, some panel members recommend that accuracy and calibration be tested by obtaining judgments for seed quantities, the values of which will become known after the expert provides his distribution.

6. The white paper should reference more recent literature. A list of suggested references by topic is appended to the end of this report.

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 is 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.

Transparency is important to: 1) characterize the range of expertise; 2) identify the experts' rationales for the quantitative judgments (for credibility and to decide when

Deleted: ies

Deleted: (

Deleted:)

Comment [c1]: Tom Wallsten suggests this language:

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

Deleted: The white paper should address methods for evaluating and ensuring the quality of expert judgments including tests for internal consistency, coherence in judgments over multiple factors, and, when possible, performance (measured using quantities the values of which are known or will soon become known)."

Deleted: , including tests for coherence of judgments over multiple factors and performance (tested using judgments for seed quantities, the values of which will become known).

Deleted: biases

Deleted: employer's or other's

Deleted: identify

new understanding renders the results obsolete); 3) evaluate strengths and weaknesses of the study in the future; 4) evaluate and enhance credibility by demonstrating that the approach was applied rigorously; 5) 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 say more about how to capture each expert's assumptions and "basis of judgment." It should also discuss how deepening the interactions between elicitors and experts makes documentation more difficult and expensive. In other words, the report should assess each method in terms of how much of the process is intrinsically a black box, e.g., the extended in-person interviews often used for EE may be the best approach in terms of getting the most accurate estimates, but the interaction between the elicitor and the expert is more difficult to chronicle and may influence the results in ways that are difficult to identify. Other approaches, such as a remotely-conducted Delphi or survey, may provide more transparency about the process, if not necessarily about experts' rationales. It may be useful to create a table that lists the aspects that can be easily conveyed transparently and those that cannot. That is, the white paper should discuss transparency in the context of tradeoffs. In short, the more help that the experts get, the more difficult it is to document and convey all of the interactions.

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, formal, structured or unstructured – and therefore that the guidance offered below applies uniformly – and is not intended to be a critique of formal elicitation of expert opinion.

The panel notes that for an EE study to succeed, the experts selected must be

Deleted: effort
Deleted: is
Deleted: credibility
Deleted: increase
Deleted: terms of
Deleted: should be methodology (how the problems were framed, how judgments were elicited and combined), the distributions provided by the experts, characteristics of the experts, the assumptions they hold, definitions of the quantities being elicited, the relevant quantities upon which the quantity being elicited is being conditioned, and the specific interactions between elicitor and expert.¶ ¶
Deleted: does a good job in urging transparency in methodology, judgments, and characteristics of experts (this does not
Deleted: requiring
Deleted: require identifying which expert provided which specific probabilities). However, the white paper
Deleted: on
Deleted: the
Deleted: "
Deleted: relied on by each expert.
Deleted: It should also discuss how deepening the interactions between elicitors and experts makes it more difficult to be fully transparent.
Deleted: For example, if the elicitor has to point out to the expert that the expert's responses are incoherent (i.e., violate
Deleted: certain
Deleted: particular properties of probability distributions), this interaction and how the expert modifies his or her distributions is very difficult to chronicle.
Deleted: A
Deleted: , in contrast,
Deleted: is much easier to describe with full transparency.
Deleted: would
Deleted: consider
Deleted: ing
Deleted: . Although the white paper makes the correct point that the "basis for judgment" (the experts' assumptions) should be made transparent, it doe ... [1]

credible, the set of experts must be acceptable to stakeholders, and the process for selection should be clearly documented and replicable. To enhance the quality of the results, experts should 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, the values of which become known, can be assessed). To enhance the transparency and credibility of the study, it is helpful if the experts can articulate the basis of for their judgments.

Although it seems intuitive that the set of experts should span the set of reasonable perspectives in the domain, the panel cautions that it is difficult to evaluate this criterion in advance (e.g., to determine whether an outlying perspective is "reasonable"). Some panelists are concerned that in some domains the set of reasonable perspectives may not be adequately represented without including more than ten experts (hence requiring approval from OMB).

With regard to the question of whether EPA should conduct the expert selection process or contract it out (p. 71 of the white paper), the panel notes that contracting out may increase the perception of greater objectivity in cases where EPA is viewed by some to have a prior bias, particularly when the task if delegated to a credible independent entity (e.g., the National Academies of Science). This may be important for certain high-profile, highly-contested issues. Whether EE is conducted by EPA, or by a contractor, the process should be judged on its perceived quality. This may be enhanced by prior review of the EE study design (including public comment) before experts are selected.

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 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 choose 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

Deleted: that are well-calibrated

Deleted: . To enhance the quality of the results, experts should have the ability to characterize their beliefs in terms of probability distributions

Deleted: may be

Deleted: areand be capable of

Deleted: ing

Deleted: To enhance the transparency and credibility of the study, it may be helpful if the experts are capable of articulating the basis of their judgments

Deleted: . The panel suggests that the white paper should distinguish criteria for establishing the pool of experts from criteria for selecting from the pool, given that

Deleted: it

Deleted: the pool is larger than the number of experts required. Another approach to establishing the pool is to invite stakeholders to nominate experts who meet the criteria specified above.¶ ¶

Deleted: ¶

Deleted: . 1. 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.¶

Deleted: 2

Deleted: ; she must choose a single policy

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 discuss performance-based methods (e.g., Cooke's "classical method" 1991) and 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).

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.

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 is acceptable.

The white paper should give more attention to conditionality among the model quantities. Conditionality is important for at least two reasons. First, for experts to provide judgments about the value of some quantity, they must understand which of the factors on which its value is conditional are specified (and their specified values) and which are unspecified. The influence of unspecified factors on the quantity being elicited becomes part of the uncertainty in the value of that quantity. Second, when experts are asked to provide judgments about multiple quantities, dependencies among these quantities may be relevant; i.e., using independent marginal distributions (ignoring correlation) for multiple uncertain parameters in a model can produce misleading outputs. Experience suggests that correlation coefficients should not be directly elicited (see Evans et al., 1994 and Clemens et al., 2000).

Deleted: 3
Deleted: recent
Deleted: FULL CITATIONS NEEDED.
Deleted: parameters
Deleted: parameter
Deleted: quantitiy
Deleted: parameters
Deleted: parameters
Deleted:
Deleted: (CITE EVANS WORK ON CHLOROFORM?). C
Deleted: .
Deleted:

The "clairvoyance test," which "demands that all of the significant assumptions and conditions that could impact the expert's response are well-specified" attempts to capture the first issue, but it 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 experts' judgments about dry deposition velocity (ddv) of aerosols in the EU-USNRC study (CITATION?. It is known that ddv depends on at least 80 physical parameters ranging from the mean free path of Brownian motion to the mixing layer of the atmosphere, but it is not known how it depends on all these. The study in question distinguishes ddvs according to chemical species, surface (e.g., grass, urban, skin), aerodynamic diameter, and wind speed. An expert is not asked to build a model for ddv, but is asked about a potentially measurable quantity conditional on others, e.g., "Suppose we measure the ddv of aerosols with aerodynamic diameter of 1 μm on grass with wind speed 2 m/s. Please provide a probability distribution for the result we will obtain."

The expert is asked to conditionalize his uncertainty in a way that is conformable to the model to which his judgment will be input. It is known that ddv can vary by an order of magnitude according to the species of grass. The expert is not told the species of grass, rather he is told that uncertainty arising from this factor should be "folded into his distribution." Similarly, many other variables may be important. The omniscient being implied by the clairvoyance test presumably would know the values of these, but neither the analyst nor the expert does. Maintaining a consistent conditionalization across a large study is critical (and difficult to accomplish). For example, when eliciting judgments about atmospheric dispersion and wet deposition, the conditionalization must be consistent with that for dry deposition.

Problem structure and consistent conditionalization may be facilitated by use of an influence diagram. The influence diagram illustrated in Figure 6.1 of the white paper should be replaced with an improved example that is adequately labeled so that it can be understood without reference to additional text.

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; while scenario design may affect experts' judgments about quantities (because the quantity may be conditional on factors that are specified by the scenario), EE is not an appropriate tool for obtaining expert judgment about analytic 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.

Charge question C.4 & 5 – findings and recommendations

Deleted:

Deleted: (e.g., electrostatic potential, humidity, temperature, surface roughness, insolation, wind profile)

Deleted: do

Comment [c2]: Word missing

Deleted: The need to explicitly maintain a consistent conditionalization across multiple factors is easily neglected when thinking about 'clairvoyance."

Deleted: The problem of maintaining a consistent conditionalization across multiple factors is easily neglected when thing about "clairvoyance."

Deleted: Problem structure and consistent conditionalization may be facilitated by use of an influence diagram.

Deleted: use of an influence diagram such as the one illustrated in Figure 6.1 of the white paper. (The panel recommends that all figures be adequately labeled so they can be understood without reference to additional text; Fig. 6.1 does not meet this standard.)

Deleted: ;

Deleted:

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. As noted above, the tone of the white paper and the conclusions is too much one of advocacy for EE rather than a balanced review of its advantages and disadvantages, and comparison with other approaches.

An important topic that receives little attention in the white paper is that of the consistency of judgments from a single 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). This raises the question of the extent to which experts' judgments should be subjected to consistency tests and how violations of these tests should be rectified. Experts should be made aware of these violations and asked to adjust their distributions, though success in satisfying multiple consistency tests may require some guidance from the elicitor or others. A danger when there is extensive interaction of this type is that transparency of the process is compromised and the experts' reports may be significantly influenced by the elicitor.

The literature on cognitive biases is much richer than is indicated in the white paper. In addition to well-known estimation biases such as anchoring and availability heuristics, 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 heuristics 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 a glossary of key terms with practical definitions. A suggested list of terms is attached.

Charge question D – development of future guidance

Deleted: his

Deleted: he

Deleted: his knowledge about

Deleted: multiple known

Deleted: Presumably

Deleted: e

Deleted:

Deleted: REFERENCES

Deleted:)

Deleted: in detai

Deleted: S

Deleted: trategies to elicit expert judgments beyond familiarizing them with some of these biases merit some discussion.

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. Considerable experience with structured expert judgment exists in other fields, including nuclear, aerospace, volcanology, health, environmental transport, and finance. The challenge is to bring this experience to bear on the specific problem areas within EPAs mandate. 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 on 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."

Deleted: that

Deleted: that is should

Deleted: in

Deleted: It

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

Accurate
Aggregation
Assumption
Assumptions
Availability
Averaging
Bias
Conditional Probability
Data gap
Data quality
Decision options
Dependence
Domain expert
Elicitation
Elicitor

Encoding
Estimates
Event
Extrapolation
Heuristics
Input
Model
Model choice
Objective
Overconfidence
Paradigm
Parameter
Precision
Quality
Quantity
Relationship
Representativeness
Robust
Subjective
Subjective Judgment (?)
Subjective Probability
Weighting

Suggested additional references for inclusion in a revised White Paper

[should be added to white paper and discussed, organized by topic]
EE studies (EU-USNRC)
EE guidance documents
Reliability engineering journal special issue
Cognitive biases in estimating quantities and probabilities
Other

References

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

Clemen, R.T., and R.L. Winkler (1999), Combining Probability Distributions from Experts in Risk Analysis, *Risk Analysis* **19**, 187-203.

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., *Experts in Uncertainty*, Oxford University Press, 1991.

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.

Hsu, M., M. Bhatt, R. Adolphs, D. Tranel, and C.F. Camerer. 2005. Neural systems responding to degrees of uncertainty in human decision-making. Science 310:1680-1683.

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

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

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

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

O'Hagan, Anthony Caitlin E. Buck, Alireza Daneshkhah, J. Richard Eiser, Paul H. Garthwaite, David J. Jenkinson, Jeremy E. Oakley, and Tim Rakow. 2006. *Uncertain Judgements: Eliciting Experts' Probabilities* .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.

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

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.

Page 7: [1] Deleted JKH 4/21/2009 5:32:00 PM

Although the white paper makes the correct point that the "basis for judgment" (the experts' assumptions) should be made transparent, it does not explain how this can be done under conditions of fairly limited resources. It should also be emphasized that putting huge resources into chronicling the interactions between elicitor and expert may not be worthwhile in light of the impossibility, in many cases, to be fully transparent. The suggestions for documentiation seem extreme in terms of time and costs involved.