INTERNATIONAL MARITIME ORGANIZATION

E

MARITIME SAFETY COMMITTEE
89th session
Agenda item 16

MSC 89/INF.12
7 March 2011
ENGLISH ONLY

FORMAL SAFETY ASSESSMENT

Information on expert elicitation

Submitted by the United States

SUMMARY

Executive summary:	This document contains information on the use of expert judgment and expert elicitation. Formal Safety Assessments (FSAs) often rely upon expert judgment when there is a shortage of historical data. The United States believes that this information could be useful to member governments and/or organizations that plan on conducting a FSA or are currently working on one.
Strategic direction:	12.1
High-level action:	12.1.1
Planned output:	12.1.1.1
Action to be taken:	Paragraph 47
Related documents:	MSC 89/16; MSC 83/INF.2; MSC 87/WP.7; MSC 87/18; and MSC/Circ.1023-MEPC/Circ.392

Background

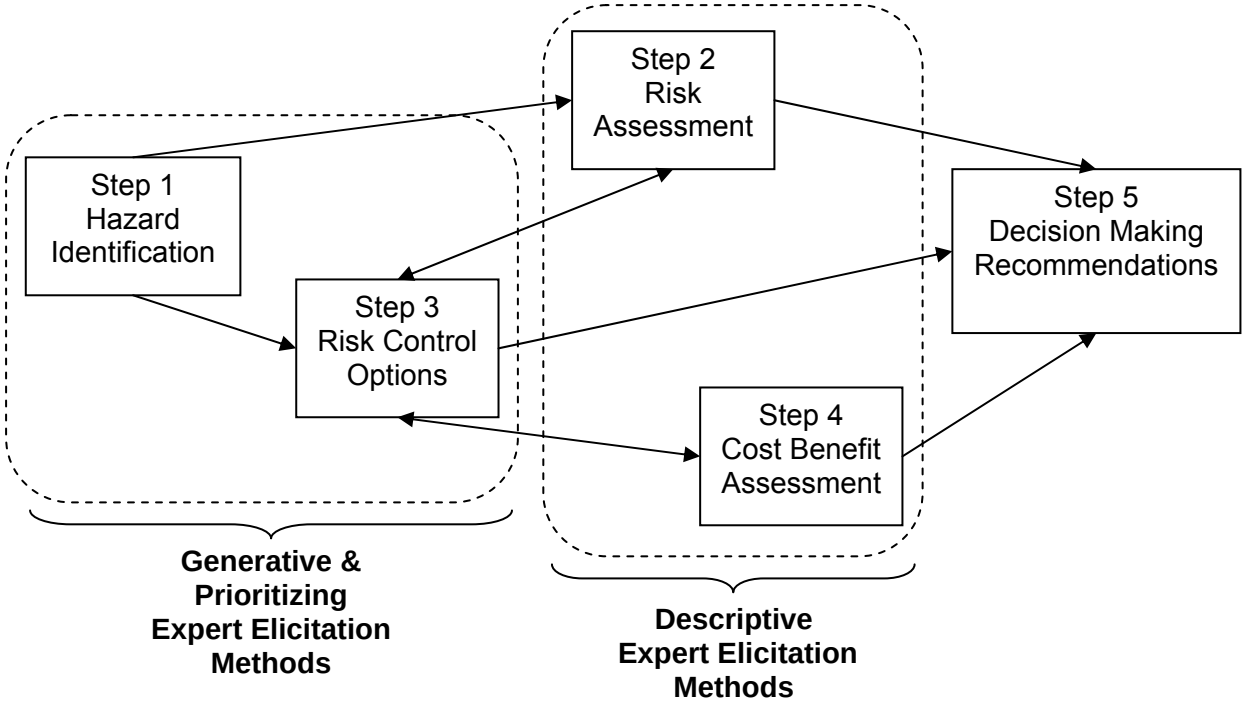
1 Earlier this year the FSA Correspondence group deliberated on the terms of reference established by the Marine Safety Committee (MSC) at its eighty-seventh session. As a participating member of the correspondence group, the United States provided suggestions and comments on areas listed in document MSC 87/17, paragraph 49, specifically concerning items 1, 4, 5, 6, 9, and 11. This work, and the resulting discussions, initiated a research project by the United States to expand our knowledge on expert judgment and expert elicitation.

2 While conducting this research, it became apparent that there was not sufficient guidance on expert elicitation in the FSA Guidelines (MSC 83/INF.2). During the FSA process, relevant and/or accurate historical data/information is often difficult to locate. In such cases, expert judgment must be used to evaluate the potential hazards present in a given system. Without historical data, expert judgment is often the only source of information and has a vital role in the FSA methodology.

Discussion

3 It is inherently difficult to conduct a risk analysis when there is a scarcity of historic or experimental data. In the absence of such data, organizations often turn to experts to provide some basis upon which to conduct analyses. Expert judgment can be defined as information about a phenomenon, event, quantity, or parameter of interest (e.g., probability of hazard scenario occurrence) that is based upon the experience and knowledge of the expert. While expert judgment presents an effective means by which to close gaps in historical or experimental data, it can have some considerable limitations. Unless judgments made by experts are systematically elicited and use a well-defined process that addresses potential cognitive and motivational biases, the elicited expert judgment may be inaccurate or biased, leading to potentially misleading information.

4 As has been demonstrated, FSAs often rely upon expert judgment as an important source of evidence or data. The figure below describes the FSA methodology flow chart (slightly rearranged, but with the same dependencies). Two distinct functional categories of expert elicitation applicable to specific steps of the FSA have been indicated below.



.1 **Generative Expert Elicitation Methods:** For steps 1 (Hazard Identification) and 3 (Risk Control Options) of the FSA methodology, the following are some common methods employed for such activities:

.1 brainstorming;
.2 nominal group; and
.3 technique scenario planning

.2 **Prioritizing Expert Elicitation Methods:** Sometimes the FSA methodology calls for the prioritization of the results from steps 1 (Hazard Identification) and 3 (Risk Control Options). The following are some of the common methods employed for such activities:

.1 nominal group technique;
.2 multi-voting;
.3 paired comparison (e.g., Analytic Hierarchy Process);
.4 order ranking scales; and
.5 Policy Delphi Method

.3 **Descriptive Expert Elicitation Methods:** Alternatively, the results of steps 1 and 3 can be prioritized through description. Once a list of hazards has been identified in step 1 of the FSA, then these hazards are quantified (or described) using risk assessment methods in step 2 of the FSA. This description allows prioritization of the risks to determine where additional resources or regulatory schemes may be necessary. Likewise, once a list of risk control options has been identified in step 3 of the FSA, then these options are quantified (or described) using cost-benefit assessment methods in step 4 of the FSA. Again, this description allows prioritization of the risk control options to determine which are most cost-effective. For steps 2 (Risk Assessment) and 4 (cost benefit assessment) of the FSA methodology, expert elicitations taken from the top half of the taxonomy described in Appendix 9 are employed to both describe or assess risk and cost benefit. The following are some of the common methods employed for such activities:

.1 direct elicitation (point or distribution);
.2 paired comparison method (e.g., analytic hierarchy process, Bradley Terry model, Thurstone model); and
.3 discrete or continuous scaling methods

5 Generally, most expert elicitations follow a similar process, with anywhere from four to eight steps. For the purposes of this document the United States has chosen to review a generic expert judgment elicitation process that uses the following seven steps:[1]

STEP 1: FRAME THE PROBLEM

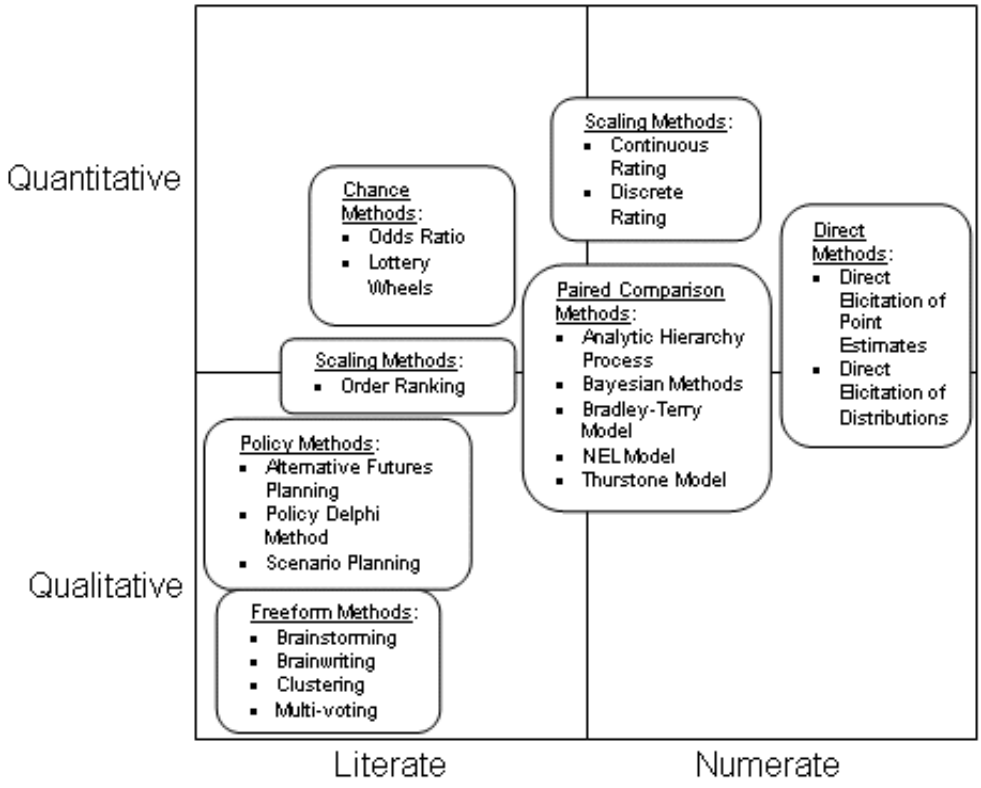
6 The majority of this scoping will have already occurred in the problem definition stage of the FSA methodology. Additionally, it may be appropriate to create a problem statement for the purposes of conducting the expert elicitation. Additionally, the problem may need further decomposition to enable effective expert elicitation.

STEP 2: PLAN THE ELICITATION

7 The nature of the expert elicitation is affected by scope, schedule, and budget. In the planning and execution stages of the expert elicitation, it is useful to treat the elicitation like a project and apply project management techniques. A plan of action should be developed (with time and cost estimates, as well as responsibility assignments, for each step in the process). The schedule should allow sufficient time for the design and testing of the elicitation protocol, the time needed to gather and train the experts (often there are issues around expert availability), the time needed to gather and disseminate baseline reports, the time needed for the actual elicitation (including any requirements for iterative or interactive processes), and the time needed to administer, document, and communicate the elicitation processes.

[1] Cooke, R. & Goossens, L. (2000) Procedures guide for structured expert judgment, EUR 18820 EN describes a 15-step process.

8 In order to facilitate planning, it will also be necessary to determine which type of elicitation method is most appropriate. The choice of elicitation method will depend upon both the type of data or information desired and the type of expertise available. There are a wide variety of methods by which to elicit expert judgments. The following figure is taxonomy of the various expert elicitation methods.



9 There are five general classes of elicitation methods (i.e. direct elicitation methods, paired comparison methods, scaling methods, chance methods, and other methods). The classes have been placed in the taxonomy based upon the necessary input expertise and the desired output form of the data.

10 First, determine the type or form of data that is sought. The desired form of the data roughly corresponds to Steven's scales of measurement[2] – nominal and ordinal data being primarily qualitative and interval and ratio data being predominantly quantitative data.

11 Next, determine the type of expertise that is available or necessary. "Numerate" experts have facility with and possess the ability to discuss and describe quantities, probabilities and numbers. "Literate" experts possess the ability to discuss and describe qualities using words and narrative. Some experts may be both numerate and literate. Additionally, in some contexts, experts may be more comfortable providing relative estimates. In such cases, paired comparison methods and scaling methods using order ranking would be appropriate. When experts are capable of providing absolute estimates, direct methods and scaling methods using discrete or continuous ratings would be appropriate.

[2] Stevens, S. (1946). On the theory of scales of measurement. *Science*, 103(2684), pp. 677-680.

12 Since much of expert elicitation used in FSAs concerns developing estimates of probability, consequence, cost, or benefit probability estimates (i.e. quantitative information), the following table provides a comparison of the three basic types of expert elicitation methods used in this manner.[3,4]

Method	Description	Advantages	Disadvantages
Scaling Methods	Experts are asked to provide judgments using some form of linguistic-numeric scale.	Experts find ratings (and rankings) relatively easy to use. If redundant ratings are used, measures of concordance may be calculated to determine expert agreement. For *n* scenarios, this method requires *n* estimates (or 3**n* estimates if 5%, 50%, and 95% are desired to capture uncertainty).	Most scales are limited to five or seven categories due to human cognitive capacity. Scales may not effectively capture desired orders of magnitude. There is limited evaluation of scale validity and reliability.
Paired Comparison Methods (also known as Conjoint Analysis)	Experts are asked to provide judgments by comparing pairs of hazards and comparing each pair using some discriminator.	Experts are capable of reliably discriminating differences between pairs of scenarios, etc. Several of these methods provide estimates of expert's internal consistency. This method allows experts to provide relative comparisons between scenarios, rather than providing absolute judgments.	Requires considerable time to compare all possible pairs (for n different scenarios, there will be between (n-1) and n*(n-1)/2 comparisons). However, there are ways to incorporate less than complete sets of comparisons.
Direct Elicitation Methods	Experts are asked to provide judgments directly.	Provides point estimate or distribution, which can be directly used in risk analysis. For *n* scenarios, this method requires *n* estimates (or 3**n* estimates if 5%, 50%, and 95% are desired to capture uncertainty). (Delphi method allows anonymous and iterative process that avoids negative group dynamics.)	Experts may have difficulty estimating rare events due to cognitive bias toward underestimation. (Delphi method is limited in the number of judgments that can effectively be elicited, does not allow group interaction, and may not motivate full participation.)

13 Knowing the desired form of the expert judgment data, the type of expertise available, as well as the advantages and disadvantages of the various methods, a specific expert elicitation method may be chosen.

[3] Dalton et al. (2009) provide a framework for comparing between and selecting from among four different methods of expert elicitation.

[4] Meyer and Booker (2001) provide an extensive description of the various components of elicitation (elicitation techniques, modes of communication, situations, response modes, and aggregation schemes), including the advantages and disadvantages of each. They also include a more extensive listing of pertinent literature about these subjects.

14 Additionally, the mode of the expert judgments can vary regardless of which method is chosen. For example, depending upon the desired amount of interaction; the Delphi method, the Nominal Group technique, Consensus techniques may be used. The Delphi method is typically iterative, but indirectly interactive – experts do not meet face-to-face, but there is anonymous communication among experts. The Nominal Group technique requires independent estimates to be developed prior to expert interaction. Consensus group processes often have expert interaction immediately from the beginning of the elicitation.

15 Once an expert elicitation method has been chosen, an elicitation protocol should be developed and tested. This should be viewed as a proof of process to ensure it achieves the intended results. The protocol should consist of the instrument that will be provided to the experts to elicit their judgments (i.e. survey, questionnaire, etc.), instructions for the experts, and instructions for the facilitator or risk practitioner conducting the elicitation. It is important that the protocol be tested prior to the actual elicitation (in step 5) because there will invariably be areas for improvement (e.g., reduction of the number of questions to avoid expert fatigue, rewording of questions and instructions to help avoid bias, clarifying questions to provide better understanding). It is common practice to test the protocol on colleagues and peers (rather than the experts themselves).

16 One area of particular importance is to ensure the protocol is designed to minimize motivational and cognitive biases. There are essentially two families of cognitive operations:[5] intuitive (or system 1) and reflective (or system 2). System 1 is more primitive (but not less capable) and is characterized as automatic, effortless, and rapid. System 2 is characterized as controlled, effortful, deductive, and slow. Expert judgment belongs to system 2. In risk analysis and assessment, intuitive answers are not readily available and we tend to rely on expert judgment to provide reasoning and rationale upon which to base decisions.

17 However, despite the fact that expert judgment is intentional and more time-consuming, it is still subject to cognitive biases. Additionally, experts may be directly or indirectly motivated to express something other than their true response to the elicitation. Cognitive and motivational biases and heuristics can also be controlled for (or at least mitigated) by careful expert training and elicitation protocol design. There are many types of cognitive and motivational biases (e.g., anchoring and adjustment, availability, framing). The nature of these biases is beyond the scope of this document, but ample literature exists to provide reference and guidance.[6]

18 While it would be important to be able to evaluate each expert's performance (as described in step 5), it is not always possible. In such cases, efforts should be made during the design stage to ensure the expert elicitation process adheres to the principles of rational consensus[7] as much as possible. The following hierarchy of desirable features of the elicitation process has been provided.

---

5 Kahneman, D. & Frederick, S. (2001). Representativeness revisited: Attribute substitution in intuitive judgment. In Gilovich, T.; Griffen, D.; & Kahneman, D. (Eds.), *Heuristics of intuitive judgment: Extensions and applications.* New York: Cambridge University Press.

6 See any of the many excellent edited books and articles on the subject of heuristics and biases – see for example:
Tversky, A. & Kahneman, D. ((1974). Judgments under uncertainty: Heuristics and biases. *Science*, 185, pp. 1124-1131.
Kahneman, D. & Tversky, A. (Eds.) (2000). *Choices, values and frames*. Cambridge: Cambridge University Press.
Kahneman, D.; Slovic, P.; & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
Gilovich, T.; Griffin D. & Kahneman, D. (Eds.) (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge: Cambridge University Press.

7 See pp. 80-86 in Cooke, R. (1991). *Experts in uncertainty: Opinion and subjective probability in science*. Oxford: Oxford University Press.

Principle	Description
Fairness	All experts must be treated fairly (prior to, during, and after the elicitation) such that discrimination on expert performance may occur.
Empirical Control	Elicited judgments must be subject to empirical control (in principle) by providing means of evaluating expert performance (e.g., accuracy and precision).
Neutrality	The method for eliciting, combining and aggregating expert judgments should motivate unbiased responses.
Reproducibility	Scientific peers should be able to review and reproduce the results of the expert elicitation and risk analysis (e.g., data and models should be fully specified).
Accountability	As a form of evidence or data, the source and rationale for the expert judgment should be provided and documented.

19 The way in which this hierarchy works is such that each successive level builds upon the previous. For example, accountability would be the first level of quality to achieve within an expert judgment elicitation process. Next, reproducibility would be added to accountability. The first three principles (accountability, reproducibility, and neutrality) can be applied almost regardless of the context. The remaining two principles (empirical control and fairness) are much more difficult (but not impossible) to attain, particularly in rare event situations. Finally, after accountability, reproducibility, neutrality, and empirical control, fairness would be layered onto the elicitation process. It may seem odd that fairness would be the last element to consider, but until all of the others have been included (especially empirical control), fairness cannot be included. Therefore, in cases where expert performance cannot be evaluated directly, expert elicitations that did not include any of these features would be considered inferior to those that adhered to some or all of these principles.

20 Finally, once the desired timeliness, accuracy/quality, and form of the expert judgments have been determined, these requirements will dictate schedule and budget. In project management, the triple constraint is the trade space between schedule, budget, and scope (or quality), such that changes to any one of these constraints will impact the others. So, if it is determined to conduct an expert elicitation rapidly and without budget, the results will likely be of limited value due to the inherent inability to produce quality expert judgments under these conditions.

STEP 3: SELECT THE EXPERTS[8]

21 It is extremely important to identify the requisite expertise that determines whom is an expert. There are several factors to consider when selecting experts. Experts selected for the elicitation should posses both a high professional standing and widely recognized competence.

22 Evidence of relevant experience and training include the following:

.1 certifications such as academic degrees or professional training;

.2 professional reputation of the expert (as a potentially reliable guide);

.3 impartiality; and

[8] Experts, also known as subject matter experts (or SMEs), are people who have a background in the subject area (including training and experience) and are recognized by their peers or those conducting the risk analysis as qualified to answer the questions posed.

.4 multiplicity of viewpoints (i.e. consideration of multiple forms of data and perspectives).[9]

23 Experts should represent a diversity of technical perspectives on the issue of concern. Experts should be willing to be identified publically (although exact identification with judgments may be withheld except for competent peer review),[10] provide their rationale supporting their judgments, and disclose any potential conflicts of interest. Additionally, experts must be willing to devote the requisite time for completing the elicitation and study.

24 The number of experts selected depends upon the nature of decision context and the nature of the problem, including the degree of uncertainty expected. As a general rule of thumb, six to eight experts (and no fewer than four) should be obtained and at least some of the experts should be from outside of the organization conducting the FSA. A pool of candidate experts (who posses requisite expertise and have demonstrated interest and commitment to participate) should be reviewed by a committee and a sufficient number of the best experts will be selected from that pool.

STEP 4: TRAIN THE EXPERTS

25 There are several reasons for conducting pre-elicitation training of experts:

.1 familiarize the experts to the problem under consideration and ensure they share the same baseline of information (e.g., basic domain knowledge or probabilistic and uncertainty training);

.2 introduce the experts to the elicitation protocol, procedure, and process;

.3 introduce or reinforce the experts on uncertainty and probability encoding and provide them practice in formally articulating their judgments and rationale; and

.4 provide awareness of the potential for cognitive biases that may influence their judgments.

26 If possible, it is desirable for the experts to share a common understanding of exactly what information is being elicited. While experts will approach the elicitation in a variety of differing perspectives based upon their diversity of training and experience, it is paramount that they all address the same problem as posed by the elicitation. This can be accomplished through pre-elicitation training. Also, it is important to allow the experts to gain experience with the elicitation protocol (i.e. the questionnaire, survey, interview, etc.) in advance of the actual elicitation so that when it comes time to provide their judgments, they are consistently supplied.

[9] Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes*, 53, pp. 252- 266.

[10] Cooke and Goossens (2000) recommend the following procedure: 1. Publish expert names and affiliations in the study, 2. Retain all information (including identified judgments) for competent peer review, but not unrestricted distribution, 3. Allow de-identified judgments to be available for unrestricted distribution, 4. Document and supply rationales for all judgments, 5. Provide each expert feedback on their own performance, 6. Request expert permission for any published use beyond above.

27 Pre-elicitation training may also include tuning expert numeracy – for example, many experts are not familiar with describing their degrees of belief and uncertainty in terms of quantiles (e.g., 5%, 50%, 95%). Allowing all experts to participate in a group training session provides each the benefit of hearing the others' questions (and responses) and ensures all have a common understanding of what will be asked of them.

STEP 5: ELICIT JUDGMENTS FROM THE EXPERTS

28 The elicitation session is at the heart of the expert elicitation process and represents the execution phase of the plan – hopefully, a well-designed plan is well executed. Generally, elicitation sessions should provide an:

.1 orientation of the problem to be confronted;

.2 description of logistics and instructions for the experts (including allowing for questions and answers, as well as agreement upon definitions and assumptions);

.3 the elicitation itself (where care should be taken to ensure that each expert is queried in a uniform manner); and

.4 debriefing (to gain feedback from the experts on the elicitation process and to provide experts feedback on their performance in as much detail as possible and appropriate). Frequently, the elicitation sessions are "live" in that both expert and analyst are present when the actual judgments are rendered. However, this does not necessarily have to be the case – one example that does not require direct interaction is the Delphi method.

29 Ideally, in situations where the experts are present during the elicitation, both normative and subjective analysts should be on hand to facilitate the elicitation. The normative analyst will be experienced with subjective probabilities and have experience in expert elicitation. This person will be able to address procedural issues that may arise. The subjective analyst will be experienced with the field or domain of interest (i.e. maritime safety or environmental protection) and may have been involved in framing the problem and in developing the specific questions. This person will be able to clarify questions experts have about the nuances of particular elicitation scenarios. Occasionally, a single analyst may be able to serve both roles.

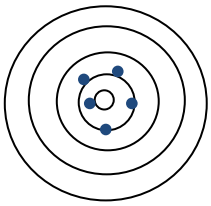
30 The elicitation session is where the experts are asked for their independent judgments. The mode of the elicitation can take many forms, but it is important to capture individual judgments prior to any interactive portion (if deemed appropriate). Experts should be directed to provide their rationale (i.e. sources of information, any modeling or analysis or calculations that may have occurred, as well as lines of reasoning). While the expert judgments are an important aspect of expert elicitation, the rationale by which those judgments were arrived at can be equally important. Depending upon the circumstances, it may be advantageous to allow experts time in advance of the elicitation to develop their rationale. Additionally, in some cases, it may be deemed appropriate to bring experts together to be presented with collective or de-identified individual judgments so that experts can interactively and collaboratively defend and improve their judgments. This may allow for a wider sharing of information and rationale, as well as an opening of perspectives. However, it may also create consensus (where none existed) due to groupthink or dominant participants.

31 Expert fatigue results when experts remain on task too long during the elicitation. Another sort of expert fatigue can result when experts are elicited too frequently (and they are not sufficiently motivated to fully participate). When the elicitation session runs too long, the quality of the expert judgments will eventually be diminished over time through an excessive cognitive load. Also, experts may experience boredom and disinterest if the elicitation instrument becomes routine and repetitive. Ideally, expert elicitation sessions should be kept to two to three hours in duration. Every attempt should be made to prevent expert elicitation sessions from exceeding three or four hours in duration (per elicitation session). If additional elicitation is necessary, breaks should be provided to allow rest and restoration of cognitive abilities. Expert fatigue should be kept in mind during the planning step.

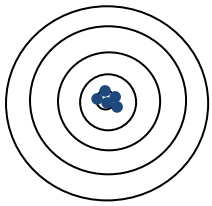
STEP 6: ANALYSE (AND AGGREGATE) THE JUDGMENTS

32 Steps should be taken to assess the performance of experts whenever possible. Evaluating expert performance is necessary in order to objectively evaluate the quality of the "experiential insight" of each particular expert. Typically, the phenomenon of interest is unknown when the risk assessment is performed and expert judgments are elicited. However, one method for judging the performance of the experts introduces additional "seed" variables of interest where the true value is known (e.g., in cases where historic data may be available or in cases where additional reports were unavailable to the experts). The experts' judgments regarding the "seed" variables are then compared to the true values and this will form an indication of the experts' performance. Measuring expert performance is important because it can be used as a means of combining expert judgments (i.e. higher weighting for superior performance) and it serves to enhance the credibility of the study.

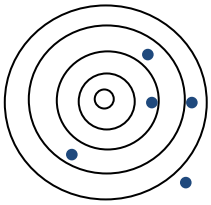
33 To evaluate the performance of an expert's judgment or opinion against a true or known value, one can relate these to the notions of precision and accuracy. The figure below describes the various combinations of accuracy and precision.



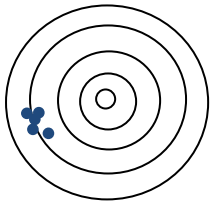
Low Precision, High Accuracy



High Precision, High Accuracy



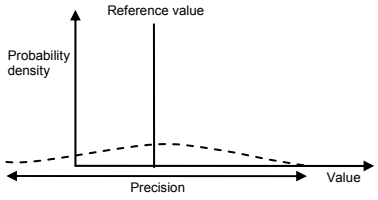
Low Precision, Low Accuracy



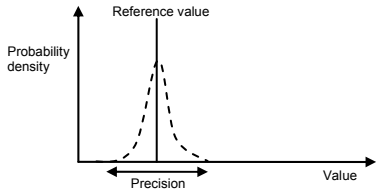
High Precision, Low Accuracy

34 Generally, accuracy relates to the closeness of the judgment to the true value (i.e. the "bulls eye" or center of the target); whereas, precision relates to the repeatability or reproducibility of the judgments under similar circumstances and states of knowledge (i.e. the tightness of the "cluster").

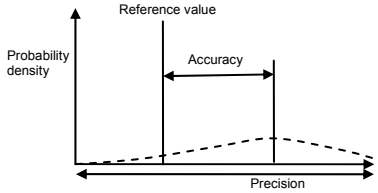
35 Expressed in a slightly different manner, since experts are rarely certain of their judgments, we often express their judgments as a distribution where their accuracy is reflected by the difference in the central tendency of the judgment (e.g., the mean) and the true value (e.g., the seed variable). In this representation, precision would be represented as the variance or spread of the judgment distribution (e.g., the standard deviation). The following four figures provide the different combinations of accuracy and precision an expert may exhibit.



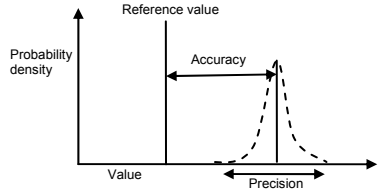
Low Precision, High Accuracy



High Precision, High Accuracy



Low Precision, Low Accuracy



High Precision, Low Accuracy

36 Ultimately, experts that demonstrate both high accuracy and high precision are sought; however, in order to employ the best experts, analysts must consider a means of evaluating their performance.

37 There are several options for combining expert judgments:

.1 intuitive methods;

.2 mathematical methods; and

.3 behavioural methods.

38 For a complete review of many of the possible options, you are encouraged to read one of the survey documents.[11, 12, 13]

Intuitive Aggregation

39 Intuitive methods are when the analysts allow the decision makers to observe all expert judgments (be they point estimates or probability distributions) and then intuitively combine them into a single estimate. Some might consider this the absence of an aggregation method, but it allows the decision maker the opportunity to view expert judgments separately (including disagreements and differences). It also allows the decision maker considerable flexibility.

11 Clemen, R.T. and R. L. Winkler (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis* 19(2), pp.187–203.

12 Genest, C. and J.V. Zidek (1986). Combining probability distributions: a critique and annotated bibliography. *Statistical Science* 1, pp. 114-148.

13 Hora, S. (2010). An analytic method for evaluating the performance of aggregation rules for probability densities. Operations Research, 58(5), pp. 1440-1449.

Mathematical Aggregation

40 Mathematical methods for aggregating expert judgments involve an analytical process or model to combine the judgments of individual experts into a single judgment for the group of experts. One of the simplest examples of mathematical aggregation is to take the arithmetic mean (sometimes called the linear opinion pool) of the expert judgments. This is represented as:

$$p(\theta) = \sum_{i=1}^{n} w_i * p_i(\theta)$$

where n is the number of experts, $p_i(q)$ is expert i's judgment about the probability of phenomenon θ, w_i is the weighting applied to expert i, and the $p(\theta)$ is the combined probability distribution, and weights sum to 1, $(\sum_{i=1}^{n} w_i = 1)$.

41 One of the most challenging considerations in mathematical aggregation is that of assigning appropriate weights to experts. In the arithmetic averaging method given above, the weights are often assumed to be equalitarian (i.e. all experts have equal weights, $\frac{1}{n}$). This may be reasonable when there is reason to believe all experts possess the same expertise or, alternatively, if there is no basis to make any assessment as to the expertise of individual experts. However, if there is a means by which to assess expert performance, scoring rules may form the basis by which expert weights are determined by pooling expert judgments (i.e. the judgments of experts who demonstrate greater accuracy will be weighted more heavily than experts with lesser accuracy).

Behavioural Aggregation

42 Behavioural methods comprise the third set of approaches for combining expert judgments. Behavioural methods require that the experts interact (e.g., directly face-to-face, using some asynchronous means, etc.) in an effort to either arrive at some form of agreement or consensus, thus a behavioural aggregation, or to share information and rationale to improve each expert's individual judgment.

43 There is some evidence to suggest that mathematical approaches outperform behavioural approaches and that both of these are preferable to the intuitive approaches due to inherent biases and cognitive limitations that surround intuitive aggregation. Additionally, there is an emerging body of literature that states simpler methods are preferred over more complex methods – for performance reasons, as well as for ease of use and defensibility reasons, particularly in public policy settings.

44 Regardless, the selection of aggregation method should occur prior to the elicitation so that the analyst does not introduce a form of selection bias (e.g., selecting the aggregation method that most favors expected or desired results). Therefore, selection of aggregation method (whether intuitive, mathematical, or behavioural) should be performed during the planning stages.

45 Once aggregated (or not), the expert judgments should be packaged in a form consumable by the decision maker. Naturally, this will depend on the literacy and numeracy of the decision maker. Often, expert judgments will characterize not only the expert's understanding of the available body of evidence (i.e. what they "know"), but will also characterize their uncertainty about that particular judgment (i.e. what they "do not know"). It is important to capture both aspects of expert judgment (i.e. knowledge about subject as well as epistemic/aleatory uncertainty).

STEP 7: DOCUMENT AND COMMUNICATE THE RESULTS

46 Given that most FSAs will be subject to peer review, it is extremely important that the expert elicitation process as well as the judgments themselves be thoroughly documented. The following is a step-by-step listing of documentation that should be included:

.1 **Frame the problem:** A concise problem statement should be provided.

.2 **Plan the elicitation:** A project plan and an elicitation protocol (with instructions) should be developed.

.3 **Select the experts:** A list of experts, their affiliations, and curriculum vitae (to document their expertise) as well as any conflict of interest statements, participation agreements, and/or compensation agreements should be provided.

.4 **Train the experts:** Typical information to include would be lesson plan(s), copies of training materials, a list of attendees, and a summary of questions that arose or feedback that was provided to improve process.

.5 **Elicit judgments from the experts:** During the expert elicitation, all expert judgments and the rationale for those expert judgments should be collected. Additionally, notes about the elicitation process and information about expert performance should be captured.

.6 **Analyse and aggregate the judgments:** A description of the aggregation method, documentation of any behavioural expert interaction, and the aggregated expert judgment should all be captured.

.7 **Document the expert elicitation:** Documentation is necessary to allow for a historical record. The documentation of the expert elicitation should allow any reviewer to reconstruct the logic and outcomes of the expert elicitation. This also helps ensure transparency. Documentation should occur at all seven steps of the expert elicitation process (as described above) and compiled for future reference.

Action requested of the Committee

47 The Committee is invited to note the information set out in this document.