Process Manual

Elicitation Tool

© Australian Centre of Excellence for Risk Analysis
Contents

1. Reading this manual .................................................................................................................. 5
  1.1. Who should read this manual ............................................................................................ 5
    1.1.1. Workshop facilitators .................................................................................................. 5
    1.1.2. Domain experts ......................................................................................................... 5
  1.2. How to read this manual ..................................................................................................... 5
    1.2.1. I know what I’m doing and just want to get started ..................................................... 5
    1.2.2. I’m a bit unsure and need some background information first .................................... 5
  1.3. Feedback on this document ................................................................................................. 6

2. Why conduct an elicitation workshop ....................................................................................... 7

3. Key concepts ............................................................................................................................. 8
  3.1. Interval estimates ................................................................................................................ 8
    3.1.1. Overconfidence and the 4-step question format .......................................................... 8
    3.1.2. Interval concepts ....................................................................................................... 9
  3.2. Different sources of uncertainty ......................................................................................... 9
    3.2.1. Uncertainty due to language ....................................................................................... 9
    3.2.2. Uncertainty due to lack of knowledge ....................................................................... 10
    3.2.3. Uncertainty due to natural variation ......................................................................... 11
  3.3. DELPHI workshop format .................................................................................................. 11
    3.3.1. Anchoring .................................................................................................................. 11
    3.3.2. The Halo effect ......................................................................................................... 12
  3.4. Frequentist question formats ............................................................................................. 12
  3.5. Process overview ............................................................................................................... 13
  3.6. Typical workshop agenda .................................................................................................. 13

4. Before the elicitation ................................................................................................................. 15
  4.1. Defining the problem ......................................................................................................... 15
    4.1.1. As a logical series of events ....................................................................................... 15
    4.1.2. As a network of related nodes ................................................................................. 16
4.2. Identifying participants ........................................................................................................... 17

4.3. Deciding on a workshop format ............................................................................................. 18

4.3.1. Small groups (<10) in the same room .............................................................................. 18

4.3.1. Larger groups (≥ 10) in the same room ............................................................................. 19

4.3.2. Virtual panel: Participants (<10) in different locations, connected via real-time conference call ......................................................................................................................... 19

4.3.3. Virtual panel: Participants in different locations, connected via email and responding independently ........................................................................................................................................ 20

4.4. Writing and framing elicitation questions ................................................................................. 21

4.4.1. Eliciting information regarding categorical data .................................................................. 22

4.5. Identifying calibration questions ............................................................................................. 22

4.6. Augmenting ET with additional questionnaires ...................................................................... 23

4.6.1. Collecting information not handled by ET ....................................................................... 23

4.6.2. When participants’ cannot easily access ET ...................................................................... 23

4.6.3. When initial estimates are gathered prior to the elicitation workshop .............................. 23

5. During the elicitation .................................................................................................................... 25

5.1. Setting the ground rules ......................................................................................................... 25

5.2. Participant registration and familiarization ............................................................................. 26

5.3. Emphasise that the best estimate is the median value .............................................................. 26

5.4. Integrating calibration questions into the elicitation ................................................................. 26

5.4.1. _Before the workshop_ .................................................................................................... 27

5.4.2. At the beginning of the workshop ..................................................................................... 27

5.4.3. During the workshop ........................................................................................................ 27

5.5. Establishing a shared understanding of each question ........................................................... 28

5.5.1. Linguistic interpretation ..................................................................................................... 28

5.5.2. Underlying assumptions .................................................................................................... 28

5.5.3. Interpretation of estimates .................................................................................................. 28

5.5.4. Acquiring information on natural variation ....................................................................... 29

5.6. Facilitating participant discussion regarding initial estimates ............................................... 29
5.6.1. Case 1: When the participants are in agreement .............................................. 30
5.6.2. Case 2: When there is a clear difference of opinion ........................................... 30
5.6.3. Case 3: When there is considerable variation among estimates .......................... 31
5.7. Overcoming participant resistance to providing any response .................................. 32

6. After the elicitation ........................................................................................................ 34
   6.1. Interpreting and reporting the results ..................................................................... 34
      6.1.1. Case 1: The participants reach a consensus ...................................................... 35
      6.1.2. Case 2: The participants are divided in opinion ................................................. 36
      6.1.3. Case 3: Participant responses vary considerably ............................................... 37
   6.2. Communicating with the participants ..................................................................... 37
   6.3. Peer review of the results ....................................................................................... 38

7. References ..................................................................................................................... 39

8. Appendix—ET Documentation ..................................................................................... 40
1. Reading this manual

The purpose of this document is to provide guidance on how to conduct an expert elicitation using the ACERA Elicitation Software Tool (referred to as ET in this document). It outlines the process of preparing for, conducting, and analysing the results of an elicitation. The scope of ET and of this document is limited to elicitations where numerical estimates are being sought. This manual should be read in conjunction with the ET User Manual (ACERA, 2010a).

1.1. Who should read this manual

To run an elicitation workshop two key roles need to be performed, that of workshop facilitator and domain expert. Whilst it is possible for one person to run a workshop, in many cases two people working in collaboration will be necessary. This manual is primarily targeted at the workshop facilitator, although domain experts may also benefit from familiarising themselves with the elicitation process.

1.1.1. Workshop facilitators

Workshop facilitators are experienced at working with groups to generate a shared outcome. Typically, they work to an agenda, although they are flexible in the way in which this is implemented. Good facilitators can tailor their style to suit different audiences (e.g. domain experts versus key stakeholders), and are sensitive to the dynamics of a workshop (e.g. the context and motivations of participants, energy levels and current levels of engagement). Some of the elements they may employ include seating arrangements, introductions and declarations of interests, and rules for listening and speaking.

1.1.2. Domain experts

Domain experts are recognised and respected authorities within their domain of expertise. Their primary roles are to assist the workshop participants and facilitator to interpret context, background information and the meaning of questions. Whilst the workshop facilitator ensures the smooth progression of the workshop, the domain expert also gives the workshop an air of credibility. Domain experts are pivotal to establishing a shared understanding for each elicitation question, and for negotiating appropriate assumptions necessary to answer each question.

1.2. How to read this manual

1.2.1. I know what I’m doing and just want to get started

If this is the case read the document from Section 4 onwards. For technical help on how to use ET refer to the User Manual (ACERA, 2010a).

1.2.2. I’m a bit unsure and need some background information first

If this is the case you are likely to find Sections 2 and 3 helpful. You may also find some useful references in Section 7. Sections 4, 5 and 6 will guide you through the different stages of running the elicitation. Remember for technical help on how to use ET refer to the User Manual (ACERA, 2010a).
1.3. Feedback on this document

Please forward any suggestions or feedback regarding this document to Andrew Speirs-Bridge
speirs@unimelb.edu
2. Why conduct an elicitation workshop

Expert elicitation is a means of obtaining and synthesising expert opinion. It is used when no data are available (or easily obtainable), and uncertainty regarding a particular outcome needs to be qualified or quantified. Expert judgement is subject to a number of heuristics and biases (see Section 3) therefore special procedures are need to minimise their impact on the data gathered. It is important to acknowledge that expert opinion is less than perfect. However structured workshops aim to improve the quality of expert judgements over naïve or unstructured approaches.

In the context of ET the primary purpose of conducting an elicitation is to assess real life risk of public interest. To date ET has been used in a number of workshops including an import risk assessment, and workshops to investigate the potential spread of livestock diseases in Australia. Whilst ecological risks will continue to be the focus of the ACERA workshops, this does not preclude the investigation of risks outside this domain. ET may be used for expert panel risk assessment in any domain where expert opinion can be articulated in the form of interval estimates (see Section 3.1).

It is important to understand that data derived via elicitation is not equivalent to data obtained from empirical studies. Expert elicitation data can be considered a measure of uncertainty, where the uncertainty is due to lack of knowledge of the experts. This is in contrast to data resulting from well-designed empirical studies, which may be considered a measure of certainty, where the degree of certainty is influenced by sample size, measurement error, and natural variation. Consequently elicited data is judged to be less trustworthy than data from empirical studies (NHMRC, 2000).
3. **Key concepts**

This manual assumes that the reader understands the key concepts regarding interval estimates, the different sources of uncertainty, and the DELPHI workshop format. These concepts are briefly outlined below along with references to more detailed explanations. The frequentist question format is also introduced in this section, though workshop facilitators may choose not to use this.

### 3.1. **Interval estimates**

Estimates are often associated with a degree of uncertainty, both within science, and in real life. When asked to make an estimate of arrival time at a friend’s house, you might reasonably respond with ‘about five-ish’. This implies that you may arrive either shortly before five or soon after. More formally, this might be expressed as being no earlier than 4:45 and no later than 5:15. This is a naturalistic example of an *interval estimate*. As a familiar concept, interval estimates are also useful in the process of expert elicitation. Not only do they ask a question in a familiar format but they can also capture additional information regarding the degree certainty associated with the response.

#### 3.1.1. Overconfidence and the 4-step question format

While interval estimates are valuable when eliciting uncertain quantities, research suggests that they are prone to considerable overconfidence (Soll & Klayman, 2004). An interval estimate is considered overconfident when the interval is too narrow given the degree of uncertainty there is regarding the elicited quantity. However, it has been demonstrated that by simply changing question format, overconfidence can be substantially reduced (Speirs-Bridge et al., In Press). To ensure that participant overconfidence is minimised we use the 4-step question format outlined in Figure 1.

![Figure 1. The 4-step question format.](image-url)
3.1.2. Interval concepts

The concepts of precision and accuracy (Yaniv & Foster, 1995); hit and miss; and skewed and symmetrical distributions are important to appreciate when dealing with interval estimates. An accurate interval estimate is one that contains the true value, and is referred to as a Hit. However it is possible to achieve high degrees of accuracy by providing extremely wide intervals. Extremely wide intervals are said to lack precision, and the excessive width may not necessarily reflect plausible degrees of uncertainty. As a consequence there is a trade off between interval accuracy and precision, with overly precise intervals often resulting in a Miss, and under-representing plausible degrees of uncertainty. Figure 2 illustrates these concepts, and also highlights that interval estimates may have a skewed distribution, and are not required to be symmetrical around the best estimate.

Figure 2. Precision and accuracy in interval estimates.

3.2. Different sources of uncertainty

When considering the uncertainty associated with a judgement it is important to acknowledge that there are three potential sources of uncertainty.

3.2.1. Uncertainty due to language

Elicitation questions may be open to multiple interpretations due to the language in which they are expressed (Wallsten et al., 1986). Understandings vary because words have more than one meaning, they allow borderline cases, and their interpretation depends on background knowledge or context.
### Example Question

<table>
<thead>
<tr>
<th>Example Question</th>
<th>Comments</th>
</tr>
</thead>
</table>
| How much will it rain tomorrow? Provide your response in millimetres. | This question is ambiguous in a number of ways:  
• It is dependent on when participants answer the question as to what the true value will be.  
• The question also fails to specify where the rainfall will be measured and for what period of time it is to be measured for.  
However it does clearly specify the response units. |

This source of uncertainty can be controlled in two ways. The first is to word the questions carefully so that there is little scope for multiple interpretations. Section 4.4 offers guidance on this. The second is to facilitate a discussion with the participants so they can arrive at a clear, shared interpretation of the question and shared set of underlying assumptions. Section 5.5 offers guidance on this.

#### 3.2.2. Uncertainty due to lack of knowledge

Typically, an elicitation exercise will focus on the true value of a variable. The interval that arises from the 4-step procedure asks the participant to consider how sure their knowledge is, essentially quantifying their lack of knowledge about a ‘fact of the matter’. The resulting interval estimate, represents the lowest, highest and best values the participant believes are plausible.

<table>
<thead>
<tr>
<th>Example Question</th>
<th>Comments</th>
</tr>
</thead>
</table>
| What will be the annual rainfall in Melbourne for 2010? Provide your response in millimetres. | This question is less ambiguous than the previous one:  
• The period of measurement is precise,  
• and the location of measurement has been specified.  
However participants may still want to agree on assumptions regarding how rainfall is measured for Melbourne. For example is it just from one city centre weather station, or is it averaged across a number of Melbourne metropolitan weather stations? |
Lack of knowledge is also reflected by differences of opinion among participants. When making their initial estimate each participant uses their own private knowledge to determine their estimate. Given that each participant is likely to have different private knowledge (when eliciting an uncertain judgement) it follows that there should be some variation among the participants’ best estimates.

3.2.3. Uncertainty due to natural variation

Most quantities and probabilities vary naturally, independent of our knowledge about them. This kind of variation should be the focus of separate questions to the ones regarding a fact of the matter. It is important to clearly distinguish to participants between questions regarding ‘matters of fact’ and questions regarding natural variation.

<table>
<thead>
<tr>
<th>Example Question</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>What will be the variation in annual rainfall for Melbourne between the period 2000 and 2099? Provide your response in millimetres.</td>
<td>This question explicitly asks for the natural variation. However due to the wording participants may still be confused about what they are being asked for. Is it the difference between the mean annual rainfall for each of the 100 years, or is it the lowest, highest and mean annual rainfalls over the 100 year period?</td>
</tr>
</tbody>
</table>

Natural variation may occur over time or location. For groups with the appropriate statistical training it may also be appropriate to ask for an interval estimate of the standard deviation for a fact of the matter.

3.3. DELPHI workshop format

The elicitation process outlined in this document takes place in the context of a facilitated Risk Analysis Workshop using the DELPHI process for complex decision-making (Burgman, 2005). For ET a variation on the DELPHI workshop format is implemented, in which participants provide individual judgements, then review and discuss one another’s estimates, and finally make a second, private estimate. Just as the 4-step question format helps reduce overconfidence, the DELPHI process aims to reduce Anchoring and Halo effects often associated with expert panels. In addition DELPHI helps to alleviate the effects of group-think (Janis, 1982) and power differences between individuals.

3.3.1. Anchoring

When asked to give a numerical estimate, panel members will often use an existing value as an anchor and adjust up or down from this value. Tversky and Kahneman (1982) demonstrated this by randomly assigning participants to a value of either 65% or 10%, and then asking them a question regarding the percentage of African countries in the United Nations. Participants assigned to the 65% group consistently predicted 45%; whereas participants assigned to the 10% group consistently predicted 25%.

In the context of ET, anchors may take the form of the response units for a question (e.g. using millimetres versus metres anchors the order of magnitude of an estimate), and the specified upper and lower bounds for a response (e.g. using upper and lower bounds of 0 and 1m² implies that
responses over 1m² are implausible). One way in which to address this is to use standardised response units such as percentages (where the anchors are 0 and 100), probabilities (0 and 1) or by using the frequentist question format described in Section 3.4.

### 3.3.2. The Halo effect

This refers to the tendency of participants to make global assumptions about a panel member based on a limited amount of information (Nisbett & Wilson, 1977). Related biases emerge because we tend to have more confidence in, and attribute more weight to people who are confident, articulate and authoritative, even though these characteristics may have little bearing on a person’s domain expertise. Similarly, we may be swayed by the opinions of people who influence our careers, especially if they can see that our opinion diverges from theirs.

ET reduces Halo effects by allowing participants to remain anonymous if they choose. The facilitator may encourage panel members to identify themselves and defend their estimates when there is a clear difference of opinion. However irrespective of any discussion, participants’ final estimates are private and confidential.

### 3.4. Frequentist question formats

When manipulating probabilistic data people are most accurate when they use natural frequencies as opposed to probabilities or percentages (Gigerenzer, 2002). For example rather than combining percentages, Gigerenzer suggests that working with figures out of 1000 is more intuitive. We refer to this phrasing as a frequentist question format. This is best illustrated using the examples in the table below.

<table>
<thead>
<tr>
<th>Question Format</th>
<th>Question Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequentist</td>
<td>Imagine 1000 randomly sampled animals taken from the population. How many have the disease?</td>
</tr>
<tr>
<td>Percentage</td>
<td>What percentage of the population of animals has the disease?</td>
</tr>
<tr>
<td>Probability</td>
<td>When randomly sampling the population, what is the probability that an individual will have the disease?</td>
</tr>
</tbody>
</table>

Each question asks for the same information but in a different format. Whilst Gigerenzer’s research suggests the frequentist format is superior, this needs to be tempered with the following points:

1. Our experience has been that when there are more than three 0’s or more than three decimal places in a response, then responses become prone to transcription errors.

2. Some expert groups may prefer percentage or probability format questions because this is what they are familiar with. For example, epidemiologists are taught statistical procedures that manipulate probabilities, and so may prefer response formats using a probability format.

For more guidance on writing questions refer to Section 4.4.
3.5. **Process overview**

The elicitation process is divided into three key stages each of which discussed in its own section of this document:

- Before the elicitation (Section 4)
- During the elicitation (Section 5)
- After the elicitation (Section 6)

Figure 1 also provides a pictorial representation of the end-to-end elicitation process.

3.6. **Typical workshop agenda**

A workshop may last from two hours to two days, and involve between three and 40 expert participants. During the workshop, participants are asked to respond to questions (as few as 10, as many as 50). Two types of questions may be asked: elicitation questions (Section 4.4) where the answer to the question is unknown; and calibration questions (Section 4.5) where the answer to the question is known (but not widely known). It is recommended that between five and 10 calibration questions are included in each workshop.
The workshop agenda outlined here is appropriate when all participants can meet at a single location for one or two days.

A. **Preparation**: participants are selected and invited. Background information is disseminated, and the location and workshop venue are planned.

B. **Welcome and overview**: participants are briefed on the purpose of the workshop and given an explanation of the how the elicited information will be used.

C. **Introductions**: participants are introduced to each other, provide their demographic information, and register on ET.

D. **Familiarisation**: two to three practice questions are posed so that participants become familiar with the software and facilitation process.

E. **Elicitation**: the facilitator asks calibration and elicitation questions using the process described below:
   
   a. The underlying assumptions are agreed for each question. Language-based misunderstandings are resolved and a shared understanding of the question is reached across the workshop participants.

   b. Each participant enters his or her initial interval estimate. Errors are trapped and nonsensical entries corrected.

   c. Each expert reviews a visual representation of his or her interval adjusted to an 80% confidence level. He or she gets the opportunity to modify their estimate and once happy, the participant submits their response to ET.

   d. Once participants have provided an initial response to a question, the facilitator displays the estimates of all the participants together on a single graph. Each interval is identified with a participant ID number, which participants may choose to keep private.

   e. A discussion is then facilitated with the group in regard to the differences and/or similarities among the estimates.

   f. Following this discussion, the participants are provided with an opportunity to revise their 80% interval. This is referred to as their final estimate.

   g. The initial and final values for each expert are then stored but not displayed for a second time.

F. **Wrap up**: The facilitator thanks participants for their attendance, elicits feedback from the participants regarding the workshop, and summarises initial impressions of the responses (if any).

G. **Post-hoc analysis**: Export participant responses from ET for analysis.
4. Before the elicitation
Preparation prior to the elicitation is critical to its success. In this section key areas to consider in advance of the elicitation are discussed.

4.1. **Defining the problem**
The elicitation problem domain must be well defined, even if the quantities being elicited are uncertain. Some suggestions as to how to define a problem are presented below but these are by no means exhaustive.

4.1.1. **As a logical series of events**
Using this approach the elicitation starts at the beginning of a logical series of events and questions are posed for each step in the series. The illustrative example used below is related to a risk assessment conducted regarding the importation of mangos from India to Australia. The specific bio-security risk being assessed is the probability of an invasive pest (the Mango Weevil) establishing in Australia as a consequence of importing Indian mangoes. Each question is posed as a conditional probability. That is, assuming the step before has happened (or has not happened), what is the probability of the next step.

1. Consider 1000 randomly selected mango orchards in India. How many of these orchards are infested with the weevil?
2. Consider 1000 randomly selected fruit picked in an orchard affected by the weevil. How many of these fruit are infested?
3. The fruit are graded and packed into boxes containing 20 fruit each. Consider 1000 randomly selected boxes of mangoes sourced from infected orchards in India. How many are infested with the weevil?
4. Boxed mangoes are inspected, stored and then shipped to wharves in Australia. Consider 1000 randomly selected boxes sourced from infected orchards on arrival in Australia. How many are infested with the weevil?
5. Consignments are inspected by AQIS at the port of entry. Consider 1000 randomly selected boxes on arrival in Australia that are infested with the weevil. How many infested boxes will be detected by inspectors?
6. Consider 1000 randomly selected mangoes imported into Australia from India. How many do you expect to be distributed to areas in Australia that support susceptible hosts?
7. Given that 1000 infected fruit are delivered to an area with suitable hosts, how many weevils will successfully disperse to a local mango tree?
8. Given that 1000 weevils establish on mango trees in an Australian area with suitable hosts, how many will establish reproducing populations and spread to adjoining trees?
4.1.2. As a network of related nodes

A problem can often also be defined in terms of a pathway diagram or network model. These diagrams translate naturally into a formal model such as a Bayesian Decision Network. The pathway diagram with associated questions for the mango import risk assessment is illustrated in Figure 4.

Figure 4. Pathway diagram for the mango import risk assessment.
4.2. Identifying participants

Most definitions of expertise include skills, training, experience and professional recognition. We define an expert as someone who has knowledge not available to all, and who may make a useful, substantive contribution to a judgment (ACERA, 2007). We focus particularly on cases where quantitative judgments are required, and for which one can be reasonably secure that a fact of the matter exists. That is, this tool is not appropriate for questions of value and preference.

Expertise is usually highly domain specific: expertise and skills in one field do not transfer well to another field, or to tasks in the same field with which an expert is unfamiliar. Care must be taken to ensure that participants have training and/or experience that qualifies them as having useful, substantive knowledge. In elicitation tasks, often experts are asked to state their knowledge in unfamiliar forms. Knowledge about a subject area does not translate to an ability to accurately convey that knowledge in quantitative or probabilistic terms. Calibration questions (see Section 4.5) may be used to weight some individuals more heavily than others, or to exclude some participants entirely.

From a practical perspective the following approaches can be used to identify candidate elicitation participants (EPA, 2009):

- Identify experts who have published the most literature in the question domain.
- Look for experts who are representatives on relevant committees.
- Look for people with substantial, direct experience with the problem at hand (irrespective of their formal qualifications).
- Ask relevant professional, industry and other relevant organisations to nominate candidates.
- Ask experts to nominate peers.

When the elicitation is likely to influence sensitive policy decisions it is important that the process be transparent and defensible. To assist with this it may be prudent to exclude experts with a known association with the organisation conducting the elicitation (EPA, 2009). In this way the impartiality of the panel can be demonstrated.

Having identified a pool of candidate experts the next step is to generate a list of invited participants. Whilst practical constraints such as cost and availability will play a significant part in this selection process it is also important to understand how group dynamics influence judgements. Surowiecki (2004) suggests that “smart groups need to be diverse, independent and decentralised” (p. 22). Groups that lack these qualities are subject to “groupthink” (Janis, 1982), where group consensus is quickly reached, but poor quality judgements may be made due to group homogeneity.

Taking this into account it is important to stratify the selection of experts to:

- Draw from a diverse range of organisations (e.g. government, academic and commercial).
- Include participants with known differences of opinion regarding the question domain.

It is also important to appreciate that a smart group’s judgement improves as more opinions are added—a smart group can perform better than any single expert (Surowiecki, 2004). Counter
intuitively, group judgement can also be improved by adding dilettante participants (Surowiecki, 2004). Using the Mango Weevil example, group judgements may be improved by adding entomologists to the group, even when they have no specific expertise regarding weevils. Diversity can also be introduced by adding a range of expertise to the elicitation. In the Mango Weevil elicitation, entomologists, bio-security experts and dilettante participants were all included. In summary when identifying experts:

• Ensure that all relevant expert groups are represented.
• Include dilettante participants to increase the diversity of opinion.
• Assemble the largest group possible of relevant experts (see Section 4.3 for guidance regarding practical limitations).

4.3. Deciding on a workshop format

ET has been designed to accommodate a number of workshop formats including virtual panels. In practice, workshop format will often be determined by pragmatic factors such as the logistics and cost of bringing the participants together. Four workshop formats supported by ET are described below, and the advantages and disadvantages of each format are discussed.

4.3.1. Small groups (<10) in the same room

Facilitation: By restricting the workshop to a small co-located group of participants, the task of facilitating the workshop is simplified. Group discussions are likely to be easier to manage, and offering one-on-one support to participants for data entry into ET is feasible.

Venue: The ideal venue for this style of workshop is a small computer lab or training room where each participant is allocated their own computer with Internet access. An alternative solution is for each participant to provide his or her own laptop computer, at a venue with a wireless network for Internet access. In both situations the facilitator will require their own computer (also with Internet access) and the ability to display shared results screens to the participants via a data projector.

<table>
<thead>
<tr>
<th>Advantages:</th>
<th>Disadvantages:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Simplified facilitation</strong>—with only a small group to manage, workshop facilitation is simplified.</td>
<td><strong>Increased potential for bias</strong>—with only a small number of participants the potential for a diverse range of opinions is reduced.</td>
</tr>
<tr>
<td><strong>Shared understanding of each question</strong>—reaching a shared understanding and agreed set of assumptions for each question is easier face-to-face.</td>
<td><strong>Cost and logistics</strong>—even with a small group the costs and logistics of bringing the participants together at a suitable venue may be prohibitive.</td>
</tr>
<tr>
<td><strong>Extended duration</strong>—it may be possible to run a face-to-face workshop over a period of up to two days. This allows for an increased number of elicitation questions, and extended discussion regarding each question.</td>
<td></td>
</tr>
</tbody>
</table>
4.3.1. Larger groups (≥ 10) in the same room

**Facilitation:** In this format the facilitator chairs the workshop, and participants travel to the workshop location. An agenda is set for the workshop, but the facilitator can be flexible in how the elicitation progresses, potentially adding questions as the need arises, or revisiting questions when new information comes to light.

**Venue:** The ideal venue is a large computer lab or training room where each participant can be allocated their own computer with Internet access. An alternative solution is for each participant to provide his or her own laptop computer, at a venue with a wireless network for Internet access. In both situations the facilitator will require their own computer (also with Internet access) and the ability to display shared results screens to the participants via a data projector.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced bias—The increased diversity of private knowledge in the larger group is likely to reduce the group’s bias.</td>
<td>Challenging facilitation—Group facilitation becomes more challenging, and group discussions more difficult to manage. An experienced facilitator becomes necessary.</td>
</tr>
<tr>
<td>Flexible facilitation—The face-to-face workshop format allows more flexibility in the way in which the workshop progresses, adapting to participant input.</td>
<td>Most costly solution—Costs incurred for the venue, participant time and travel may be prohibitive.</td>
</tr>
<tr>
<td></td>
<td>Timing less flexible—The logistics of coordinating the attendance of the participants becomes more challenging</td>
</tr>
</tbody>
</table>

4.3.2. Virtual panel: Participants (<10) in different locations, connected via real-time conference call

**Facilitation:** In this format the facilitator chairs a conference call, and participants join the call from remote locations. Both participants’ and facilitator require access to their own computer with Internet access. For this workshop format to work it is important to have a detailed agenda distributed prior to the start of the workshop. Using the agenda, the facilitator can ensure all participants are at the same point in the elicitation, and give participants a sense of progress through the overall elicitation.

**Telephone conference:** It is important that high quality audio is maintained during the workshop. To ensure this it is recommended that conference phones and commercial conferencing services be used. Things to avoid include using mobile phones and speakerphones with no mute button. Both of these can introduce ambient noise into the conference making it difficult to follow conversations.

**Internet conference:** Commercial services and applications now allow conference calls to be delivered via the Internet. This may have the additional benefit of allowing for visual information to be easily shared among participants. However if all the participants are not familiar with the particular service or application being used the additional technical complexity may off-set the benefits of being able to share visual information.
Advantages: | Disadvantages:
---|---
**Reduced cost**—use of conference call facilities will be cheaper than assembling the panel in one location. | **Increased potential for bias**—with only a small number of participants the potential for a diverse range of expert opinions is reduced.

**Increased facilitation difficulty**—working with participants in remote locations (via a conference call) makes the task of facilitation harder. Without visual cues it becomes difficult to judge the energy, and level of engagement of participants. In addition the facilitator may need to be more skilled to build and maintain rapport for the duration of the workshop.

**Limited duration**—conference calls lasting more than two hours are likely to result in excessive participant fatigue. This may result in poor quality judgements and participant drop-out.

**Different time zones**—when the participants are located in multiple time zones, finding a mutually agreeable time for the conference call may become challenging.

4.3.3. Virtual panel: Participants in different locations, connected via email and responding independently

*Facilitation:* This approach changes the task of real-time facilitation to one of off-line coordination of the elicitation. Recommended steps are outlined below:

1. The facilitator emails participants a link to ET and instructions on how to register.

2. Participants register and login to ET and see the list of elicitation questions.

3. Via email participants seek clarification and establish a shared understanding regarding the meaning and context of each question via email. The facilitator needs to carefully consider how best to manage the flow of emails particularly if the anonymity of participants is to be maintained.

4. Participants enter their initial responses into ET within an agreed timeframe—we suggest that this be no more than one week after participants receive the registration email.

5. The facilitator releases the initial responses, and initiates an email discussion regarding the range of opinions. Once again the facilitator needs to carefully consider how best to manage the flow of emails.

6. Participants enter their final responses into ET within an agreed timeframe—we suggest that this be no more than one week after the initial results are released to the participants.
7. The facilitator thanks participants via email, encourages participants to feedback their experiences of the elicitation process, and also early feedback on the results (if appropriate).

<table>
<thead>
<tr>
<th>Advantages:</th>
<th>Disadvantages:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low cost—no travel expenses are incurred to bring the participants together. Participants may volunteer their time free of charge.</td>
<td>Limited opportunity for discussion—whilst discussion via email is feasible, it is likely to be less expressive and more difficult to follow than when face-to-face.</td>
</tr>
<tr>
<td>Greatest reach—with low cost and no time-zone issues international participation becomes increasingly viable.</td>
<td>Limited rapport building—it is more challenging for the facilitator to build rapport with the participants. As a consequence participation may become less involved, and participants may drop out before the completion of the elicitation.</td>
</tr>
</tbody>
</table>

<p>| |</p>
<table>
<thead>
<tr>
<th></th>
</tr>
</thead>
</table>

4.4. **Writing and framing elicitation questions**

In writing and framing questions it is important to minimise uncertainty due to language. Questions should be unambiguous and written in plain English, as far as possible. They should be developed in collaboration with the domain expert to ensure the consistent use of terminology, and domain specific conventions. When composing questions it is important to make them as familiar to the domain experts as possible, and try to use response formats that are intuitive to the participants. For example if participants are converting between probabilities (because that is the convention in their expert domain) and a frequentist format out of 1000, then the likelihood of errors being introduced into their estimates is increased.

Frame elicitation questions within the context of a mental model that is familiar to participants. This may either be an established model or one introduced at the beginning of the elicitation. Ensure that there is a shared understanding of the model prior to proceeding to the elicitation questions. Make explicit any underlying assumptions that are necessary to appreciate the model.

Present questions in a logical order. Group questions together that have a similar response format, and try and to avoid regular changes between response formats. Typical questions formats include:

- Given \( X \) conditions, out of 1000 randomly selected \( Y \), how many would you expect to be \( Z \)?
• Given the historical data $A$, what you predict the $B$ of $C$ to be, in the time period $D$?
• Given the outlined scenario what would you predict the value of $Q$ to be?

### 4.4.1. Eliciting information regarding categorical data

Whilst ET is designed to elicit numerical interval estimates it is possible to capture categorical data by using the frequentist question format:

- Identify a set of conditions that adequately characterise a category and use the following question format:
  - Out of 1000 randomly selected $Y$, given $X$ conditions, how many would you expect to be $Z$?

- For example given a site with the following characteristics:
  - A coastal cliff top location in Victoria Australia within 1m of a footpath.
  - Soil depth=8cm; Ph=7.3; Total N=0.16%;
  - Herb Cover=30%; Shrub Cover =10%; Stone Cover=60%;
  - Out of 1000 randomly selected sites, with the above characteristics, how many would you expect to have plants infected with Cinnamon Fungus (*Phytophthora cinnamomi*).

### 4.5. Identifying calibration questions

A calibration question is one where the true value is known to the elicitation organisers, but which is unlikely to be known to the participants. Whilst it is possible to run a risk assessment workshop without calibration questions this is less than ideal. Calibration question allow users of workshop output to weight participants differently, or to exclude some participants altogether (Cooke, 1991). It is important to establish with participants that the questions are ‘reasonable’ in the sense that they would expect an experienced, credible expert in the domain to perform better than a dilettante. Responses to calibration questions can be used to infer the quality of the responses to the elicitation questions. Participants who provide intervals that include the true value for the calibration questions can be considered well calibrated, with participants who provide relatively narrow intervals which contain the true value considered the most informative. Whilst this inference will always be subject to error, it is preferable to having no indication of each participant’s performance.

Ideally, each elicitation should include between five and 10 calibration questions. Well-calibrated participants provide 80% confidence intervals responses to the calibration questions that include the true value about 80% of the time. For example in an elicitation where there are 10 calibration questions, well-calibrated participants will provide 80% intervals that include the true value for 8/10 calibration questions. Participants who get less than 8/10 can be considered overconfident, whereas participants with more than 8/10 are under-confident.
4.6. **Augmenting ET with additional questionnaires**

Three situations have been identified when it may be necessary to augment ET with additional questionnaires.

4.6.1. **Collecting information not handled by ET**

There may be situations where it is important to capture additional data from participants. This may be in the form of demographic information (see Section 5.2), or additional data of relevance to the elicitation. The Participant ID allocated by ET at registration may be used as a primary key for associating this data with each participant’s estimates. In this way ET can help protect the privacy of participant information during data entry and analysis.

4.6.2. **When participants’ cannot easily access ET**

There may be occasions when it is not possible for all the participants to enter their judgements directly into ET. For example participants without formal scientific training may be anxious, or inexperienced in the use of computers. Similarly, in some situations, especially in developing countries, computers or Internet access may simply be unavailable. In situations like this it is possible to use paper (or electronic document based) questionnaires to gather initial responses using the 4‐step question format. These responses can then be entered into ET by the facilitator, on each participant’s behalf, or directly imported if the Excel version of ET has been used (ACERA, 2010b).

It is critically important for the success of the elicitation protocol that participants are able to sight the combined view of estimates for each question, prior to entering their final estimate. It is feasible that these could be printed out and distributed to participants. If participant judgements are collected using alternate means there is still value in entering the data into ET as if each participant were typing their own entries. This provides a simple means of data entry and aggregation into electronic format for further analysis.

Key drawbacks to this approach are:

1. When entering initial estimates using a paper based questionnaire participants do not receive feedback on how their level of confidence (as captured in the 4‐step question format) impacts the derived 80% interval.

2. Manually entering participant estimates is time consuming, and prone to transcription error.

4.6.3. **When initial estimates are gathered prior to the elicitation workshop**

It is possible to use paper (or electronic document based) questionnaires to fast track the workshop process. Initial interval estimates can be gathered prior to the workshop using a questionnaire. These can be then be loaded into the system prior to the workshop, to focus the discussion on the differences among the estimates. This can significantly improve workshop productivity and reduce its duration. However this is at the cost of creating a shared understanding of each question prior to making an initial estimate. If this is to work it is important either that the questions are framed so that they are not open to a broad range of interpretations, or that the domain expert contacts each
participant prior to the workshop to discuss with them their understanding of the questions and their context.
5. During the elicitation

This section offers guidance on how to facilitate an elicitation. Whilst this information is biased towards the face-to-face workshop formats describe in Section 4.3, it may also be of value for the virtual panel elicitation formats.

5.1. Setting the ground rules

In any facilitated workshop it is important to establish ground rules for participation. Whilst this list is not exhaustive it helps contextualise what is meant by a ground rule, and offers some guidance for less experienced facilitators.

1. Listen actively—respect others when they are talking and do not interrupt.

2. Participate to the fullest of your ability—workshop success depends on the inclusion of all participants.

3. Do not dominate the discussion—encourage other participants to share their views.

4. The goal is not to always agree—it is about hearing and exploring divergent perspectives and considering alternatives.

5. Speak about differences of opinion directly with the participant(s) involved — avoid complaining to others behind their backs.

6. Refrain from personal attacks—focus on ideas and core issues rather than personalities.

7. Be candid, honest, and open—do not blame, attack, embarrass, or put-down other participants.

8. Be conscious of body language and nonverbal response—they can be as disrespectful as words.

9. Take responsibility for observing the ground rules.

10. Feel free to come up with your own ground rules, but make sure everyone in the workshop agrees with the new rules.
5.2. **Participant registration and familiarization**

In order to capture the panels’ judgements it is necessary for participants to first register on ET. Currently ET only captures basic personal information: First Name; Last Name; Email Address. Additional demographic information may be captured using paper or additional electronic forms. These can use the Participant ID allocated by ET as a primary key for associating data. Information that we have found valuable in the past includes:

- Age
- Gender
- Level of qualification
- Years of relevant experience
- Membership of professional bodies
- Number of peer reviewed articles

It may also be helpful to have a couple of practice questions prior to the elicitation/calibration questions. This provides participants’ the opportunity to become familiar with ET and the elicitation process. Three example questions are provided below.

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
<th>Degree of Difficulty</th>
</tr>
</thead>
<tbody>
<tr>
<td>What percentage of the Australian population is female?</td>
<td>50.3%</td>
<td>Easy</td>
</tr>
<tr>
<td>In what year did Kevin Rudd enter parliament?</td>
<td>1998</td>
<td>Medium</td>
</tr>
<tr>
<td>What percentage of the Australian landscape is urbanised?</td>
<td>0.3%</td>
<td>Hard</td>
</tr>
</tbody>
</table>

5.3. **Emphasise that the best estimate is the median value**

From a statistical point-of-view the best estimate in the 4-step question format could be interpreted as being the mean, median or mode. It is important therefore to clarify with participants that the best estimate is intended to be the median value. One way to communicate this in lay-terminology is to explain to participants that the best estimate is as equally likely to above or below the true value.

In the 3-point question format (Soll & Klayman, 2004), from which the 4-step was developed, the question wording was more explicit. For example “I think the year is as likely to be after this year as before it” (p.307). However early pilot testing of the 4-step protocol found this wording to be somewhat confusing and cumbersome, and so we settled on the terminology of “best estimate”, which is less specific but more easily comprehended.

5.4. **Integrating calibration questions into the elicitation**

Experts who receive regular feedback become better calibrated (Cooke, 1991). Whilst this tends to be expert domain specific (e.g. weather forecasters, options traders, and racing handicappers) we believe that receiving calibration feedback may be helpful in bio-security and related domains.
Three protocols outlined below show different strategies for the inclusion of calibration questions, along with the advantages and disadvantages of each approach.

### 5.4.1. Before the workshop
Participants are asked to register on ET and enter their responses to a set of calibration questions prior to the start of the workshop. Without “naming and shaming” the participants, each receive feedback on their calibration prior to being asked any elicitation questions. Explicit instructions are offered to overconfident participants regarding increasing the width of their intervals.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants receive feedback on their calibration prior to answering the elicitation questions.</td>
<td>Participants may research their responses and as a consequence provide a biased indication of their degree of calibration.</td>
</tr>
<tr>
<td></td>
<td>Participants may enter inconsistent data due to their lack of familiarity with ET.</td>
</tr>
</tbody>
</table>

### 5.4.2. At the beginning of the workshop
Participants are asked to register on ET and enter their responses to a set of calibration questions at the start of the workshop. Without “naming and shaming” the participants, each receive feedback on their calibration prior to being asked any elicitation questions. Explicit instructions are offered to overconfident participants regarding increasing the width of their intervals.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants receive feedback on their calibration prior to answering the elicitation questions.</td>
<td>Participants may feel uncomfortable about having their performance measured, and withdraw from the elicitation.</td>
</tr>
</tbody>
</table>

### 5.4.3. During the workshop
An alternative protocol is to include the calibration questions amongst the elicitation questions without bringing them to the attention of the participants. Calibration questions should be evenly distributed throughout the elicitation. Without “naming and shaming” the participants, each should receive feedback on their calibration after the workshop.

<table>
<thead>
<tr>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>The calibration of the participants’ private knowledge is assessed under the same conditions under which their elicitation responses are captured.</td>
<td>Participants do not receive feedback on their calibration answers prior to the elicitation questions being asked, and as a consequence do not get an opportunity to assess their response style.</td>
</tr>
</tbody>
</table>
5.5. Establishing a shared understanding of each question

Using a workshop to elicit participant estimates offers a significant advantage over using a questionnaire. This is because there is an opportunity to establish a shared understanding of each question prior to participants making their initial estimates. Sources of variance due to a lack of a shared understanding among participant responses may include:

- Linguistic interpretation of the question.
- Underlying assumptions associated with the question.
- Interpretation of the meaning of the lowest, highest and best estimates.

Using the example question below each of these sources of variance is examined in detail.

- What is the population of the world? Provide your response in Millions of people.

5.5.1. Linguistic interpretation

“Is this question asking for an estimate based on how many people are living now or how many people have ever lived?” Despite best efforts of the facilitator in writing and framing their questions, there will always be the possibility for multiple interpretations. Ask at least one participant how they personally interpret each question and then poll the other participants to see if there are any alternate interpretations. Clarify any differences of opinion, and then agree on the meaning for the purposes of the elicitation. This process will often lead to a set of underlying assumptions being identified. Facilitators should actively seek instances of ambiguity, vagueness, context dependence, under specificity, and seek to eliminate them.

5.5.2. Underlying assumptions

“Is the estimate to be based on the last available census data, or model projecting population growth based on available data?” An agreed set of underlying assumptions should be established and documented for each question. This discussion often follows on naturally from the discussion regarding linguistic interpretation. If all participants agree on the linguistic interpretation it is still helpful to poll opinion regarding the underlying assumptions participants are making regarding the question. Sometimes it is not until the participants actually try to enter their estimates that they identify the underlying assumptions that they are making. Make sure these are communicated and agreed on by all participants, and if necessary allow participants to re-enter their data, adjusting their estimates based on new, agreed assumptions. Underlying assumptions may also hold for a group of similar questions. For example if the underlying assumption for the example question is that the estimate is based on projected population growth, then this assumption should also hold for questions such as “How many people are in Asia?”

5.5.3. Interpretation of estimates

“Is the difference between the lowest and highest estimate due to natural variation or to a lack of private knowledge?” Participants should be given explicit instructions indicating that the difference between the lowest and highest estimate should represent their lack of knowledge associated with the quantity being elicited (unless otherwise specified). Continuing with the same example a well-calibrated and precise participant might give the interval estimate:
• Lowest = 6799.9
• Highest = 6800.1
• Best = 6800 Million

However participants are unlikely to give such narrow estimates, and a well-calibrated participant might provide a much wider interval.

• Lowest = 6790
• Highest = 6810
• Best = 6800 Million

5.5.4. Acquiring information on natural variation

Once participants understand that as a default intervals represent lack of knowledge, additional questions may be posed regarding natural variation. For instance, if participants agree that the (true) mean rate of global population growth is 200,000 every day, the facilitator may then ask questions regarding natural variation:

• If we were to observe global population growth every day for 100 days, on how many days would the mean value exceed 250,000?

• And on how many days would the mean value fall below 150,000?

5.6. Facilitating participant discussion regarding initial estimates

Another advantage of running a face-to-face workshop is the increased ease with which it is possible to review and discuss the participants’ initial estimates. In general three styles of results may be returned.
5.6.1. Case 1: When the participants are in agreement

In Figure 5 participants appear to be in agreement that the best estimate is ~0.2. However there is some difference in the participants’ degree of confidence (as indicated by interval width) that may also be worth exploring.

![Figure 5](image)

**Figure 5.** The participants appear to be in agreement.

Typical questions that a facilitator might pose for a data set like this include:

- Can you think of any reasons why the value might be much higher/lower than 0.2?
- Why have participants 1 and 11 given such wide intervals? Does this represent natural variation, or does it represent a lack of knowledge? In this case we want to capture uncertainty due to lack of knowledge.

When asking participants to defend their estimates it is important to reassure them that their final estimate will remain private and confidential.

5.6.2. Case 2: When there is a clear difference of opinion

In Figure 6 there appears to be a two-way split with the group. One group believes the value is ~0.25 whereas the other believes it is closer to ~0.7.

![Figure 6](image)

**Figure 6.** The participants are divided in opinion.
Typical questions that a facilitator might pose for a data set like this include:

- Can someone explain to me why he/she believes the value is around 0.25?
- Can someone else explain to me why he/she believes the value is around 0.7?
- Does anybody believe that the average estimate represented by the black interval may be more representative of the true value?
- Based on the previous discussion is anyone considering changing their estimate from one group to the other, and if so why?

In some instances, plausible reasoning supports different positions and the group cannot be reconciled to a single position. Analysis in this case should retain both options as alternative explanations, and base subsequent analyses on two sets of assumptions.

5.6.3. Case 3: When there is considerable variation among estimates

In Figure 7 there appears to be no consensus among participants. In addition many of the intervals are extremely wide.

Figure 7. There appears to be no consensus among participants.

Typical questions that a facilitator might pose for a data set like this include:

- Participant 4 why have you provided such a wide interval?
- Participant 8 what justification do you have for being so certain in your response?
- Can anyone tell me why there is so much variation among participants?
- Does anyone believe that the average estimate, represented by the black interval, is the most plausible estimate in this case?
5.7. **Overcoming participant resistance to providing any response**

There are number of reasons why participants may be reluctant to provide any responses during an elicitation. The table below lists reasons we have encountered in the past and potential responses.

<table>
<thead>
<tr>
<th>Participant Reason</th>
<th>Facilitator Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>This question is too hard to answer.</td>
<td>By necessity many of the questions in this elicitation may be hard to answer. However it is important to understand that it can be helpful to identify the questions where there is greatest or least uncertainty. The goal of this workshop is not to establish a consensus opinion, but to get a better understanding of where the uncertainty lies within this domain.</td>
</tr>
<tr>
<td>More data are required to answer this question.</td>
<td>Is there a set of assumptions that we can agree on, which will make it possible to answer this question? We never have enough data. In this instance, we are depending on your experience and intuition to provide us with a best guess, until such time as the data become available.</td>
</tr>
<tr>
<td>The response depends on the underlying assumptions.</td>
<td>Is there a set of assumptions that we can agree on, which will make it possible to answer this question?</td>
</tr>
<tr>
<td>I do not have the time to do this (this excuse is used primarily when participants are responding via email).</td>
<td>We appreciate that your time is extremely valuable. We have indicated that responding to these questions should take approximately X mins/hours. Can you set aside mins/hours and provide us with the best response you can, given that constraint?</td>
</tr>
<tr>
<td>I am not qualified to answer these questions.</td>
<td>Whilst you may not feel qualified to answer these questions, you were invited to participate because we value your opinion in this matter. No one is likely to be truly expert in this domain, because the uncertainty is high. Remember that judgements expressing a great deal of uncertainty are valuable to us, and using the 4-step question format you can express your uncertainty as wide intervals with low levels of certainty.</td>
</tr>
<tr>
<td>I do not like making estimates without first being able to research my response. This excuse is</td>
<td>We are interested in your intuitive judgements in this particular domain. We are not aware of any</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Used primarily when participants are responding in a workshop setting.</td>
<td>Existing research that will positively influence your judgements in this matter, so we want you to leverage your general expertise in formulating your responses.</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>I do not want to give an answer that undermines my professionalism.</td>
<td>When dealing with uncertain events it is important to remember that no one knows for sure what might happen. By being unwilling to explore plausible alternatives you are in fact demonstrating a lack of professionalism. Whilst we appreciate this is a difficult task, you have been invited to participate because we value your opinion, and are keen to understand your thinking on this matter.</td>
</tr>
<tr>
<td>I am not used to making judgements under pressure in a public forum. I prefer to work in the comfort of my own office, when I have time to think.</td>
<td>We appreciate that making judgements in a public forum can be challenging, however this is also an opportunity to brainstorm new ideas with your colleagues. Remember the ground rules are that everyone’s opinion is valued, but at the same time differences of opinion are as important to capture as consensus.</td>
</tr>
<tr>
<td>This elicitation is a waste of time. The people running it do not know what they are doing. Why should I put any effort into it?</td>
<td>Are there practical improvements you can suggest? We value your opinion in this domain of expertise, how can we improve on what we are doing? We appreciate that your time is valuable. What changes would be necessary to improve your engagement in the elicitation process?</td>
</tr>
</tbody>
</table>
6. After the elicitation
The tasks following on from the elicitation are described in this section.

6.1. Interpreting and reporting the results
In interpreting and reporting the results of an elicitation it is important to appreciate that the estimates gathered represent the opinion of each participant, which is not equivalent to empirical data gathered through observation. In the broadest possible terms expert elicitation quantifies uncertainty, whereas empirical measurement quantifies certainty. As a consequence of this, reporting the results in a format similar to one illustrated in Figure 8, Figure 9 and Figure 10 along with a summary of the qualitative information gathered during the workshop may be the most faithful way of representing the results.

We appreciate that there will be situations where some numerical aggregate of the result be presented, in which case we suggest that this be in the form of an interval and not a point estimate. Cataloguing statistical methods for aggregating the results of the elicitation currently falls outside the scope of this document. EPA (2009) provides an excellent introduction and more detailed references regarding simple averaging (the technique used by ET), weighted averaging (Cooke, 1991), and Bayesian aggregation models (see ACERA project 0901).

Whilst statistically combining the results across participants from an elicitation can be problematic (see review in EPA, 2009), others suggest that groups of experts can perform better than any single expert (Surowiecki, 2004). The DELPHI method integrated into ET can be considered a behavioural means of building participant consensus, however as indicated earlier elicitation is more about quantifying the extent and magnitude of uncertainty than creating consensus. In the sections that follow the pros and cons of combining the results across three illustrative cases are reviewed.
6.1.1. Case 1: The participants reach a consensus

In Figure 8 the participant responses appear to converge on a similar value. In cases like this the simple average may adequately represent collective opinion of the participants.

![Figure 8](image_url)

Figure 8. An example of where the participants reach consensus. The black interval represents the simple average of the participant intervals.

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Little data is lost when combining the results using simple averaging. The summary results accurately reflect the group decision</td>
<td>The consensus may be due to shared bias among the participants. Check to see if there is sufficient participant diversity.</td>
</tr>
</tbody>
</table>
6.1.2. Case 2: The participants are divided in opinion

In Figure 9 there appears to be a clear difference of opinion among the participants. In cases like this the simple average is likely to misrepresent the collective opinion of the participants.

Figure 9. In this plot there appears to be a difference of opinion among participants. One group believes the best estimate lies between 0.2 and 0.3, whereas the other group believes it lies between 0.65 and 0.75.

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simplifies representation of the data, however this is an extremely poor reason for combining the results in this case.</td>
<td>Combining the results of the two groups using simple averaging (the black interval) misrepresents the judgement of the group.</td>
</tr>
<tr>
<td>In situations like this it is important to gather qualitative data (via facilitated discussion) as to the source of the difference of opinion.</td>
<td></td>
</tr>
<tr>
<td>• If the difference of opinion is due to divergent linguistic interpretations of the questions then clarify these and revisit the question.</td>
<td></td>
</tr>
<tr>
<td>• If the difference of opinion is due to natural variation then emphasise to the participants that their interval should represent lack of knowledge.</td>
<td></td>
</tr>
<tr>
<td>• If the difference of opinion is due to lack of knowledge then document the distinction between the two groups.</td>
<td></td>
</tr>
</tbody>
</table>

In situations where there are divergent, plausible opinions about facts of the matter, we recommend retaining all the alternatives in subsequent analyses.
6.1.3. Case 3: Participant responses vary considerably

In Figure 10 the participants offer a wide variety of judgements. In cases like this the simple average may or may not adequately represent the collective opinion of the participants.

![Figure 10](image)

In cases like this the simple average may or may not adequately represent the collective opinion of the participants.

Figure 10. In this plot there appears to be no obvious consensus among participants.

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>In this case the simple average may represent a better estimate than any single participant can offer. However this will only be true if participants represent a diverse, independent and decentralised sample of domain expertise.</td>
<td>These results may be indicative of participants having insufficient private information to be able to make an accurate judgement. If this is the case then the simple average does not adequately represent the true extent of the uncertainty associated with this judgement.</td>
</tr>
</tbody>
</table>

6.2. Communicating with the participants

It is important to recognise the commitment all participants have made to the elicitation. They have allocated time to the process, and may well have travelled to take part in the elicitation. It is therefore essential that all participants receive feedback regarding their individual contribution to the elicitation, as well as be debriefed regarding the overall results.

As described in Section 5.3 participants should receive feedback on their responses to any calibration questions included in the elicitation. Participants should also have the opportunity to debrief their personal experiences of the elicitation. Some participant’s may find the workshop process quite distressing, for this reason it is recommended that the facilitator makes himself or herself available for a short time after the completion of the elicitation to allow distressed participants an opportunity to debrief in private. This is in addition to encouraging participants to phone or email with any additional information or concerns. Formal evaluation questionnaires may also be helpful to provide the facilitator with feedback on his or her performance during the elicitation process.

Participants should be reassured that all individual responses will remain private and confidential in summary reports of the elicitation. Summary information may include the names of participants and the justification for their inclusion in the results. However when referring to individual
responses only the de-identified participant ID will be used. As a rule of thumb, participants may choose to defend their estimates in the context of the elicitation process, but should not be forced to defend them after the fact when summary results are published.

### 6.3. Peer review of the results

The purpose of a peer review is two-fold. The first is to ensure that the process used to elicit the data was conducted in a professional and objective manner. The second is to verify the results as plausible given the context of the elicitation and the underlying assumptions documented during its course.

A peer-review may be deemed necessary when the results of the elicitation are to be used as input into decision-making or policy development of public interest. In other situations a peer-review may be considered unnecessary and key-stakeholders will be satisfied with the due diligence of the elicitation process, and the integrity of the resulting data. EPA (2009) offers some guidance and additional references regarding conducting peer reviews.
7. References


8. Appendix—ET Documentation

To ensure that the documents can be appropriately targeted to their respective audiences, the following document suite has been produced for ET.

User Manual—this document is targeted at participants and facilitators, and explains how to use the native functionality of ET.

Process Manual—this document is targeted at facilitators (and to a lesser extent domain experts), and explains how to prepare for, run, and interpret the results of an elicitation.

Administration Manual—this document is targeted at system administrators, and explains how to manage, import and export data from the ET database.

Requirements Documents—this document is targeted at system developer, and specifies the ET system requirements.

Excel ET—this is an Excel spreadsheet that can be used to collect the elicitation responses from remote participants. This spreadsheet is prepared by a facilitator and distributed to participants as an off-line questionnaire via email. These responses can then be imported into the ET database to assist with data aggregation and visualisation.