Definitions

The definition of a Quantity of Interest (QoI) to be elicited in a SHELF workshop should satisfy three conditions.

1. The definition must be clear and unambiguous.
2. It should be such that the QoI will have a unique value.
3. It should be formulated so as to make the experts' judgements as simple as possible.

In this document we take each of these conditions in turn, explaining why they are necessary and giving practical advice on achieving them.

"Clear and unambiguous"

If there is any ambiguity in the definition there is a risk that experts will interpret it differently, and so will make judgements that refer to different quantities. Although this is likely to be discovered during the group discussion, there will be time lost to identify it and then to resolve the problem by developing a new unambiguous definition. Also, the experts' individual judgements will not be an accurate representation of their initial positions regarding the (redefined) QoI.

To highlight how easy it is to leave ambiguity even in an apparently clear definition, consider the following examples.

Example 1

The QoI is

"The dollar/pound exchange rate in July next year"

Some of the ambiguities in this definition are quite obvious.

- What dollar? There are of course numerous dollar currencies around the world. There are even several 'pound' currencies in addition to the UK's 'pound sterling', for instance the Egyptian pound.
- What/whose exchange rate? Different organisations – bureaux de change, banks and credit card companies – operate with different currency exchange rates.
- When in July? The rates vary daily, and in some cases from one moment to the next.
- Dollar to pound or pound to dollar? In some official rates, one will be the inverse of the other, but this is certainly not the case for commercial organisations like bureaux de change.

Here is a definition that removes those ambiguities.

> *"The Bank of England daily spot rate for US dollars against the pound Sterling on 1st July next year"*

Example 2

> *"The unemployment rate in Gothenburg"*

Some relevant questions here are:

- How is unemployment: measured? There are various ways to measure unemployment, and official figures from different countries or organisations do not always use the same methods. The differences are often a source of political argument!
- In what units? Unemployment is usually quoted as a percentage (e.g. 7.1%), but it can for instance be a pure number (0.071), or a number per thousand (71).
- Which Gothenburg? There is a city in the US state of Nebraska called Gothenburg, as well as the second-largest city in Sweden.
- City or metropolitan area? There can usually be different definitions for the boundaries of a city.
- When? The unemployment rate will certainly vary over time.

We can answer these questions with a definition such as

> *"The percentage of the labour force resident in the metropolitan area of Gothenburg, Sweden, who were actively seeking work at April 5, 2016"*

Notice, however, that there are terms embedded in this definition that may themselves need definition – 'labour force' and 'actively seeking work'. The particular definitions of these terms used in Swedish official statistics may be familiar to the experts, but if there is any concern that they may not then they should also be defined (clearly and unambiguously!).

"Unique value"

The definitions in the preceding two examples clearly identify single, unique values for the corresponding QoIs, thereby satisfying our second condition. However, there are situations where we need to answer other questions to achieve uniqueness.

Example 3

A project to protect tigers in Northern India is considering trapping tigers and transporting them to wildlife refuges. In order to understand the scale of the task, the project would like to know …

> *"The weight of a tiger"*

How might this vague quantity be defined properly? As in the first two examples, there are several questions.

- What units? For instance, we could express the weight in kilograms or pounds.
- Which sub-species? Although the tiger is a single species, several sub-species are recognised, such as Bengal or Siberian tigers. They might have very different weights.
- At what age? Adults, juveniles and cubs will certainly differ in weight.
- What gender? Male tigers are typically larger than females.
- When? A tiger's weight may vary with the season. The time of year should be specified if this is relevant.

The project team may revise their definition to reflect the fact that it is Bengal tigers that they will be trapping in Northern India and if weight is to be a limiting factor for the project that it is adult male tigers that are of most concern.

> *"The weight (kg) of a wild adult male Bengal tiger in Northern India"*

As in Example 2, it may be necessary to further define what is meant by 'Northern India', unless the experts can be assumed to have a common understanding of this term. But in this example there is another important question.

- Which tiger?

Individual tigers will certainly vary in weight! So the question is still ambiguous because the weight as defined above does not have a unique value. It is very common for quantities that are relevant to the wider question to be subject to random variability. In order to identify a suitable definition, we need to look carefully at uncertainty and variability.

Uncertainty and variability

We consider a quantity A that will have a value for each member of some *population*. In the case of Example 3, the population is the collection of all wild adult male Bengal tigers in Northern India. To clarify the notion of the population, here are some more illustrations.

- A pharmaceutical company is interested in the efficacy of a new treatment for people with a particular disease. The population is the set of all people with this disease who will receive the new treatment. The quantity A is the magnitude of the effect of the treatment on each patient.
- A sociologist is interested in the usage of email amongst people aged 70 and over in France. The population comprises all the people aged 70 or more in France. The researcher's quantity A is the number of email messages sent by each person in a given week.

Where there is variability of this kind, we cannot define the QoI to be simply the quantity A, because it does not have a unique value. We need to be more explicit. It may be that we are interested in individual values of A in the population. We might define X to be the value of A for a specified individual member of the population or, more often, a randomly selected member of the population.

However, the QoI that is relevant to the wider context will generally be some property of the population A values. The pharmaceutical company may be interested primarily in the average treatment effect. The sociologist is likely to be interested in the proportion of people who sent no email messages in the specified week, and perhaps also in the proportion who sent more than a small number such as 10. For the tiger project, the principal concern might be the maximum weight. Alternatively, they may be willing to make special arrangements for a few unusually large animals, and wish to know what weight is exceeded by only 5% of the population values.

There is invariably uncertainty regarding the population values of such a quantity, and hence in particular there will be uncertainty about the value of any QoI such as the value of A for a random member of the population, or the average value of A in the population. It is important to be aware of some confusions that can arise in these cases.

Consider a QoI X that is the population average value of the quantity A. When asking experts to judge the plausible limits of X, it is common for them to think of the range of plausible values for an individual A. For instance, an expert might judge that the weight of an adult male Bengal tiger could vary from 120 to 350kg, and it is tempting to then assign similar values as the plausible limits for the average. It will often be necessary when dealing with an average for the facilitator to point out that if, for instance, individual values might range from 120 to 350 then the average simply cannot plausibly be close to either of these limits. The plausible range for an average should generally be considerably shorter than for an individual.

Now suppose that X is the value of A for a randomly chosen member of the population. Uncertainty about X has two components – there is the randomness associated with a random draw from the population values of A, and there is also intrinsic uncertainty about those values – and this makes it difficult for an expert to make judgements about such a QoI. But definitions “should be formulated so as to make the experts’ judgements as simple as possible.” We now examine this third condition.

“As simple as possible”

SHELF templates take the expert through a carefully constructed sequence, designed to elicit the kind of judgements that experts make most accurately, but the definition can play its part in making the expert’s judgement task as simple as possible. For instance, it should be framed using terms that are familiar to the experts.

Some quantities are better defined on a logarithmic scale. Instead of the original QoI X, we ask the experts to make judgements about Y = log(X). For example, suppose that X is the number of plastic items dropped directly into the Mediterranean Sea from shipping in 2017. Even an expert may be so unsure what this number will be that they judge it could take any value from a few thousands to tens of millions. Then transforming to the logarithmic scale, base 10 so that if X = 1,000 then Y = 3, and if X =10,000,000 then Y = 7, may make it easier for the expert

Taking this idea of transformation further, we consider cases where the QoI may be usefully redefined in terms of two or more other quantities, each of which is easier for the experts to formulate judgements about. This restructuring of the quantity is a powerful technique known as *elaboration*.

Elaborating a single quantity

In general, an elaboration is a way to derive the probability distribution of the original quantity of interest X from the distributions of two or more quantities, Y1, Y2, …. These then become the new QoIs. Once we have elicited the experts' knowledge about them in the form of probability distributions, we can infer the probability distribution for the original X. (We note here that in fact what is needed is a *joint* distribution for Y1, Y2, …; this is discussed in detail in the document "Multivariate Elicitation".)

An effective elaboration is such that Y1, Y2, … are easier for the experts to think about and so we can more easily and accurately elicit a distribution for X indirectly by eliciting distributions for Y1, Y2, ….

The following series of examples will explain and illustrate the ideas of elaboration.

A simple elaboration

A businessman needs to drive from central London to the centre of Bristol to attend a meeting. Let X be the time required (in hours) for this journey. It is natural to think of X in terms of the length L of the journey (in miles) and the average speed S (in miles per hour). We could therefore elicit distributions for L and S, and then using the equation X = L / S we can derive the distribution for the original X. Thus L and S are the two new QoIs (which were denoted by Y1 and Y2 in the general formulation of elaboration).

This elaboration achieves the desired result of L and S being easier to think about separately, rather than trying to elicit beliefs directly about X.

We could elaborate further by noting that the journey is composed of two slow sections getting out of London and into Bristol, together with a fast section between the two cities on a motorway. If we define L1 to be the combined length of the slow sections, L2 to be the length of the motorway section, S1 to be the average speed in the slow sections and S2 to be the average speed on the motorway, then

$$X = (L1 / S1) + (L2 / S2)$$

This may be a better elaboration than the first because it will be easier to think about S1 and S2 individually than about the average speed S for the whole journey.

A chain

In a risk analysis for trans-national spread of a disease of cattle, the QoI is the number, X, of infected animals imported in a given year to country B from country A, where the disease is endemic. In order to assess X directly, an expert would need to consider the number of cattle exported to B from A in that year, and the proportion of these that will be infected when they leave A. Furthermore, since the journey can take several days the expert would need also to consider whether some cattle that were infected at the start of the journey in A may have ceased to carry the disease before arrival at B.

The following elaboration separates the different factors and allows them to be assessed independently:

$$X = V \times I \times (1 - C),$$

where V is the number of animals exported from A, I is the proportion of those animals that are infected, and C is the proportion of infected animals whose disease will have cleared before arrival.

In risk assessments, it is common for elaborations like this to arise, where the outcome whose risk is to be assessed arises from a chain of circumstances (in this case, infection and then non-clearance of the disease).

Extending the argument

A company needs to decide whether to expand its capacity by building a new factory, but the economic future is uncertain. The QoI is the value X of the company's capital at the time when the final payment on the factory must be made. A group of financial experts is convened, but they are unwilling to make judgements about X because it will be strongly affected by the outcome of an impending national election. If the current governing party is not re-elected, the value of the company's investments is likely to fall.

Two probability distributions are therefore elicited from the experts, a distribution for X conditional on the governing party being re-elected, and another conditional on them not being re-elected. Another adviser adds their judgement of the probability that the governing party will be re-elected, which can then be used to combine the two conditional distributions into the required (marginal) distribution of X.

This elaboration is called *extending the argument* – probability distributions for the QoI are elicited conditional on two or more uncertain contingencies, and probabilities are elicited for the contingencies. It is useful whenever the experts feel that the uncertainty over the various contingencies makes it difficult to make judgements about X.

Variability and uncertainty

Consider the earlier example of eliciting beliefs about the value X taken by a quantity A for a random member of its population, where we noted that the judgement task was complicated by the need for the expert to account for both variability due to random selection and uncertainty about the population values of A. To separate these two components, first suppose that the expert knows all the values of A for the members of the population. Then the expert's probability distribution for X will be identical to the physical distribution of population values. For instance, if the expert knows that 30% of the values in the population are less than some value x, then the expert's probability that X is less than x will be 0.3, because this is the definition of a random draw from the population. The first step of the elaboration, then, is that the expert's probability distribution for X *conditional* on the population values is equal to the physical distribution of those values. We do not need to elicit this conditional distribution.

Of course the expert does not know the physical distribution of the population values, and this is where we need to elicit the expert's judgements. Because of the large number of uncertain values, this is a complex task, and suitable elaboration is discussed in the document "Multivariate Elicitation".

To elaborate or not to elaborate?

In principle, elaboration is a powerful device to simplify the expert's task of making probability judgements about X. In the first two examples above, the expert is asked to make judgements instead about two or more quantities Y1, Y2, … that are easier to think about. In extending the argument the simplification comes by allowing the expert to make judgements about X *conditional* on each of a set of contingencies, and these distributions are combined through additionally eliciting probabilities for the contingencies. In the final example, the task is simplified by recognising that if we condition on the composition of the set of population values then the distribution of X is implied by the random selection, and then the expert can focus on uncertainty about that composition.

However, in every case, instead of needing to elicit one probability distribution, that of X, we now need to elicit several. So elaboration increases the elicitation load. It is not necessarily true that because judgements about each of the distributions required by the elaboration will be easier for the expert, and therefore hopefully lead to distributions that accurately reflect the expert's knowledge, the resulting distribution of X will also be a better, more accurate representation. And if it is better, it is not always clear that the improvement justifies the extra effort.

An important consideration is that the quantities whose distribution are to be elicited may not be independent, in which case the task of eliciting those distributions becomes much more difficult. Independence, and ways to achieve independence through elaboration, are discussed in detail in the document "Multivariate Elicitation".

Whether to use elaboration, and if so whether to use a simpler or more complex elaboration, is a matter for the elicitation team to decide.