Many Quantities of Interest

It is not unusual for a project to have a large number of quantities of interest (QoIs) that are uncertain and so require probability distributions to be specified, and yet there is not enough resource (time, money or energy!) to elicit a distribution for every quantity in a full SHELF elicitation workshop. We therefore need strategies to (a) identify which QoIs should have a full SHELF elicitation, and (b) specify probability distributions for the other QoIs in ways that require less resource than the SHELF process.

We will begin by addressing (b): if we can only use the full SHELF method for a subset of the QoIs, how can we elicit probability distributions for the remainder, in a less resource-intensive way than SHELF demands, whilst accepting that any such method will be less rigorous and more prone to errors of judgement or biases?

Minimal assessment

The simplest and crudest method is for the project team to make quick, approximate judgements themselves, without reference to external experts. Minimal assessment for a quantity X consists of two simple judgements.

1. An estimate m_X. The team may interpret this as a median, or simply think of a 'best estimate' in some unspecified sense. It is not intended to be a precisely defined judgement.
2. An uncertainty measure s_X such that the team judges that X is about twice as likely to be in the range $m_X \pm s_X$ as to be outside that range. Again, this is not intended to be a precise probability judgement.

A distribution is then fitted to these judgements in any convenient way that matches their 'definitions'. For instance, it might be a normal distribution with mean m_X and standard deviation s_X.

Compared with a full SHELF elicitation, minimal assessment is clearly deficient in almost every respect. The judgements are made very quickly and imprecisely, without involving experts outside the project team. Nor is an evidence dossier used, although the team are expected to be familiar with at least some of the available evidence. Yet minimal assessment does provide a distribution that is centred around the value of X that the team consider to be most likely, with a spread that reflects their uncertainty (which should be larger than would result from the gathering of evidence and the expert discussion that takes place in the SHELF).

Probabilistic Delphi

The Delphi method has been a well-established way of enlisting expert judgement since the 1960s. Experts do not meet, but are sent a questionnaire asking them to estimate one or more uncertain quantities. The facilitator collects the estimates from all those experts that respond to the questionnaire and sends a second questionnaire asking the same questions but now giving the experts a summary of the estimates from the first questionnaire. The experts are now expected to consider revising their own original estimates after seeing the estimates from the other experts. Further rounds may follow, and in principle the Delphi method is supposed to continue until consensus is reached, where all experts agree on a common estimate. Anonymity is a key feature of Delphi – the experts do not know who is in the group of respondents – which was a conscious decision by the developers of the method to avoid the problems inherent in group discussion and judgement (and which the facilitator must manage in a SHELF workshop).

An obvious limitation of the traditional Delphi method is that it does not elicit the experts' uncertainty. Therefore a version known as Probabilistic Delphi is proposed, in which the questionnaire asks experts to make the same set of judgements as in a SHELF method. They should make these judgements in the same sequence, beginning with credible limits. The facilitator gathers all the respondents' judgements, and in sending out the next questionnaire provides feedback summaries of those judgements.

An evidence dossier should be prepared and sent to the experts with the first questionnaire, but this method still does not have two of SHELF's key features, the discussion between experts and the final group judgements appropriate to a Rational Impartial Observer (RIO). To partially address the first of these deficiencies, experts are asked to provide explanations, known as 'rationales', for their judgements and these are also forwarded to all experts with the next questionnaire.

It is not feasible to reach a RIO consensus with Probabilistic Delphi, nor to iterate until all the experts agree on all the different judgements. Therefore, after two or three questionnaire rounds their final judgements are averaged. A simple way to do this is for the facilitator to fit a distribution to each expert's final judgements and then to average the distributions.

Probabilistic Delphi does not require experts to come together, which can be a significant saving in cost and time. The time requirement for the experts themselves is reduced, and in practice they can make judgements about multiple quantities of interest in a single questionnaire. The price to be paid for this gain in terms of resource cost and convenience is the loss of two of the SHELF method's key features – the discussion between experts and the final group judgements from the RIO perspective.

It should also be noted that Probabilistic Delphi cannot be carried out quickly. A dossier must be produced, experts need time to complete each questionnaire and the team or facilitator need time between each questionnaire round to collate responses and rationales, and to prepare summaries for feeding back to the experts.

Training is particularly important with this method. Experience has shown that it is not enough to send the experts instructions, of the kind that the facilitator would give in a SHELF workshop. The extent to which this is a problem depends on context.

- We are considering using Probabilistic Delphi in the context of some QoIs being elicited using the full SHELF method, so the experts in the corresponding workshop(s) will be fully trained and experienced. These same experts can then be given a Probabilistic Delphi questionnaire to make judgements about other QoIs while this training is still fresh in their memories.
- However, in a different context we also recommend Probabilistic Delphi as a next best alternative when the SHELF workshop is simply not feasible. The most important instance of this is when there is no common language in which the experts are all reasonably fluent, in which case they cannot all contribute to group discussion and group judgements. Since questionnaires can be translated into different languages if necessary, Probabilistic Delphi is feasible in this situation. Training is now a significant problem, however. Experts should be required to follow the online training course on probabilistic judgements that is available from the SHELF website, but this method will still be more prone to errors of judgement and bias than SHELF.

Prioritising

The full SHELF method should be applied to the most important quantities of interest, in terms of their influence on the principal outcome of the wider project. For example, in planning a building project, for which uncertain quantities of interest concern the resource requirements for different parts of the work, the outcomes will include the time to complete the work, the manpower requirements and the cost. The team may decide that the total cost is the principal outcome.

The importance of each QoI is then assessed by a simple one-way sensitivity analysis. We begin by applying minimal assessment to every QoI. Then for each QoI in turn the principal outcome is computed with this QoI, say X^*, set to $m_{X^*} - s_{X^*}$ and all other quantities set to their m_X, and then computed again with the value of X^* changed to $m_{X^*} + s_{X^*}$. The importance of X^* is then the difference between these two values of the principal outcome.

It is not uncommon for some QoIs to have very small importance, implying that over the range of likely values they have negligible impact on the principal output. For these QoIs, we clearly do not need to spend more resource on eliciting their distributions – their minimal assessments are an adequate representation of their uncertainty. All other QoIs are important and therefore candidates for a full SHELF elicitation workshop.

If there are still too many important QoIs, the full SHELF process should be used for those with the highest importance. Probabilistic Delphi is an option for any remaining QoIs for which it is felt that minimal assessment is not adequate. After the SHELF workshop(s), a Probabilistic Delphi questionnaire is sent to the same experts, who have therefore already been fully trained and have experience in making the necessary probabilistic judgements.

Finally, if there are still too many quantities of interest in this intermediate category for it to be practical to elicit them all by Probabilistic Delphi, then we suggest that the project team should make a final revision of their minimal assessments of the remainder, in the light of the discussions and expert judgements in the SHELF and Probabilistic Delphi elicitations. In this case, it is desirable that the QoIs that are chosen for Probabilistic Delphi are representative of or related to those not chosen.

We therefore have the following hierarchy.

Method	**QoIs**
SHELF workshop	Most important
Probabilistic Delphi	Less important, but representative or related
Revise minimal assessment	Non-negligible importance
Minimal assessment	Negligible importance

Caveats regarding one-way sensitivity analysis

The proposed one-way sensitivity analysis is designed to be quick and simple, so as to maximise the availability of resources for elicitation. However, there are situations in which it can give quite misleading indications of importance.

Correlated quantities

One of these situations is when QoIs are correlated. Consider the use of a climate model to predict future global average temperatures. A key set of uncertain inputs to such a model would be the CO_2 emissions each year. If X_n is the emissions figure in year n and the model is being used to forecast the next 50 years, then the total emissions would be $X_1 + X_2 + \ldots + X_{50}$. Now the most uncertain of these inputs will be X_{50}, which will therefore often be assigned the largest importance by the sensitivity analysis, closely followed by X_{49}, X_{48}, etc. However, the annual emissions will be strongly correlated, with each year changing relatively little from the previous year. Letting Z_n be the change $X_n - X_{n-1}$, and if X_0 is the current year's emissions figure, which is known, then the total emissions can be written as $X_0 + nZ_1 + (n\text{-}1)Z_2 + \ldots + 2Z_{n\text{-}1} + Z_n$. Now Z_1 (which is equivalent to X_1) is seen to be the most important uncertain quantity; the order of importance is reversed.

The one-way sensitivity analysis is most reliable when the QoIs are independent. Therefore, correlated quantities should wherever possible be expressed using elaboration in terms of independent components before prioritising with minimal assessment and one-way sensitivity analysis. In the climate example, the elaboration in terms of the emissions increments Z_n should be used if those increments are judged to be independent.

Nonlinear response

One-way sensitivity analysis is also most reliable when the principal output is simply a linear function of the QoIs. Suppose that as the QoI X^* is varied the principal output has similar values at $m_{X^*} - s_{X^*}$ and at $m_{X^*} + s_{X^*}$, and that both are lower than the value at m_{X^*}. Then clearly the importance of this QoI according to the sensitivity analysis will under-estimate its true influence on the principal output.

Elicitation often provides input distributions for very complex, nonlinear models, and the quick and simple sensitivity analysis approach described above may be too quick and simple. If the model or analysis is not too complex, so that the principal output can be evaluated relatively quickly for a range of input settings, then a more robust sensitivity analysis can be considered. For instance, when evaluating the importance of X^*, outputs can be obtained at other values of X^*, for instance m_{X^*}, $m_{X^*} - 2s_{X^*}$ and $m_{X^*} + 2s_{X^*}$, as well as at $m_{X^*} - s_{X^*}$ and $m_{X^*} + s_{X^*}$, and the importance defined as the difference between the highest and lowest outputs.

Another suggestion when the computation is not too onerous is to evaluate the importance of a given X^* with the other QoIs fixed at settings other than their m_X values. Values for each of those other QoIs could be randomly set to their m_X, $m_X - s_X$ or $m_X + s_X$, with the importance of X^* calculated as above for several such random settings. It could then be assigned the maximum or average of these random importance values.

A final caveat

Even if devices like these can be applied, it should be remembered that minimal assessment is a crude, quick and simple process, and therefore the subsequent one-way sensitivity analysis can at best only give a rough indication of the relative importance of different QoIs.