Eliciting Beliefs in Health Technology Assessment to Characterize Uncertainties

Marta Soares, MSc, Research Fellow, Centre for Health Economics, University of York, York, UK

KEY POINTS . . .

In Health Technology Assessment (HTA), expert judgements are often sought for informally.

There is a need for using more formal approaches to the elicitation of expert opinion in this context, in a way that is more explicit, valid, and accountable.

This article describes some of the important aspects of formal approaches to elicitation; specifically those that aim to elicit uncertain judgements.

Context

Health technology assessment (HTA) aims to evaluate the consequences of using alternative health technologies in order to better inform decisions on which technology to use at a societal level. The assessment is based on evidence relating to quantities of interest for the appraisal, for example, the frequency of occurrence of important health events or the extent to which treatments can affect their occurrence. The evidence is then gathered together using a decision model, a mathematical model, which relates these quantities, and estimates outputs such as overall survival, quality adjusted survival, and/or total costs, for each treatment of interest. By comparing across treatments, one can establish incremental effectiveness, incremental costs, and incremental cost-effectiveness that can then be used to inform decision making.

Despite most of the evidence used to inform such assessment coming from empirical studies (where a sample of individuals is observed for one or more outcomes of interest), judgements are also often used. One example may be seen in informing the values of quantities for which there is sparse or no evidence. In this case, we may seek the knowledge of an 'expert' over a specific quantity to be translated into quantitative values – this process is called elicitation. Anecdotally, elicited judgements in HTA are often sought informally, where an expert is contacted by the researcher and asked directly about his/her beliefs on the quantity of interest. There is, however, the need for using more formal approaches to elicitation instead; this is because the quality of such exercises rests on their methodological validity, given that it is impossible to ascertain whether results are themselves valid (the true knowledge of experts is not observable).

This article describes formal approaches to elicitation; specifically those that aim to elicit uncertain judgements, i.e. those that aim to produce a description of the range of values experts believe best represent reality concerning a specific quantity.

Uncertainty: Importance for HTA and Elicitation

Reflecting upon existing uncertainties in HTA is recognised as fundamental, as it allows an explicit assessment of how confident the decision maker can be in his/her decision. This is true of evidential uncertainties, i.e. where evidence collected within an empirical study from a sample and thus values for the whole of the population of interest are not observed but are inferred. This also applies, however, to epistemic uncertainties in the beliefs of experts (or uncertainty in knowledge). Thus, for elicitation to capture such knowledge, it needs to not only describe the experts' best guess, but also any uncertainty experts may have over this value. Where multiple experts are used both within and between expert uncertainty will need to be considered and analysed.

Elicitation Approaches

More formal approaches rely on a careful design and structured implementation of the elicitation exercise. There are several recommended stages to an elicitation process: the design of the exercise, its conduct, the synthesis of multiple experts' beliefs and assessment of the adequacy of an elicitation exercise. We will next briefly describe each stage. Note that we do not aim to be comprehensive, but rather to refer to literature that complements the information provided.

Design of an Elicitation Exercise

Firstly, in designing an elicitation exercise, one needs to define whom to elicit from. We should seek to select substantive experts in the particular area of interest, but without competing interests (so as to reduce motivational bias). In HTA, we may aim to elicit from health carers, which are unlikely to have strong quantitative skills. Such experts can be assumed to provide reasonable estimates of observable quantities, such as proportions, but it is improbable that they are able to elicit more complex quantities such as means, variances, or other parameters of statistical distributions. Secondly, one must define how many experts to include in an exercise. Generally, multiple experts will provide 'more' information than a single expert; however, there is a lack of guidance about the appropriate number of experts.

What to elicit is also of importance. There may be different possible ways of eliciting information on each of the uncertain quantities in which we are seeking expert knowledge. For example, in informing the occurrence of an event of interest, such as myocardial infarction, one could ask experts about the proportion of patients expected to have the event in a year, or the expected time until first myocardial infarction. Ideally, the focus should be on eliciting quantities with which the experts are familiar, for example proportions. It will be useful to pilot the elicitation exercise with one or two experts so that its design may be appropriately finalised.

In what concerns the exercise of eliciting uncertainty, experts can be asked to elicit credible/confidence intervals directly (the range of values that an expert believes possible within a specified degree of credibility, usually 95%), but it is generally accepted that asking for other, less extreme, percentiles may characterise the distribution better. A method that has been applied previously in HTA is the histogram technique or probability grid. The expert is presented with possible values (or ranges of values) of the quantity of interest, displayed in a frequency chart on which he/she is asked to place a given number of crosses in the intervals or bins (Box 1).

Conduct of the Exercise: Preparation and Training of Experts

Training and preparation of experts is a key part of any elicitation exercise. Performing face-to-face exercises is usually seen as preferable, possibly because preparation and training can be better delivered. During expert training, it is important to explain the concept of uncertainty: eliciting measures of uncertainty can be complicated, particularly as one wants to ensure that quantities reflect uncertainty in the expected value rather than variability or heterogeneity. It will be useful to present contrasting examples of uncertainty and variability.

Understanding how experts elicit unknown quantities is also useful, in particular how they may use specific principles or methods in order to make the assessment of probability simpler. These heuristics are useful but can sometimes lead to systematic errors. An example of this is anchoring, where the expert relies on an anchor value that does not provide any information about the actual value. Another is hindsight bias that reflects the tendency to overestimate the predictability of past events. Heuristics should be considered when eliciting probabilities, and attempts to minimise bias taken (for example, avoid introducing anchors in formulating elicitation questions).

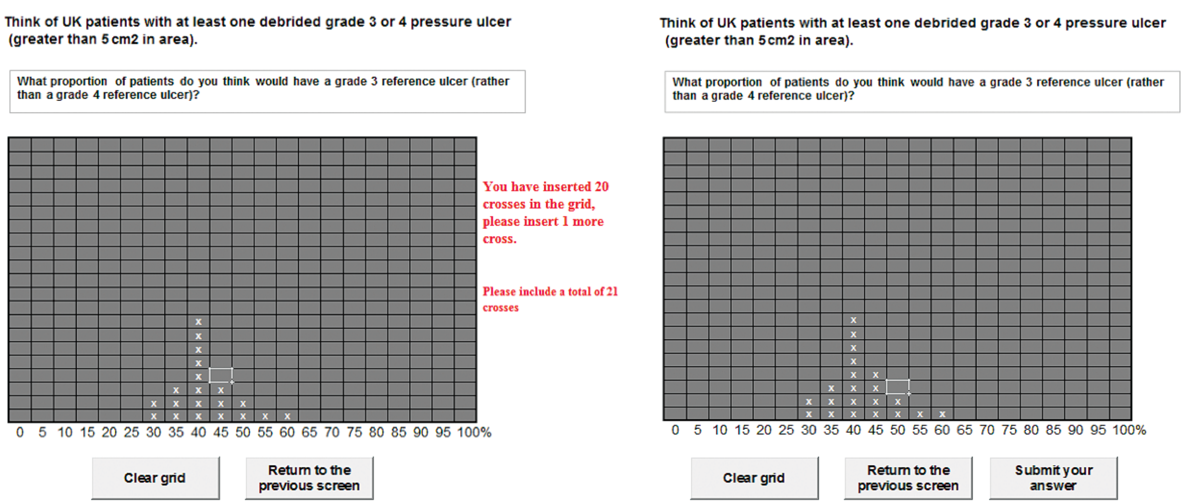
Finally, it may also be helpful to train experts in the method of elicitation chosen and the instrument being used - this is particularly useful when the experts have limited experience with elicitation. Experts will often respond better to questions and give more accurate assessments if they are familiar with the purpose and methods used for the elicitation exercise. Frequent feedback should also be given to the expert during the elicitation process and, if possible, experts should be allowed to revise their judgements.

Synthesizing Multiple Experts' Beliefs

When judgements from several experts are required it is often desirable to obtain a unique distribution that reflects the judgements of them all. Methods to achieve this fall into two categories: behavioural and mathematical. Behavioural approaches focus on achieving group consensus – a group of experts is asked jointly to elicit its beliefs by achieving a level of agreement. >

Box 1. An example of the Histogram Method

For the histogram method, the range of values the quantity of interest may take needs to be partitioned into intervals, and for each interval, information is collected on the probability of observing values. The figure below shows an example of the graphical set-up of the data capture histogram, where a probability parameter was elicited (varying from zero to 100%). This was adapted from an exercise developed by Soares et al. In this exercise, individual experts were asked to place 21 crosses on a grid defined to have 21x21 cells. The expert can express certainty by stacking all of the crosses in the same value (vertical column), or express full certainty that a value is not possible by not attributing any crosses to it. By attributing one cross to each possible value the expert expresses that any value could be possible, that is, full uncertainty.



There are known limitations to consensus methods, namely that an agreement may not be easily achieved, that dominant individuals may determine the view of the whole group and/or that this method tends to generate over-confident replies. Alternatively, mathematical approaches can be used, where individual beliefs are elicited and combined analytically to generate a single distribution. Weighting methods are most commonly used to combine probability distributions analytically, specifically opinion pooling (linear or logarithmic). Weights may be defined through the use of seed questions, which will subjectively determine the contribution of a particular expert judgement to the pooled elicited quantity. An example of the use of linear pooling is described by White et al, 2005, who elicited expert opinion on treatment effects and the interaction between three trials. Experts were asked to assign a weight of belief (up to 100) to intervals of annual event rates. Experts' weights were then combined by taking the arithmetic mean of individual assessments (linear pooling with equal weighting of experts).

Thus, for elicitation to capture such knowledge, it needs to not only describe the experts' best guess, but also any uncertainty experts may have over this value.

Assessment of the Adequacy of an Elicitation Exercise

An elicitation exercise can be judged on its internal consistency, or on experts' performance. Internal consistency is particularly relevant when eliciting probabilities, as experts' assessments need to be consistent with the laws of probability. Achieving coherence may, however, involve complex reasoning, and it can be useful to give qualitative feedback to the expert, with any discrepancies fed back to the experts and appropriate adjustments to assessments made.

Experts' performance can be measured through calibration. The basic premise is that a perfectly calibrated expert should provide assessments of a quantity that are exactly equal to the frequency of that quantity. Experts are thus asked to provide estimates of known parameters, and the distance between elicited estimates and the known distribution can indicate performance. Such measures can then be used to adjust estimates of future unknown quantities. There are many limitations to methods of generating performance weights, and it is unclear whether using such weights improves estimates. These methods have thus been seldom used in HTA.

Conclusion

There are many possible uses for elicitation in HTA. It may inform assumptions and judgments about the conceptualization and structure of a model, as well as to the quality and relevance of data used in the model. But elicitation may be especially useful in generating evidence where data is absent or is inadequate to inform model uncertainties. For example, elicitation can be used to improve generalizability, where the evidence available does not reflect the specificities of the setting or country for decision is to be made.

Elicited evidence is prone to bias, which can only be minimised if a more formal process of elicitation is conducted. This short paper summarised the main choices that an analyst will face when designing and conducting such an exercise. We hope that this will help disseminate the use of these more formal methods in HTA decision models.

Acknowledgements

Thanks to Pedro Saramago, PhD, Research Fellow, Centre for Health Economics, University of York, York, United Kingdom, for doing a technical review of this article.

Suggested Further Reading

Bojke L, et al., Eliciting distributions to populate decision analytic models. Value Health 2010; 13:557-64.
Cooke RM, Experts in Uncertainty: Opinion and Subjective Probability in Science. 1991, New York: Oxford University Press.
Garthwaite PH, Kadane JB, and O'Hagan A, Statistical methods for eliciting probability distributions. J Stat Assoc 2005;100:680-700.
Jenkinson D. The Elicitation of Probabilities-A Review of the Statistical Literature. BEEP working paper. 2005, Department of Probability and Statistics, University of Sheffield: Sheffield.
Leal J, et al. Eliciting expert opinion for economic models. Value Health 2007;10:195-203.
Meyer MA, Booker JM, Bradshaw JM. "A flexible six-step program for defining and handling bias in knowledge elicitation" in Current Trends in Knowledge Acquisition (1990) Wielenga B, Boose J, Gaines B, Schreiber G, Someren MV. ISBN 90 5199 036 7. ■

Additional information:
This article appeared in Volume 2, Issue 3 (August/September 2014) of ISPOR's Latin America Consortium's News Across Latin America

To view the issue of *News Across Latin America* in which this appears, go to: http://www.ispor.org/consortiums/LatinAmerica/articles/vol1issue4_outcomesresearch_english.asp