Introduction to Structured Expert Elicitation: A Risk Analysis Perspective[1]

John S. Evans (Cyprus International Institute and Harvard University)

Abstract

Expert judgment is a method for quantitatively characterizing the state of knowledge about an uncertain quantity. The method seeks to elicit a subjective probability distribution for the quantity of interest from each of several experts; to summarize these distributions and to provide insight about the extent of uncertainty, the sources of the uncertainty, the extent of agreement/disagreement and reasons for any disagreement among the group of experts consulted.

The method is used when the value of the quantity of interest is critical to policy decisions and ordinary statistical approaches cannot provide an answer. This commonly occurs when the issue is not internal validity but instead is about how to extrapolate or generalize findings from the setting in which they were obtained to another setting. Here the central question is about the validity of the analogy which underlies the extrapolation.

The central idea underlying the approach is that experts have information about this and that by using structured protocols to elicit this information one can characterize the state of knowledge, quantify the uncertainty, and identify key sources of uncertainty in the quantity of interest.

Approaches for the elicitation of expert judgment differ somewhat from analyst to analyst, but all include several elements – (i) framing the question; (ii) reviewing the evidence; (ii) identifying and recruiting experts; (iii) development of an elicitation protocol; (iv) informing experts about well-known biases in human judgment; (v) elicitation of judgments; and (vi) analysis, summarization, and presentation of results.

---

[1] Prepared for "Methods for Research Synthesis: A Cross-Disciplinary Workshop," Harvard Center for Risk Analysis, October 3, 2013.

Normally expert judgment elicitations are conducted by teams including at least one substantive expert and one normative expert. Typically the teams work with the problem owner to frame a well-posed question – i.e., a question that could be answered unambiguously by a clairvoyant[2].

Once the question is framed, the elicitation team devotes considerable effort to finding and understanding the evidence relevant to answering the question. Experts are then identified. In some cases the problem owner is simply asked to identify experts. In other cases, more formal methods are relied on – such as publication counts, peer nomination, or identifying scientists who serve on relevant editorial boards or NAS committees. The question of how to identify what "expertise" one is seeking is central and often has no clear answer. Any process for selection of experts must also deal with the question of "balance" – e.g., disciplinary, institutional, and political. Whether to seek, and if so, how to achieve, balance is sometimes controversial. Another issue is determining how many experts are needed / are optimal.

Once a group of experts has been identified, recruited and has agreed to participate, the team turns its attention to development of the elicitation protocol. Most protocols begin with a relatively unstructured discussion of the subject – "Tell me about … What evidence would be ideal? … What evidence exists? … What's missing? …" They then move on to highly structured questions about the quantity of interest. Each elicitation team must decide whether to ask a direct question – e.g., "What is the value of X?" or instead whether to decompose the problem into its logical elements and ask a series of questions about the values of key intermediate quantities – e.g., "What does X depend on? How? What is the value of A? What is the value of B?"

Typically the development of the final protocol involves collaboration with the experts. One approach is to hold a workshop, normally a day or two in length. During the workshop the preliminary protocol can be presented to the group of experts, discussed and refined. An additional benefit of the workshop is that it provides an opportunity to present information to

[2] Note that some analysts, e.g., Cooke, dislike the "clairvoyance" terminology and instead frame the issue as whether the quantity of interest is well-enough characterized that its value could be determined, hypothetically, by an experiment.

the experts about the many known biases in human judgment (availability, anchoring, overconfidence, …) and the approaches that will be used in the elicitations to minimize these. An additional advantage of the workshop is that it allows experts to share recent information about the subject of concern and to freely discuss issues in the interpretation of available evidence.

Once the final protocol has been established the project enters the elicitation phase. In most elicitations, the project team conducts individual interviews with each expert in the expert's office. This allows the expert access to his files and data. Most elicitation teams prefer this intimate face-to-face approach and believe that it fosters open and honest communication. However it is costly. Recent advances in remote communication (SKYPE, GoToMeeting) and in elicitation support software have made it possible to conduct the elicitations at a distance and offer the prospect of significant cost and time savings.

One aspect of the protocol not yet discussed is how to elicit the quantities of interest and whether to use elicitation aids such as probability wheels. To minimize overconfidence, most decision analysts prefer to begin by asking the subject first to think of the extremes – "What is the highest (lowest) value that you can imagine? How could this occur?" Once the expert has given his estimates, it is not uncommon to then ask them to imagine that they had been away from the field for several years and upon return learned that the true value was just a little higher (lower). The expert is then asked whether he can provide an explanation of how this might happen. Normally the elicitation then moves from the extremes toward the center of the distribution. While there are cases where probability wheels and other elicitation aids are helpful, many analysts find that when working with epidemiologists and other quantitatively oriented scientists such tools are not necessary or helpful. Typically at some point during the elicitation experts are shown their results; asked whether these reflect their views; and are given the opportunity to adjust their answers to correct any obvious errors or inconsistencies.

A key question in the reporting of results is whether, and if so how, to combine the judgments of the participating experts. Regardless of whether the results will be combined, it is important for any summary of results to begin by presenting the individual probability distributions given

by each of the experts. If the experts agree on both the central tendency and extent of uncertainty in the quantity of interest little remains to be done. If instead the experts provide substantially different assessments it is important to discuss the reasons for these differences. This is where careful documentation of each expert's rationale is essential.

If there is substantial disagreement among the experts the question arises of whether, and if so how, to combine their judgments. Many approaches have been proposed – including equal weighting, peer weighting, and performance weighting. Cooke's so-called "classical" approach uses information about each expert's performance on a series of calibration questions to create weights reflecting both information (accuracy) and calibration (precision). Experts differ not only in their substantive expertise but also in their ability to encode probability information. Both attributes are important in the interpretation of any probability distributions that they provide. Morgan asserts that when experts disagree there is little benefit to formal combination of results and that, in this case, providing the full set of individual results to the decision maker is all that is needed.

My talk will describe and discuss each of these issues and will conclude with a brief summary of recent efforts to use expert judgment to characterize uncertainty in problems in environmental health.