OPINION

A route to more tractable expert advice

There are mathematically advanced ways to weigh and pool scientific advice. They should be used more to quantify uncertainty and improve decision-making, says **Willy Aspinall**.

When a volcano became restless on the small, populated island of Montserrat, West Indies, in 1995, there was debate among scientists: did the bursts of steam and ash presage an explosive and deadly eruption, or would the outcome be more benign? Authorities on the island, a British overseas territory, needed advice to determine warning levels, and whether travel restrictions and evacuations were needed. The British government asked me, as an independent volcanologist, to help reconcile differing views within the group.

As it happened, I had experience not only with the region's volcanoes, but also with a unique way of compiling scientific advice in the face of uncertainty: the Cooke method of 'expert elicitation'. This method weighs the opinion of each expert on the basis of his or her knowledge and ability to judge relevant uncertainties[1]. The approach isn't perfect. But it can produce a 'rational consensus' for many hard-to-assess risks, from earthquake hazards to the probable lethal dose of a poison or the acceptable limits of an air pollutant. For broader questions with many unknowns — such as what the climate will be like in 50 years — the method can tackle small specifics of the larger question, and identify gaps in knowledge or expose significant disparities of opinion.

As a group, myself and ten other volcanologists decided to trial this methodology, making it the first time a formal elicitation procedure was used in a live volcanic crisis. We were able to provide useful guidance to the authorities, such as the percentage chance of a violent explosion, as quickly as within an hour or two. Without these prompt forecasts, more lives may have been put at risk[2]. More than 14 years on, volcano management in Montserrat stands as the longest-running application of the Cooke method.

Faced with uncertainty, decision-makers invariably seek agreement or unambiguous consensus from experts. But it is not reasonable to expect total consensus when tackling difficult-to-predict problems such as volcanic eruptions. The method's originator Roger Cooke, now at Resources for the Future, a think tank in Washington DC, says that when scientists disagree, any attempt to impose agreement will "promote confusion between consensus and certainty". The goal should be to quantify uncertainty, not to remove it from the decision process.

Of the many ways of gathering advice from experts, the Cooke method is, in my view, the most effective when data are sparse, unreliable or unobtainable.

Rational consensus

Advice during an emergency is usually the responsibility of a chief scientist, with all the stresses that involves — including the pressure to be extremely cautious. There is a better way: pooling the opinions of a group of specialists.

There are several methods of such expert elicitation, each with flaws. The traditional committee still rules in many areas — a slow, deliberative process that gathers a wide range of opinions. This has parallels with the scientific process itself. But committees traditionally give all experts equal weight (one person, one vote). This assumes that experts are equally informed, equally proficient and free of bias. These assumptions are generally not justified.

Another kind of elicitation — the Delphi method — was developed in the 1950s and 1960s. It was used in the early 1970s to try to predict the impact of rising oil prices on the US economy, and more recently to identify the top 20 research priorities in chronic health problems for the Grand Challenges in Global Health programme. This involves getting 'position statements' from individual experts, circulating these, and allowing the experts to adjust their own opinions over multiple rounds. What often happens is that participants revise their views in the direction of the supposed 'leading' experts, rather than in the direction of the strongest arguments. Another, more recent elicitation method recommended by the US Senior Seismic Hazard Analysis Committee involves asking each participating expert to predict the range of uncertainty estimates of any and every person in the field. This creates a huge spread of overall uncertainty, and sometimes physically implausible results: it once implied higher earthquake risk in Switzerland than in California. Taking a precautionary approach to guard against such implausible worst-case scenarios does not provide a rational basis for workable, effective policies.

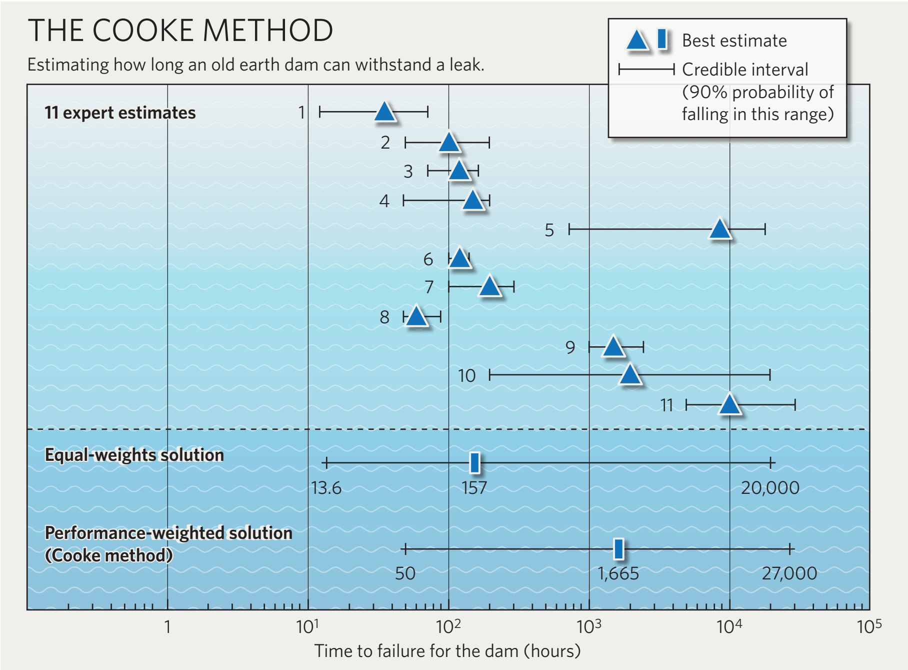
> "The goal should be to quantify uncertainty, not to remove it from the decision process."

Cooke's method, developed in 1991 at the Delft University of Technology in the Netherlands with his colleagues, instead produces a 'rational consensus'. To see how this works, take as an example an elicitation I conducted in 2003, to estimate the strength of the thousands of small, old earth dams in the United Kingdom. Acting as facilitator, I first organized a discussion between a group of selected experts about how water can leak into the cores of such ageing dams, leading to failure. The experts were then asked individually to give their own opinion of the time-to-failure in a specific type of dam, once such leakage starts. They answered with both a best estimate and a 'credible interval', for which they thought there was only a 10% chance that the true answer was higher or lower.

I also asked each expert a set of eleven 'seed questions', for which answers are known, so that their proficiency could be calibrated. One seed question, for example, asked about the observed time-to-failure of the Teton Dam in Idaho in June 1976. Again the specialists answered with a best estimate and a credible interval. Their performance on these seed questions was used to 'weight' their opinion, and these weighted opinions were pooled to provide a 'rational consensus' (see Fig. 1).

For the UK dams, the performance-weighted solution indicated a best estimate of time-to-failure of 1,665 hours (70 days) — much higher than if all of the answers were pooled with equal weights (6.5 days). As is often the case, several experts were very sure of their judgement and provided very narrow uncertainty ranges. But the more cautious experts with longer time estimates and wider uncertainty ranges did better on the seed questions, so their answers were weighted more heavily. Their views would probably have been poorly represented if the decision had rested on a group discussion in which charismatic, confident personalities might carry the day. Self-confidence is not a good predictor of expert performance, and, interestingly, neither is scientific prestige and reputation[3].

The Cooke method generally produces uncertainty spreads that are narrower than the 'democratic' pooling approaches, but wider than those provided by single experts.



SOURCE: A. J. BROWN & W. P. ASPINALL PROC. 13TH BIENNIAL BRIT. DAMS SOC. CONF. (2004)

Figure 1 | Estimates from 11 experts of the time-to-failure of an earth dam, once the core starts to leak. The performance-weighted best judgement is about 70 days — much longer than the equal-weights solution of about a week.

The speed with which such elicitations can be conducted is one of their advantages. The lethality of three biological agents was assessed and quantified in a two-day elicitation, with a few days of preparatory work. Another advantage is that it encourages experts wary of getting involved in policy advice: the structured, neutral procedure, and the collective nature of the result, reassures experts and relieves them of the burden of sole responsibility.

The facilitator needs to be impartial and able to deal with differences of opinion with tact, to prevent individual experts from hijacking the debate, and to avoid other pitfalls, such as asking misconstrued or leading questions, or prompting the experts. At the same time, he or she needs to keep all the participants fully engaged and focused.

Two tribes

Sometimes an elicitation reveals two camps of opinion, as was the case with the dam-failure example. This usually arises from ambiguity in framing the problem or because the two sub-groups have subtly different backgrounds. In this case, a few of the experts with longer time-to-failure estimates were working engineers with practical experience, rather than academics. Highlighting such clear differences is useful; sometimes it reveals errors or misunderstandings in one of the groups. In this case it triggered further investigation.

Structured elicitations have been used for difficult regulatory decisions, such as formulating US Nuclear Regulatory Commission guidelines on the safety of nuclear materials and technologies. The US Environment Protection Agency has formed a task force to recommend ways of using expert elicitations in regulatory issues whenever data are sparse, such as deciding dose criteria for air-quality standards. It has also been suggested that the Intergovernmental Panel on Climate Change use expert elicitations to ensure that areas of uncertainty poorly captured by models are better represented, such as whether the Greenland ice cap might melt[4].

The Cooke approach is starting to take root in similar problem areas, including climate-change impacts on fisheries, invasive-species threats and medical risks. In Canada, policy-makers considering the risks of variant Creutzfeldt–Jakob Disease infection through contaminated blood supplies are utilising scientific advice from a Cooke elicitation. Participants found the elicitation "extremely useful", and suggested that it would be helpful in assessing other public-health issues, including pandemic flu and vaccination risks[5].

There are still research questions to answer if performance-based expert elicitation is to be used more widely. The selection and number of experts need careful thought. The 'problem owner' may want a panel to be as impartial as possible to prevent accusations of bias, or they may restrict the selection to company employees to ensure confidentiality and relevance. My experience with more than 20 panels suggests that 8–15 experts is a viable number — getting more together will not change findings significantly, but will incur extra expense and time. However, this has not been rigorously tested. We also need more information on what makes good seed questions.

Perhaps the biggest challenge facing wider adoption of structured expert elicitations lies in communicating the resulting scientific uncertainty to policy-makers. Politicians, being averse to ambiguity in their solicited advice, do not like it when expert elicitations produce wide margins of uncertainty. 'Better' information from an expert perspective — more complete in quantifying uncertainty, but more complex overall — may confound some decision-makers rather than promoting better choices. Improvements in capturing uncertainty must be matched with improvements in communication. Fortunately, studies suggest that most decision-makers welcome more information, not less, especially if presented in context, alongside training in uncertainty concepts[6].

Uncertainty in science is unavoidable and throws up many challenges. On one side, you may have scientists reluctant to offer their opinions on important societal topics. On the other, you may have decision-makers who ignore uncertainty because they fear undermining public confidence or opening regulations to legal challenges. The Cooke approach can bridge these two positions. It is most useful when there is no other sensible way to make risk-based decisions — apart from resorting to the precautionary principle, or, even less helpfully, evading an answer by agreeing to disagree. ■

Willy Aspinall is Cabot professor of natural hazards and risk science at Bristol University, Bristol BS8 1RJ, UK. He provides expert elicitation services through his company Aspinall & Associates, Cleveland House, High Street, Tisbury SP3 6HF, UK.
e-mail: Willy.Aspinall@Bristol.ac.uk

1. Cooke, R. M. *Experts in Uncertainty: Opinion and Subjective Probability in Science* (Oxford Univ. Press, 1991).
2. Aspinall, W. & Cooke, R. M. *Proc. 4th Int. Conf. Probabilistic Safety Assessment and Management* (eds Mosleh, A. & Bari, R. A.) **3,** 2113–2118 (Springer, 1998).
3. Cooke, R. M., ElSaadany, S. & Huang, X. *Reliab. Eng. Syst. Safe.* **93,** 745–756 (2008).
4. Oppenheimer, M., O'Neill, B. C., Webster, M. & Agrawala, S. *Science* **317,** 1505–1506 (2007).
5. Tyshenko, M. G. & Darshan, S. *Summary Report of the Expert Elicitation Workshop Results for Iatrogenic Prion Disease Risks in Canada* (Univ. Ottawa, 2009); available at http://www.prionetcanada.ca/files/Expert_Elicitation_Workshop_Summary-Nov._18-093188.pdf
6. Krupnick, A. J. *et al. Not a Sure Thing: Making Regulatory Choices Under Uncertainty* (Resouces for the Future, 2006); available at http://www.rff.org/rff/Documents/RFF-Rpt-RegulatoryChoices.pdf

See Editorial, page 269.