Advances in Reliability, Safety and Security, Part 1 – Kołowrocki, Dąbrowska (eds.)
 ISBN 978-83-68136-13-5 (printed), ISBN 978-83-68136-00-5 (electronic)

Advances in Reliability, Safety and Security

ESREL 2024
Monograph Book Series

IDEA Protocol For Structured Expert Judgement Elicitation

Anca Hanea

Centre of Excellence for Biosecurity Risk Analysis, The University of Melbourne, Mlbourne. Autralia

Abstract

Expert judgement may be required to inform a range of tasks under uncertainty, including model development, estimates of probabilities and quantities, and to inform prioritisation tasks. In this workshop we concentrate on the elicitation and aggregation of expert judgements related to uncertain events and quantities. For quantitative estimates, a common approach is to elicit a point estimate. However, there are strong theoretical and practical arguments to say that the proper representation of experts' knowledge about uncertain quantities is through probability distributions. Repeated evidence also indicates that these judgements should be elicited from multiple experts. However, challenges arise when the model requires a single probability distribution, which means that the various judgements must be aggregated. This aggregation can be done by the experts themselves, through a process of interaction that is designed to encourage consensus (behavioral aggregation). Alternatively, it may be done externally, by applying an aggregation formula (mathematical aggregation). I will introduce and motivate a third (combined) way of aggregation which combines elements of the behavioral approach with mathematical aggregation. This is done through the IDEA protocol for structured expert judgement. At the end of this workshop participants will be familiar with the IDEA protocol. They will benefit from short hands-on exercises, lecture style explanations, a list of relevant literature, and relevant contacts in the field. The workshop is aimed at professionals, academics, policymakers, regulators, and (MSc, PhD) students who are, or will soon be involved in decision problems or risk analysis modelling with scarce resources, and insufficient data.

Keywords: uncertainty quantification, data-sparse environments, expert judgement , structured protocols, IDEA

1. Introduction

Expert knowledge in the form of numerical estimates and the rationale behind such estimates is often used to complement hard data collection. More often than not, this is done through structured questionnaires, or unstructured methods rather than structured expert elicitation protocols.

The use of expert judgement is however difficult when intended as scientific quantitative input to risk analysis and decision models. More difficulties arise when decisions have to be made under uncertainty and expert judgements are needed to quantify such uncertainty. Like any other type of data, expert elicited data is prone to errors and biases (most important of which are detailed below).

1.1. Individual and group biases

Individuals are known to be subject to several psychological frailties and cognitive biases, especially when reasoning under uncertainty. When expert judgements are sought, a few elements that may influence the data/answers collected from experts are listed below:

- Framing - The way a question is presented or framed can influence the response received (e.g., Tversky et al., 1981).
- Availability - Judgements of probability may be influenced by the ease with which an event is recalled. Events may be judged to have a higher probability of occurrence than is really the case if they are recent, invoke strong emotions or have been widely reported (e.g., Gigerenzer, 2004).

- Anchoring and adjustment - The tendency to base quantitative estimates on values that been previously suggested or estimated coupled with an inability to adjust sufficiently far from the anchor (e.g., Epley et al., 2006).
- Confirmation bias - People search for or interpret information in a way that confirms their prior beliefs.
- Overconfidence - When experts have unwarranted confidence in their own judgments (e.g., Oskamp, 1965).

Using a group of experts instead of just one individual draws on a wider range of experience than could be achieved with just a single individual and avoids judgements that reflect only the cognitive biases of a single individual. Of course, group assessments are also subject to their own suite of biases, including:

- Group think - Judgements are unduly affected by a desire for agreement within the group (e.g., Janis, 1997).
- Dominance - Individual experts are unduly influenced by the views of a senior or dominant group member (e.g., Maier, 1967).
- Halo effects - Perceptions of the opinion of one expert are influenced by the perception of attributes of that expert unrelated to the subject under consideration (e.g., Nisbett et al., 1977).

1.2. The expert group

One strategy to overcome (to some extent) both individual and group biases is to recruit a diverse group of experts. Even though it is hard to identify experts a-priori, one can be guided by the level of domain expertise and the ability of individuals to adapt and communicate this expertise under uncertainty. Diversity in terms of demographics (i.e., age, experience, gender) is known to be a good proxy for cognitive diversity (e.g., Page, 2008). Involving individuals with diverse experience and background increases the knowledge pool and provides a better representation of the modelled problem. Aggregates of a diverse group are proven to be more accurate and better calibrated than a single well credentialed individual (e.g., Hemming et al., 2020).

The optimum number of experts in a group is informed by the improvement in group performance (when averaging judgements) when the group size increases. Simulations studies and experiments found that between 5 and 12 experts are sufficient for obtaining a robust group judgement (e.g., Vercammen et al., 2019). Any smaller groups will be too sensitive to outliers, and any larger groups will not change the average judgement significantly. When the elicitation includes extensive discussion between experts, larger groups are also more cumbersome to manage. The current advise, for practical reasons, is anything between 4 and 10 (e.g., Hanea et al., 2021a).

Other strategies to minimize the impacts of the potential biases are incorporated in formal expert elicitation protocols, also called structured expert judgement (SEJ) protocols.

SEJ protocols follow a series of steps akin to the process of formal data acquisition, while trying to mitigate against cognitive individual and group biases. This enhances the transparency, accuracy, and defensibility of results (e.g., Dias et al., 2018).

2. SEJ protocols

A formal, structured expert elicitation consists of a collection of steps and advice defining a so-called protocol. A structured protocol includes the judgements of more than a single expert, avoids round table free discussions, and it is designed to answer hypothetically verifiable questions. Not only what, but also how we ask the experts plays an important role in a structured protocol. Allowing experts to quantify their own uncertainty around point estimates is an essential step.

Any scientific method for expert elicitation must satisfy the following set of principles (e.g., Hanea et al., 2021b):

- *Accountability / Traceability*: all data, including experts' names, affiliations and assessments, and all analyses and software tools are available to open peer review. This guarantees that the results are reproducible by competent reviewers without violating confidentiality as experts' names are never linked to individual assessments in any open documentation.
- *Empirical control*: In addition to the questions of interest (for which the answer is not yet known, but needs to be estimated), experts quantify their uncertainty with respect to calibration or seed questions, that is, questions about quantities from their field for which true values are retrievable. Experts' performance as uncertainty assessors is measured based on calibration questions.

- *Neutrality*: The method for evaluating and combining expert assessments should encourage experts to state their true opinion.
- *Fairness*: Experts are not pre-judged prior to empirical control.

Some other key elements of SEJ protocols include: eliciting individual private estimates, ask questions in certain formats, collect both quantitative and qualitative judgements, provide feedback and facilitate expert interaction, use an aggregation technique that summarises the inter- and intra-.expert variation. All of these try to mitigate as much as possible both the individual and the group biases discussed in Section 1.1. Established structured protocols that incorporate these elements are described in (e.g., O'Hagan, 2019; Hanea et al., 2021a). One of them, the IDEA protocol, is introduced in the next section.

2.1. The IDEA protocol

Briefly, the IDEA protocol involves four stages: **I**nvestigate, **D**iscuss, **E**stimate, and **A**ggregate, as outlined below:

- **I**nvestigate In this first phase, experts work individually using whatever resources are available to them to formulate point and interval (low---high) estimates of the unknown quantities. This step should be preceded by a conference call or meeting with all experts to identify ambiguities, agree on meanings, and affirm the elicitation approach and schedule. Expert-elicited estimates are then anonymised and summarised by the facilitation team, to be used as spurs for discussion in phase 2.
- **D**iscuss
  The second phase is best carried out as a face-to-face workshop. Summaries of the anonymised phase-1 responses are presented to the group, with invitations to discuss the distribution of results --- why might there be disagreement between experts, what might be reasons for particularly low or high values, etc.
- **E**stimate
  The experts are then invited to privately update their first-phase point and interval estimates. They may choose not to do so. They are not obliged to disclose either way.
- **A**ggregate
  The estimates from phase 2 are then aggregated within expert (to the extent that it makes sense to do so) and then across expert. For example, if the outcome of the exercise is to produce estimates of parameters for a structural model, for the purposes of making predictions, then a suite of predictions should be made for each expert, which should then be aggregated.

The authors of Hemming et al. (2017) provide a comprehensive description of the process, but a visual representation is presented in Figure 1. The first step involves individual assessments collected from each expert (working without consultation). Often the interval estimates that surround their point estimate are elicited in terms of distributional percentiles. In this way, the uncertainty expressed by experts can be captured by a subjective probability distribution. The upper bound is often interpreted as the 95^{th} percentile of the expert's uncertain distribution, whereas the lower bound is taken to be the 5^{th} percentile of that distribution. The point estimate (often called the best estimate) is interpreted as a median, the 50^{th} percentile. Feedback plots with anonymised estimates like the one in the middle panel of Figure 1 are then presented to the group of experts and discussion is sought. The discussion points are centred around differences and similarities between the estimates. After the group discussion, experts return to individual work where they have the chance to revise their initial estimates. In Figure 1, on the right-hand side panel, the dotted lines represent initial estimates, and the solid lines represent revised estimates.

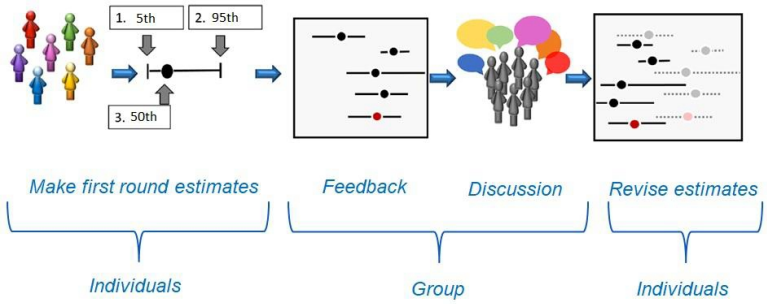

Fig. 1. The IDEA protocol steps. Adapted from: Hemming V, Walshe TV, Hanea AM, Fidler F, Burgman MA. 2018. Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. PLoS ONE.

2.2. Applications of the IDEA protocol

The IDEA protocol was developed by a multidisciplinary team of researchers at the University of Melbourne and adopted internationally throughout the years. IDEA was applied in various domains, like biosecurity, ecology, engineering, and food security, to name a few. Some applications of quantitative data obtained using the IDEA protocol are outlined below:

- Expert opinion contributed to the development of a surveillance system for black rats on an island of high conservation value (Jarrad et al., 2011). The experts estimated parameters where data were lacking for the model used to optimise surveillance.
- Experts in fish health were used to establish likelihood ratios for risk factors that might influence the occurrence of infectious *salmon anemia* in Chilean salmon farms (Gustafson et al., 2014).
- About 20 experts of diverse backgrounds and affiliations participated in an elicitation exercise on the effects of collisions on North Atlantic right whales (Fleishman et al., 2016). The experts first contributed to the development of a conceptual model of collision effects, then provided estimates of the principal parameters needed to populate the model.
- A panel of 15 active koala researchers participated in a structured expert elicitation to estimate the total koala population of eastern and south-eastern Australia and changes in populations within different bioregions (Adams-Hosking et al., 2016). The entire elicitation process took six months, including a four-day workshop for the discussion phase.
- The authors of (Hanea et.al, 2022) describe an iterative and collaborative approach that uses structured methods for eliciting, combining, and using expert judgements to build and parametrise a local risk model for an outbreak of *Bonamia ostreae* in the New Zealand oyster *Ostrea chilensis*.
- Food Standards Australia New Zealand (FSANZ) was interested to estimate, using expert judgement, the attribution of foodborne illness to specific horticultural foods for pathogens like: non-typhoidal *Salmonella* species, shiga-toxin *Escherichia coli*, *Listeria monocytogenes* and norovirus. The combination of pathogen specific foodborne illness cost estimates and expert elicitation attribution proportions was further used to provide initial estimates for costings to food commodities. The elicitation used the IDEA protocol and calibration questions to satisfy the *Empirical control* principle. Adding calibration questions to an IDEA protocol steps and timeline is not standard, but it can be done. Calibration questions should however be chosen carefully in settings where the answers may be found by the experts and the first round is done remotely over a longer period of time. The timeline for the FSANZ elicitation is outlined in Figure 2. While the calibration questions were answered in the inception meeting (to avoid finding available answers) the questions of interest were only presented to the experts during the inception meeting. They were given a month to answer the questions whenever they could fit it in their schedule. Answers were collected, de-identified and collated into feedback plots prior to the workshop where discussion per question was facilitated. After each question was discussed, the participants had a couple of minutes to update their estimates (to represent their round 2 estimates) if they wanted to.

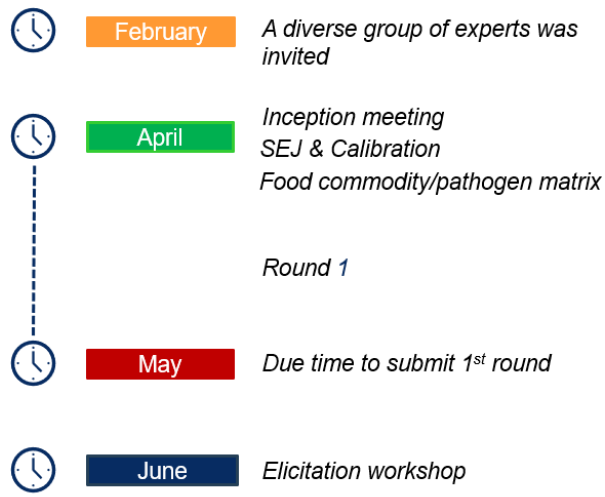

Fig. 2. Timeline for an elicitation following the IDEA protocol structure.

Using calibration questions allows us to measure performance relative to the known answers of the questions. It also allows to investigate how the discussion between the rounds influences the revised estimates from the second round. Figure 3 shows the assessments of ten experts answering a calibration question.

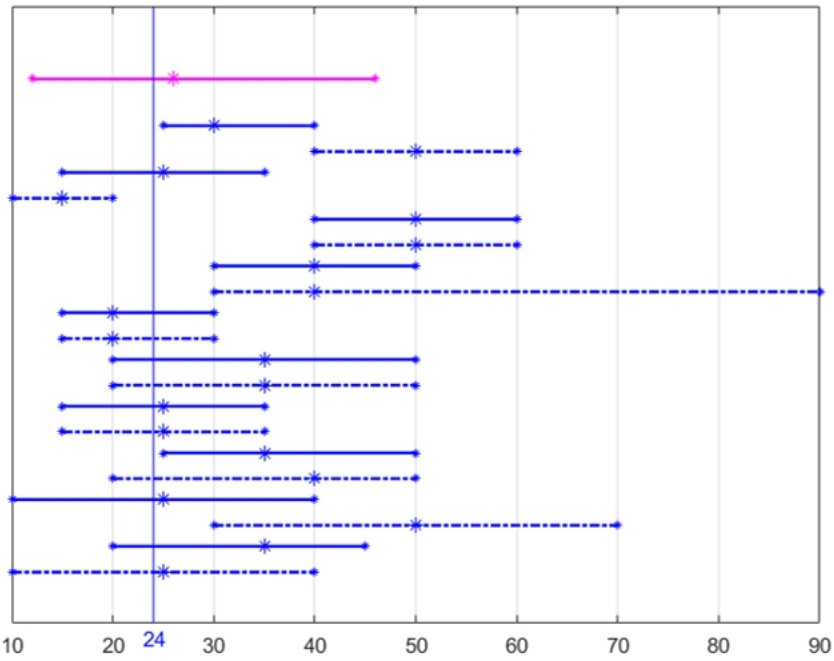

Fig. 3. Feedback plot for both rounds of elicitation, with experts' assessments represented as horizontal lines and the true value marked with a vertical line at the appropriate position.

The estimates from Figure 3, are answers to the questions formulated as follows:

> *"Food Standards Australia New Zealand purpose is to contribute to the bi-national food regulation system by developing science and evidence-based standards, coordinating regulatory responses e.g. recalls and incidents as well as providing information to the public about food standards. A food recall is undertaken by a food business to remove unsafe food from distribution, sale and consumption. For the years 2020 and 2021 what percentage of food recalls were due microbial contamination?"*

Each pair of dashed and solid lines from Figure 3 corresponds to one expert assessment for round 1(dashed) and round 2 (solid, given after the feedback and group discussion). The vertical line marking the value 24 on the x-axis represents the true answer (24% of food recalls were due to microbial contamination) not known to the experts at the moment of the elicitation. It is worth noting that most of the experts who changed their estimates after discussion (as shown by the shifted solid lines relative the dashed ones) did so towards the direction of the truth. The magenta line at the top of the graph represents the performance based weighted combination of second round estimates which represents the group's judgements.

3. Structure of the workshop

In this workshop we shall concentrate on the elicitation and aggregation of expert judgements related to uncertain events and quantities, collected through an IDEA protocol.

After lecture style explanations, participants will be asked to answer questions that will exemplify one of the potential individual biases described in Section 1.1. Due to the restricted time of the workshop only one of the individual biases will be exemplified, namely the anchoring bias. Participants will be split in two groups and will

have to answer questions formatted such that the bias is triggered in a different way in each group. Group biases will be discussed, and maybe identified in a second exercise that is meant to mimic the conditions of a real elicitation.

Fig. 4. Jar used as a toy example question for estimation.

The questions for the example elicitation however will be playful, given the diverse background and interests of the participants. The estimated unknown quantity will most likely regard general knowledge questions or visual aids, like for example participants may be asked about the number of jellybeans in a jar like the one in Figure 4. This is meant to trigger some of the thinking mechanisms experts use in real elicitation to answer questions under uncertainty. Evaluation and estimation based on the size of the beans and that of the jar can be attempted. Discussion and cross checking of rationales will be possible in a discussion stage of the IDEA protocol. Discussion will be facilitated by me and in this way, the role of a facilitator and the skills they need will be exemplified and discussed.

At the end of this workshop participants will be familiar with most of the aspects of the IDEA protocol. They will benefit from short hands-on interactive exercises and lecture style explanations that will cover all the basics of running such an elicitation. Moreover, the participants will be provided with a list of relevant literature, and relevant contacts in the field if needed.

3.1 Take home messages

The most important issues participants will be able to identify after participating in this workshop are:

- when is IDEA needed, and in general, when does one need to recourse to expert judgement and how extensive the elicitations need to be relative to the modelling needs and data gaps;
- what are the scientific principles underpinning SEJ in general and IDEA in particular, and why elicited data can be used to complement “hard” data;
- who needs to be involved in an elicitation and what are the different roles key players need to fulfill;
- how to run an elicitation with the IDEA protocol, how long it may take and what organisatorical steps need to be undertaken; and, if time allows ,
- which are the available tools that may aid running a SEJ with IDEA, especially in online settings.

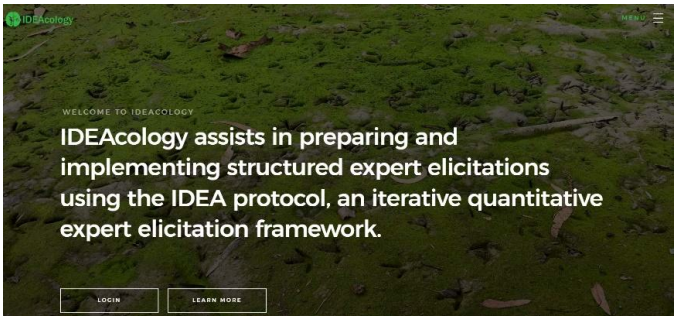

Fig.5. The IDEAcology platform for running online elicitations with the IDEA protocol.

Figure 5 shows one of the online tools participants may use when engaging in an IDEA elicitation. The IDEAcology interface (available at https://www.ideacology.com) supports the expert data collation and summary, and the expert interactions and feedback essential in a structured elicitation using the IDEA protocol. For more details about IDEAcology, I refer the participants to (Courtney Jones et al. 2023). If time allows, a small demonstration will also be given during the workshop.

3.2 Targeted audience

The workshop is aimed at professionals, academics, policymakers, regulators, and (MSc, PhD) students who are, or will soon be involved in decision problems or risk analysis modelling with scarce resources, and insufficient data. Depending on the number of participants, their level of engagement and their expectations, hands-on exercises may be done in several sub-groups or with the entire cohort. Selective material will be presented and tailored to the audience.

References

Adams-Hosking, C., McBride, M., Baxter, G., Burgman, M., Villiers, D., Kavanagh, R., Lawler, I., Lunney, D., Melzer, A., Menkhorst, P., Molsher, R., Moore, B., Phalen, D., Rhodes, J., Todd, C., Whisson, D., McAlpine, C. 2016. Use of expert knowledge to elicit population trends for the koala (Phascolarctos cinereus). Diversity and Distributions 22(3), 249–262.

Courtney Jones, S.K., Geange, S.R., Hanea, A., Camac, J., Hemming, V., Doobov, B., Leigh, A., Nicotra, A.B. 2023. IDEAcology: An interface to streamline and facilitate efficient, rigorous expert elicitation in ecology. Methods in Ecology and Evolution, 14, 2019–2028.

Dias, L.C. (Ed.), Morton, A. (Ed.), Quigley, J. (Ed.), 2018. Elicitation: The Science and Art of Structuring Judgement. International Series in Operations Research & Management Science, Springer, Cham.

Epley, N., Gilovich, T. 2006. The anchoring-and-adjustment heuristic. Psychological Science 17, 311–318.

Fleishman, E., Burgman, M., Runge, M., Schick, R., Kraus, S. 2016. Expert elicitation of population-level effects of disturbance. Popper, A. (Ed.), Hawkins, A. (Ed.), The Effects of Noise on Aquatic Life II, Springer, Cham, 295–302.

Gigerenzer, G. 2004. Dread risk, september 11, and fatal traffic accidents. Psychological Science 15, 286–287.

Gustafson, L., Antognoli, M., Lara Fica, M., Ibarra, R., Mancilla, J., Sandoval, O., Enriquez, R., Perez, A.,Aguilar, D., Madrid, E., Bustos, P., Clement, A., Godoy, M., Johnson, C., and Remmenga, M. 2014. Risk factors perceived predictive of ISA spread in Chile: Applications to decision support. Preventive Veterinary Medicine 117, 276–285.

Hanea, A., Hemming, V., Nane, G. 2021a. Uncertainty quantification with experts: Present status and research needs. Risk Analysis 42 (2), 254-263.

Hanea, A., Nane, G. 2021b. An in-depth perspective on the classical model. Hanea, A. (Ed.), Nane, G. (Ed.), Bedford, T. (Ed.), French, S. (Ed.), Expert Judgment in Risk and Decision Analysis. International Series in Operations Research & Management Science, Springer, Cham, 225-256.

Hanea, A., Hilton, Z., Knight, B., Robinson, A. 2022. Co-designing and building an expert-elicited non-parametric Bayesian network model: demonstrating a methodology using a Bonamia Ostreae spread
risk case study. Risk Analysis, 42 (6), 1235-1254.

Hemming, V., Burgman, M.A., Hanea, A., McBride, M., Wintle, B. 2017. A practical guide to structured expert elicitation using the IDEA protocol. Methods in Ecology and Evolution 9(1), 169–180.

Hemming, V., Hanea, A., Walshe, T., Burgman, A. 2020. Weighting and aggregating expert ecological judgements. Ecological Applications, 30(4):e02075. 10.1002/eap.2075.

Janis, I. L. 1997. Groupthink. Vecchio,R.P. (Ed.), Leadership: Understanding the dynamics of power and influence in organizations, University of Notre Dame Press, Notre Dame, Indiana, USA, 163–176.

Jarrad, F., Barrett, S., Murray, J., Parkes, J., Stoklosa, R., Mengersen, K., and Whittle, P. 2011. Improved design method for biosecurity surveillance and early detection of non-indigenous rats. New Zealand Journal of Ecology 35, 132–144.

Maier, N. 1967. Assets and liabilities in group problem solving: the need for an integrative function. Psychological Review 74, 239–249.

Nisbett, R., Wilson, T. 1977. Halo effect: evidence for unconscious alteration of judgments. Journal of Personality and Social Psychology 35, 250–256.

O'Hagan, A. 2019. Expert knowledge elicitation: subjective but scientific. The American Statistician, 73(sup1):69–81.
Oskamp, S. 1965. Overconfidence in case-study judgments. Journal of Consulting Psychology 29, 261–65.
Page, S. 2008. The difference: How the power of diversity creates better groups, firms, schools, and societies. Princeton University Press, Princeton, New Jersey
Tversky, A., Kahneman, D. 1981. The framing of decisions and the psychology of choice. Science, 211, 453–458.
Vercammen, A., Ji, Y., Burgman, M. 2019. The collective intelligence of random small crowds: A partial replication of Kosinski et al.(2012). Judgment and Decision Making 14(1), 91-98.