Recommendations on the Use of Structured Expert Elicitation Protocols for Healthcare Decision Making

A Good Practices Report of an ISPOR Task Force

Soares, Marta; Colson, Abigail; Bojke, Laura; Ghabri, Salah; Garay, Osvaldo Ulises; Felli, Jenna K.; Lee, Karen; Morales-Napoles, Oswaldo; IJzerman, Maarten J.; More Authors

DOI
10.1016/j.jval.2024.07.027

Publication date
2024

Document Version
Final published version

Published in
Value in Health

Citation (APA)
Soares, M., Colson, A., Bojke, L., Ghabri, S., Garay, O. U., Felli, J. K., Lee, K., Morales-Napoles, O., IJzerman, M. J., & More Authors (2024). Recommendations on the Use of Structured Expert Elicitation Protocols for Healthcare Decision Making: A Good Practices Report of an ISPOR Task Force. *Value in Health*, *27*(11), 1469-1478. https://doi.org/10.1016/j.jval.2024.07.027

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

ISPOR Report

Recommendations on the Use of Structured Expert Elicitation Protocols for Healthcare Decision Making: A Good Practices Report of an ISPOR Task Force

Marta Soares, PhD, Abigail Colson, PhD, Laura Bojke, PhD, Salah Ghabri, PhD, Osvaldo Ulises Garay, MSc, Jenna K. Felli, PhD, Karen Lee, MA, Elizabeth Molsen-David, RN, Oswaldo Morales-Napoles, PhD, Victoria A. Shaffer, PhD, Maarten J. IJzerman, PhD

ABSTRACT

Healthcare decision making, including regulatory and reimbursement decisions, is based on uncertain assessments of clinical and economic value. This arises from the evidence supporting those assessments being uncertain, incomplete, or even absent. Qualitative, structured expert elicitation (SEE) is a valuable tool for extracting expert knowledge about an uncertain quantity and formulating that knowledge as a probability distribution. This creates a useful input to decision modeling and support, particularly in areas with limited evidence, such as advanced therapy products, precision medicine, rare diagnoses, and other areas with high uncertainty.

Structured SEE protocols are used to improve the transparency, accuracy, and consistency of quantitative judgments from experts, limiting the effect of heuristics and biases. This task force report introduces 5 commonly used protocols for SEE (Sheffield elicitation framework; modified Delphi method; Cooke's classical method; investigate, discuss, estimate, aggregate protocol; and the Medical Research Council reference protocol). It describes the common elements of SEE, discusses how these protocols differ in their implementation of those elements and illustrates the use of the protocols.

The report then reviews the relevant constraints on implementing SEE within the context of healthcare decision making and considers the strengths and weaknesses of these protocols in light of those considerations. Because this is an introductory report on an emerging topic, specific recommendations on practice are not made. However, there are broad recommendations based on the suitability of the different protocols in various decision contexts. The report concludes with recommendations for further research to better guide future practice.

Keywords: Bayesian estimation, IDEA protocol, modified Delphi, SHELF, structured expert elicitation, uncertainty.

VALUE HEALTH. 2024; 27(11):1469–1478

Highlights

- Because no prescriptive guidance for the use of SEE is available, this task force report has identified 5 different protocols (Sheffield elicitation framework; Cooke's classical method; investigate, discuss, estimate, aggregate; modified Delphi and the MRC reference protocol) and compares them on the level of elicitation, individuals or groups, and the mode, with or without interaction of experts and the level of aggregation.
- To prevent behavioral and cognitive biases, healthcare decision making using SEE should carefully consider the specificities of the setting they work in before selecting a particular SEE protocol.
- The task force concludes that many studies using SEE have not documented well the SEE approach used. In particular, if no particular protocol was chosen, the task force urges decision makers to be transparent in the reporting of the methods and results.

Introduction

Healthcare decision making (HCDM), including regulatory and reimbursement decisions, is typically grounded in uncertain assessments of clinical and economic value (See Appendix 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.07.027 for a full list of acronyms). Uncertainty, in part, arises from the evidence supporting those assessments being itself uncertain, incomplete, or even absent. For example, evidence regarding the parameters used in health economic models, such as transition probabilities, is often uncertain. Therefore, resulting cost-effectiveness and budget impact for a new technology are also uncertain.

The uncertainty surrounding decision making is particularly challenging in emerging medical fields with only limited evidence, such as advanced therapy products,[1] in the field of precision medicine where evidence usually is difficult to obtain through clinical trials[2] or where regulators must extrapolate evidence obtained in larger populations to a smaller (genomically defined) subgroup or to populations that fall outside the clinical trial inclusion criteria.[3]

Uncertainty can also be the result of a small number of patients (rare diagnoses), limited duration of follow-up, or trial design (use of surrogate outcomes or lack of an active comparator).

Judgments from individuals who have expertise on relevant subject matter can help better understand uncertainty to support value assessments and facilitate decision making. However, it is important to distinguish quantitative "elicitation" from qualitative inputs from experts, for the purpose of "validation," where the focus is expert consultation or expert discussion in a deliberative process. In contrast, structured expert elicitation (SEE) is the process of extracting expert knowledge about some uncertain quantity or quantities and formulating that information into probability distributions.[4]

Although the use of qualitative inputs from experts has a wide variety of applications within HCDM (see, eg, the many applications of the Research and Development/University of California Los Angeles appropriateness method),[5] the focus here is on quantitative expert elicitation through structured processes. Quantitative SEE has a much broader grounding, including in the Bayesian paradigm of statistical inference and in decision analysis. Within these contexts, which also underpin HCDM, elicited distributions can be incorporated relatively easily alongside any empirical data on the same quantity where available.[6,7] This is consistent with the ISPOR-Society for Medical Decision Making Modeling Good Practices Task Force recommendations on parameter uncertainty.[8]

Although SEE is a valuable tool to support decisions, there is some aversion to its use in formal HCDM processes. Elicitation more broadly is not ranked highly on the evidence hierarchy, but the structured processes used in SEE can increase its quality and validity. When using SEE, its contribution to the decision at hand should be made explicit so that, where appropriate, further data can be collected.[9] Perhaps most importantly, SEE should be conducted in a way that assures both transparency and accurate probabilistic judgments, that is, it should describe the inherently unobservable beliefs of the elicitor.

To improve the transparency, accuracy, and consistency of the judgments elicited, a formal and structured process for the elicitation should be used.[4,10] A method is considered to be SEE if a prespecified protocol is prepared and used for the task. Although there is often the need to develop a protocol specific to the exercise at hand, there are also a number of generic SEE protocols described in the literature.

These existing SEE protocols can support a range of applications in health outcomes and economic research. In this report, the task force will introduce the most used protocols for SEE, describe how they differ, and illustrate various uses of SEE through examples. There is no explicit intent to compare methods due to the differences in methodological and contextual approach. Given that this is an introductory task force report on a relatively emerging topic, it does not make specific recommendations about practice. However, there are broad recommendations regarding the contexts in which particular protocols may be better suited.

Motivation and Rationale for Guidance on SEE in HCDM

SEE has been used in a wide range of applications, with considerable heterogeneity in methods,[11] prompting questions about the quality and relevance of the elicited quantities. There is no doubt that this heterogeneity presents a challenge for decision makers. Guidance and standardization of approaches will help support decision making using SEE for several reasons.

First, as with other methods such as multicriteria decision analysis,[12] SEE is prone to multiple biases given that experts are invited to provide judgments for which they rely on heuristics and, therefore, potentially retrieve biased or overconfident estimates.[13] Two common biases in SEE are cognitive and motivational biases. Cognitive biases arise when experts are not able to correctly process the information available to them. This may result from limited cognitive capacity, time pressure, lack of cognitive effort put into the task, or from experts lacking the normative skills to make appropriate probabilistic inferences. Poorly specified questions can also contribute to biases. Examples include the availability bias and anchoring.[14] Availability bias occurs when an individual substitutes the difficult question, "what is the probability of this event", with the easier "how easily can I recall instances of or imagine this event," which can lead to overconfidence. The anchoring heuristic results in insufficient adjustment when an individual fixes ('anchor') on an initial value and fail to sufficiently adjust their estimates away from it to provide an accurate judgment.

Motivational biases result from the expert being invested in a specific outcome. Even in the absence of explicit conflicts of interest, motivational biases, such as confirmation (where focus is on information that is consistent with existing beliefs and preferences) and desirability (overestimation of the likelihood of positive outcomes), can still distort judgments. Further detail on de-biasing strategies can be found elsewhere.[15]

Second, reviews of applied SEE studies confirmed the use of a range of different approaches. In the context of health modeling studies, including health technology assessments (HTAs) and cost-effectiveness studies, Cadham et al[16] found 40 SEE studies and 112 nonformal (indeterminate) methods had been used to support decision making in the period up to June 2021. They assessed the reporting quality of included studies and concluded that reporting of SEE methods used in modeling studies was poor, making it difficult to make comparisons across studies.

Similarly, van Hest et al[17] reviewed single technology assessments appraised by the National Institute for Health and Care Excellence (NICE) program in the United Kingdom between October 2018 and April 2019. They screened 25 industry submissions, 23 of which included some form of expert elicitation (including SEE). From these, there were 173 expert elicitation-informed parameters identified, gathered through 5 different forms (advisory boards for clinical and health economics, ad hoc processes for both, and formal surveys). The authors emphasized the heterogeneity in terms of the level of reporting, as well as the lack of justification and/or exploration of alternative scenarios for the use and inclusion of expert opinion. Only 4 single technology appraisals included specific detail on the method (eg, providing questionnaires or transcripts).

SEE has also been used outside of HTA. Soares et al[18] examined where SEE was used to inform estimates of expected health opportunity costs in the United Kingdom's National Health Service. The analysts sought to ascertain expert beliefs on quantities that are central to an estimate of health opportunity costs for the United Kingdom's National Health Service, specifically, associations among mortality, morbidity, and expenditures in healthcare at the system level. Another example is by Dallow et al,[19] who used SEE, facilitated using Sheffield elicitation framework (SHELF) software, to generate previous distributions for the value of treatment effects of interest.

An Overview of Available SEE Protocols

To describe the most used protocols for SEE, a brief overview of the methodology follows.

Overview of SEE Methodology

As a structured research methodology, SEE involves decisions over a number of aspects. A list of relevant elements within SEE, relating to the definition of objective, methodology, conduct, and, finally, reporting and analyses, is presented in Table 1 (See Appendix 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2024.07.027 for detailed explanation of technical terms relevant for SEE).[20] The remainder of this section, for brevity, focuses on the methods and conduct of the SEE.

As with other forms of research, it is important that the SEE design and methodology are formalized. There are 2 important

components in SEE methodology: the encoding method and the SEE process when multiple experts are engaged.

The encoding method refers to the choice of numerical representation that experts will use to express their beliefs over each uncertain quantity. It is more complex for a probability distribution than it is for a single value. Two main categories of methods exist for encoding probabilities. Fixed interval methods partition the possible values of the quantity into discrete intervals and elicit probabilities for the value of interest belonging to each interval. Variable interval methods (VIMs) elicit the value range of the quantity that would correspond to particular quantiles of the distribution encoding their beliefs.[4]

When multiple experts are engaged in the SEE, the process needs to be defined, by considering the following elements:

- Level of elicitation: whether experts are asked to elicit probabilities individually or in a group
- Mode of aggregation: how the judgments from multiple experts are aggregated when a single distribution is required. This can be done through the experts' interaction (ie, behavioral aggregation) or analytically (ie, mathematical aggregation).
- Level of interaction: what level of interaction is allowed between experts during the elicitation. This may vary from no interaction, restricted/controlled interaction, or group interaction.

Overview of Selected SEE Protocols

Most of applied SEE examples in HCDM define their own elicitation protocol, making their own choices on the various elements of methodology, conduct, and reporting.[21] Nonetheless, to facilitate SEE implementation, a number of protocols have been developed that offer more prescriptive guidance on many aspects (although typically, not on all). This report will summarize the following protocols, chosen on the basis of previous application in HCDM. It is not intended to be an exhaustive list:

- SHELF
- modified Delphi method
- Cooke's classical method
- investigate, discuss, estimate, aggregate (IDEA) protocol
- Medical Research Council (MRC) reference protocol

These protocols describe all elements of the SEE process (Table 1).[20] In addition, there are a number of generic tools that have been developed to support conducting a SEE, such as MATCH,[22] EXPLICIT,[23] and others. Given that these do not inform all relevant elements of a SEE, they are not reviewed here. The guidance provided within each of these protocols is summarized below and presented in further detail in Table 2.[24-28]

The SHELF method[29,30] includes an initial step of individual elicitation, undertaken one to one with an experienced facilitator. This is followed by consensus aggregation based on group discussion. Group discussions typically involve all experts being physically colocated (face to face) with an experienced (through training and application) facilitator.

The facilitator explores differences in the individual responses through discussion (feedback) and attempts to resolve them and achieve consensus. Participants are asked to consider how a rational impartial observer would characterize the group's views if consensus were not naturally attained. The protocol either does not specify or allows for flexibility in all other aspects of the SEE design. A software package is provided that includes fitting and feedback (on both the individual and group assessments).

An example of the application of the SHELF protocol in HCDM is Ayers et al[31] In this study, clinical experts were asked to elicit longer-term survival of individuals with multiple myeloma treated with a chimeric antigen receptor T cell therapy. A web-based application was developed to undertake the SEE, in which the Kaplan–Meier curve from an existing shorter-term study was graphically presented alongside the elicited estimates for longer follow-up times.

As per SHELF, experts were first asked to elicit probabilities individually. A consensus meeting followed where experts were presented with the (anonymized) individual estimates from each expert and then were given the opportunity to discuss and collectively provide consensus estimates.

The Delphi method[32] uses repeated cycles of feedback and individual elicitation. At each iteration, experts are provided with summary information about how their responses compare with the group's responses and are given the opportunity to revise their previous answers in light of this information. No other interaction between the experts is allowed. After several iterations, it is expected that responses converge to a consensus assessment. However, there is no established definition of consensus.

For the elicitation of single values, consensus has been defined as anywhere from 51% to 80% agreement among respondents, dependent upon sample size, research aims, and resources. The Delphi protocol either does not specify or allows for flexibility in other aspects of the design and conduct of SEE. The classic Delphi method[33] has been adapted in many ways.[32,34-36]

Because this report is focused on quantitative SEE to characterize uncertainty, we focus on a modified Delphi developed specifically for this aim: the European Food Safety Authority (EFSA) modified Delphi, also known as a probabilistic Delphi.[34] The EFSA method does not define consensus and recommends mathematical aggregation after a fixed number of rounds, therefore potentially making the methods practical to use.

Carayann and Kitsos[37] applied the modified EFSA modified Delphi protocol to elicit probability distributions from 3 melanoma experts on the proportion of individuals with advanced BRAF-mutated melanoma that develop brain or transit metastases within a particular time frame. The authors used 2 elicitation rounds: a first where experts were asked to elicit probabilities individually and a second where they were shown the anonymized judgments and rationales collected in the first round and were then asked to provide revised judgments. After the 2 rounds, the experts' individual probability distributions were averaged to provide the final aggregate distribution.

Cooke's classical method[10,38,39] defines a protocol where experts are empirically tested on a set of seed or calibration questions to determine performance (measured through a combination of statistical accuracy and informativeness). The performance of their individually elicited judgments compared with the seed questions provides relative weights for each expert. Within this protocol, no interaction between experts is required, but sessions where experts are convened in a group may be used to discuss broader issues relating to the task. Colson et al[40] used the Cooke's classical method to elicit projections (with uncertainty) of resistance rates for pathogen-antibiotic pairs. Ten seed questions were used, which were the same resistance rates but in retrospective periods (for which data were available).

The IDEA protocol[9,41,42] combines elements of the modified Delphi and Cooke's method. Experts initially express their individual estimates, which are used as a basis for structured feedback and discussion (using a facilitator) before the second (and final) individual elicitation. The final individual estimates are combined using mathematical aggregation and can include weighted pooling based on seed variables (an element of Cooke's protocol).

Table 1. SEE elements.

SEE objective	Description
Selecting quantities	Decide what to elicit, including identification of the parameter(s) of interest (eg, observable or not), the quantity that will be elicited to inform each parameter (eg, frequency, proportion), the framing of the questions, and the handling of dependence between quantities.
Method of elicitation	
Encoding method	Choose how to capture judgments from experts over probability distributions. Can include fixed interval methods (FIMs), such as a roulette or chips and bins approach, and variable interval methods (VIMs), such as bisection or asking for specified quantiles.
Level of elicitation	Elicit judgments from either individuals or a group.
Aggregation	Combine assessments from multiple individual experts, either mathematically or behaviorally.*
Interaction between experts	Decide on level of interaction allowed (eg, no interaction, restricted or controlled interaction, or group interaction).
Conduct of elicitation	
Mode of administration	Conduct the elicitation either in-person (face to face) or remotely.
Feedback provided to experts	Provide various types of feedback to experts (eg, visualizations, fitted distributions, a written description of rationales), possibly with the opportunity for revision.
Rationales	Collect, record, and possibly share information about how experts reached their assessments.
Fitting distributions	Choose whether, and how, to fit elicited assessments with parametric or nonparametric distributions.
Validation	This should include designing the exercise to capture experts' beliefs in a way that results are consistent, coherent, and fit for purpose. It can also include validating the performance of experts or combinations of experts against seed questions (ie, questions that have realizations).
Managing biases	Consider cognitive biases and heuristics (eg, overconfidence, representativeness, availability, anchoring) and motivational biases (eg, "group think," wishful thinking) in the design and conduct of the elicitation and in the framing of questions.
Piloting the exercise	Test the elicitation protocol with a nonparticipant person with subject-matter expertise.
Evidence dossier preparation	Compile existing evidence on the subject of the elicitation.
Selecting experts	Identify multiple experts with the appropriate range of relevant knowledge, background, skills, and/or experience.
Training and preparing experts	Provide experts with background information on elicitation, how to provide subjective probability assessments, and the purpose of the elicitation.
Reporting and further analyses	
Adjusting judgments	Either leave expert assessments as provided or adjust them to improve calibration and/or coherence.
Getting feedback from experts on the process	Ask experts to reflect on the exercise after completing it or after future data are collected.
Documenting	Summarize the methods, process, conduct, and results of the elicitation.

SEE indicates structured expert elicitation.
*An elicitation can be conducted either at the individual level (which may be followed by mathematical aggregation) or at the group level (behavioral aggregation). An individual elicitation leads to the need of mathematical aggregation to generate a single distribution reflecting the group's views. This means that the beliefs of individual experts are combined using a mathematical rule, an example being linear opinion pooling.[20] Behavioral aggregation relies on interaction between the experts to create a single distribution that reflects either the experts' consensus beliefs or how an independent, rational observer would summarize the collective opinions of the experts.

For instance, the IDEA protocol was used by Lan et al[43] to elicit previous distributions for the likelihood of hospital admission for patients with bronchiolitis under 2 alternative treatment options. The distributions were then used to support sample size calculations in the design of a clinical trial. Remote real-time SEE workshops were conducted using a standardized script.

First, experts used a bespoke online SEE tool to provide their individual beliefs. The facilitator then used anonymized displays of the individual-level judgments to support group discussion. After this, experts were asked to re-elicit the probabilities individually. The final elicited distributions were pooled using equal weighting.

The MRC protocol[44,45] defines a reference case methodology for HTA, consisting of individual elicitation with mathematical pooling using equal weights, that is, without performance weighting. However, it allows for the use of nonreference methods when justified by the setting. For example, the weighting of experts or group consensus results (where the group consensus follows the individual elicitation used in the reference case) can be presented.

Harnan et al[46] used the MRC protocol in a pilot study for NICE on the feasibility of innovative payment models for antimicrobials. Expert beliefs were elicited on the outcomes of infection conditional on susceptibility to treatment for inclusion in a cost-effectiveness model. The SEE was conducted individually and remotely using a bespoke R Shiny application.[47] Expert beliefs were aggregated using linear opinion pooling.

Table 2. Features of selected SEE protocols.

Feature	SHELF	Modified Delphi	Cooke	IDEA	MRC
Encoding (VIM vs FIM)	Both possible	No guidance	VIM	No guidance	FIM (preferred)
Level of elicitation (group vs individual)	Group preceded by individual	Multiple rounds of individual	Individual	Two rounds of individual	Individual (preferred)
Aggregation (consensus vs mathematical)	Consensus	Multiple iterations until an acceptable level of consensus is achieved	Mathematical	Mathematical (2nd round of individual elicitation)	Mathematical
Interaction between experts during the elicitation	Yes, under facilitation	No interaction	No interaction	Interaction may, or may not, happen	No interaction
Mode of administration (remote vs face to face)	Face to face (preferred)	Remote	Face to face (preferred)	Both possible	Face to face (preferred)
Feedback provided to experts	Yes, under group discussion	Summaries of the previous round of elicitations provided	Yes	Feedback provided in between iterations	Yes
Rationales (captured or used?)	Rationales captured and used in group discussion	Rationales are captured and typically reported alongside quantitative results.	Rationales are captured and typically reported alongside quantitative results.	Rationales captured and used in group discussion or feedback	Rationales collected and recorded
Fitting distributions	No guidance, but distribution fitting is part of the software	No guidance	No guidance	No guidance	Probability distributions allowed
Validation	Not an explicit step	Not an explicit step	Answers to seed questions are used to define expert weights.	Assumed, with past assessments used to potentially screen experts	Stresses validation by internal and external review
Managing biases	Frame questions minimizing biases—ask for the upper and lower bound first.	Frame questions minimizing biases—ask for the upper and lower bound first.	Frame questions minimizing biases—ask for the upper and lower bound first. Performance weighting enables down-weighting of "overconfident" experts.	Frame questions minimizing biases—ask for the upper and lower bound first.	Frame questions minimizing biases—ask for the upper and lower bound first.
Specialized software available?	Yes, the SHELF R-package[24]	No	EXCALIBUR,[25] ANDURIL (MATLAB based), ANDURYL (Python based, code + GUI available)[26-28]	Off the shelf tools[9]	STEER (R shiny and Excel tools available)[24]

FIM indicates fixed interval method; IDEA, investigate, discuss, estimate, aggregate; MRC, Medical Research Council; SEE, structured expert elicitation; SHELF, Sheffield elicitation framework; VIM, variable interval method.

Applicability of SEE Protocols in Different Settings

Constraints in HCDM Relevant to SEE

In principle, SEE could be applied to almost any setting. However, in practice, it is not relevant for every HCDM situation. The purpose of SEE is to quantify and better understand uncertainty. SEE may not be necessary when parameters are already well understood from existing data and models, unless this additional information can improve the precision of any estimates, or if there is a need to understand the relevance of estimates to a particular application/patient group.

Even where a decision is consequential, the potential benefits of conducting SEE should be considered before embarking on the process, if, for instance, so little is known about an emerging topic that experts are unlikely to be able to narrow down a parameter's likely range. For example, estimating infection rates for a new virus, experts may not have enough observed experience to formulate an opinion that they can express quantitatively.

The risks of not conducting SEE should also be considered. For example, if a better understanding of the uncertainty around infection rates would likely change a decision, the conduct of SEE is indicated.

Where SEE is determined to be relevant and justified, determining a suitable approach depends on several dimensions, including economic, political, scientific, and social considerations. Economic factors consider how much time and resources the decision maker is willing to invest in the SEE process. There may

Table 3. Important positive and negative features of available protocols.

SEE method	Positive features	Negative features
SHELF	• A group facilitation may be beneficial where little is known about the quantities of interest—experts borrowing strength from other experts. • Flexibility in encoding methods (VIM or FIM) allows context-specific constraints to be reflected in choice of appropriate method. • Rationales captured to understand heterogeneity in responses • Explicit consideration of question framing to reduce bias • Specialized (code-based) software freely available	• Group facilitation may be challenging for some settings, due to time constraints and availability of experts. • Potential for biases from group interaction • Focus on consensus may underestimate uncertainty at the individual level, leading to biased pooled summaries. • Face-to-face completion may be challenging in some settings, eg, rare diseases or medical devices, where there are strict time constraints and low availability of experts. • No explicit validation process
Modified Delphi	• Rationales captured to understand heterogeneity in responses • Explicit consideration of question framing to reduce bias • Controlled interaction between experts reduces the potential for bias from group interaction.	• Multiple rounds for consensus may be impractical in some settings, eg, where SEE needs to be conducted under time constraints or where there is low availability of participants. • No specific software to operationalize methods • Lack of interaction between experts may limit applicability for more complex SEE tasks. • Lack of guidance on how to select experts • No guidance on fitting distributions to elicited beliefs and aggregating across experts • No explicit validation process
Cooke	• Explicit consideration of question framing to reduce bias • Rationales captured to understand heterogeneity in responses • Validation through differential weighting may minimize contribution of less informed experts. • Specialized software freely available	• Lack of interaction between experts may limit applicability for more complex SEE tasks. • Only advocates VIM. FIM may be more intuitive for less statistically trained experts. • Face-to-face completion may be challenging in some settings, eg, rare diseases or medical devices, where there are strict time constraints and low availability of experts. • No guidance on fitting distributions to elicited beliefs and aggregating across experts • Definition of seed questions for determining differential weights may be difficult. • Validation through differential weighting may maximize contribution of informed but biased experts.
IDEA	• Rationales captured to understand heterogeneity in responses • Explicit consideration of question framing to reduce bias • Specialized software freely available	• Lack of interaction between experts may limit applicability for more complex SEE tasks. • Rationales not captured from experts; challenge to understand motivation for beliefs and understand heterogeneity • No guidance on fitting distributions to elicited beliefs and aggregating across experts • No explicit validation process
MRC	• Applicability in HCDM is explicitly considered in developing this guidance. Constraints in HCDM integrated into choice of methods. • Flexibility in encoding methods (VIM or FIM) allows context-specific constraints to be reflected in choice of appropriate method. • Rationales captured to understand heterogeneity in responses • Advocates validation through internal and external review • Specialized (code-based) software freely available	• Consensus approach only considered applicable in certain situations, eg, rare diseases. • Face-to-face completion may be challenging in some settings, eg, rare diseases or medical devices, where there are strict time constraints and low availability of experts.

FIM indicates fixed interval method; HCDM, healthcare decision making; IDEA, investigate, discuss, estimate, aggregate; MRC, Medical Research Council; SEE, structured expert elicitation; SHELF, Sheffield elicitation framework; VIM, variable interval method.

be less value in a formal quantitative SEE to support a company's prioritization of candidate molecules for early-stage research compared with a national health organization evaluating a new candidate drug for inclusion in treatment guidelines.

In emergency situations, time constraints of a decision may have an impact on the choice of method, which may in turn dictate the necessary protocols. Due to the time and resources required to conduct a SEE, the exercise is best suited for situations

that are sufficiently consequential to justify that investment of resources.

Political factors entail the extent to which the decision maker has the responsibility and capacity to act on SEE results. For example, how much trust does the decision maker have in the SEE, and how much transparency and accountability do their decisions face? An internal company decision where the company is perhaps accountable only to its investors, shareholders, and directors is likely to face different political considerations than public regulatory or regional/national postmarket access decisions.

Scientific factors relevant for determining which SEE protocol is appropriate relate to how much evidence already exists for the topic at hand. If a rich set of evidence already exists, there is likely a large pool of experts who sufficiently agree about the relatively small uncertainty surrounding relevant parameters. If a decision relates to a new technology that has not been widely deployed, there may be fewer experts with more uncertainty and greater underlying disagreement on the key parameters.

Finally, there are social factors that dictate the available pool of experts. There may be a limited set of relevant and appropriate experts accepted, both by the decision maker and other important stakeholders. There is the question of whether those experts are willing to participate in a SEE, and if so, will they be motivated to respond honestly and accurately when providing their judgments?

Each of the protocol's strengths and limitations, in light of the types of constraints described earlier, is presented in Table 3. Specifically with reference to economic, political, scientific, social, and equity dimensions, there are several conclusions that can be drawn:

Economic

Consensus approaches, as advocated in SHELF and modified Delphi, may not be feasible when there are limited resources available for SEE. However, freely available software (SHELF, IDEA, MRC, Cooke's) may guide consensus building and thus reduce the resource burden of SEE.

Political

It is important that expert selection ensures that as many relevant perspectives as possible are captured. One way to address this is to take a systematic approach to the selection of experts.[48] SEE relies on subjectivity and may be prone to biases. All the protocols suggest that appropriate framing of the questions minimizes biases. Other aspects of the SEE may be subject to biases, and these biases may be specific to the setting. Only the MRC protocol has been established specifically for HCDM.

Scientific

All the protocols attempt to elicit uncertainty from experts. Consensus approaches, such as those used in modified Delphi and the SHELF method, may reduce between expert variation. Therefore, the resulting summary distributions may be artificially narrower. The SHELF method attempts to mitigate this by considering the perspective of a rational impartial observer during consensus discussions. However, this requires an experienced facilitator.

Social

Details of expert recruitment methodology and motivation are sparse in all the protocols. Very little is known about what makes a good expert, beyond minimizing any conflicts of interest. The MRC protocol[44,45] discusses the importance of ensuring that an expert is motivated to undertake the task, but specifics about how to do so are omitted. It is likely that this includes scientific interest.

BOX 1. Recommendations on protocols that may be best suited for specific HCDM contexts

1. **Contexts where the decision is time constrained or the SEE needs to be conducted responsively, eg, conducted live at a decision-making meeting**
   - Potentially MRC protocol
2. **Early assessment context or where there is limited evidence about the quantity of interest (rare diseases or indications)**
   - Potentially SHELF or MRC protocol (using consensus)
3. **Contexts focused on commissioning further research**
   - Consensus and recursive methods may be less appropriate as these may constrain between-expert uncertainty
4. **Contexts where a large number of experts are desirable for political reasons, ie, buy-in (public health)**
   - Potentially Modified Delphi or IDEA
5. **Contexts where experts have limited knowledge of SEE (limited normative skills) and limited opportunities for significant training**
   - Protocols using the variable interval method (VIM), like Cooke's, may be less appropriate
6. **Resource constrained contexts eg, local level**
   - Clearly defined protocols accompanied with validated software may be more appropriate, eg, MRC or SHELF
7. **Contexts that require a high cognitive burden on experts (complex quantities and/or high number of quantities)**
   - Protocols using simpler elicitation methods may be more appropriate, eg, MRC protocol

HCDM indicates healthcare decision making; IDEA, investigate, discuss, estimate, aggregate; MRC, Medical Research Council; SEE, structured expert elicitation; SHELF, Sheffield elicitation framework.

Equity

This issue is not discussed in any of the protocols. Consensus approaches (modified Delphi and SHELF) aim for agreement, whereas mathematical approaches (MRC, Cooke's, IDEA) allow experts to disagree individually, and the summaries of beliefs reflect this level of disagreement through wider distributions. The Cooke's method is "meritocratic" rather than equitable in that expert weights are derived based on calibration questions. In contrast, equal weighting is an intrinsically egalitarian approach to aggregation. This is assumed in all other approaches, although the MRC and IDEA protocols include flexibility for differential weighting.

Suitability of Different SEE Protocols for a Given Context

In the current section, we consider different HCDM contexts and suggest SEE protocols that may be applicable given the expected constraints faced in these contexts. The aim is not to be too prescriptive on the choice of protocol for each setting.

Ultimately, each type of decision maker should determine the suitability themselves. We encourage decision makers to specify which protocol they will implement with an emphasis on consistency in methods across their SEE applications. A predefined protocol, or reference case, for SEE, specific to a decision-making context rather than application, ensures consistency in the design, methodology, and conduct of elicitation exercises.

Further to the general considerations discussed in 4.1, Box 1 suggests protocols that could be more applicable or feasible in common scenarios (constraints) in different HCDM contexts. This list is not exhaustive, and it is likely that many others are apparent in settings where SEE is yet to be tested (eg, grant panels or access schemes).

Box 1 #1 illustrates contexts where decisions are time constrained (eg, where the timeline for decision making is short or where SEE will be conducted as part of a live decision-making process, such as a committee meeting). In these instances, an approach, such as the MRC protocol, could be more feasible than approaches based on iteration or requiring convening experts.

There are other contexts where there is limited opportunity for in-depth training of experts, leaving them less familiar with SEE and potentially lacking normative skills (Box 1 #5). In these instances, approaches that rely on VIM, such as Cooke's, should be avoided to ensure the internal validity of results from the exercise (Table 2,[24-28] section 3).

This is likely to be the case where there are time constraints or where experts are unlikely to be re-employed across decisions (Box 1 #6). For consensus-based methods, use an experienced facilitator. Where resources or expertise in conducting SEE are limited, approaches grounded on more intuitive quantities, which do not rely on interaction, can be more easily implemented (eg, MRC and IDEA).

The cognitive burden on experts engaging in an SEE will depend on the demands imposed by the approach used, the number of quantities elicited, and the complexity of those quantities and/or the complexity in formulating judgments about these. Where the needs of the SEE are complex (many quantities, complex quantities), the researcher may choose to use a protocol that allows simpler approaches to ease the burden on participants (Box 1 #7). In such a case, a protocol allowing the use of SEE methods that require less quantities to be elicited could be considered, for example, eliciting credible intervals (a VIM) within the MRC protocol.

Where there is limited information about the condition or the impact of technologies (Box 1 #2), such as technologies for rare diseases and new (first in class) technologies, or in early assessments in the product life cycle, deliberation among experts may be an important aspect of the overall assessment. This may include conditions where the need to generate information and understanding of common elements is crucial. In these instances, SEE protocols relying on consensus may be more relevant (eg, SHELF or the MRC using behavioral aggregation).

In situations where understanding and characterizing uncertainty is important (Box 1 #3), both to inform HCDM based on current evidence and to determine where additional information may be valuable to capture with a view to revisiting decisions, it may be desirable to fully capture differences in views and divergent opinions. In these circumstances, approaches aiming for consensus (iterative or live consensus approaches) may not be as appropriate as methods such as Cooke's and the MRC protocol.

In situations where there is a need for large/general buy-in and agreement (Box 1 #4), such as where decisions are political in nature or the technology has wide public implications(eg, public health policies), the need for broad clinical agreement and acceptance may be important to support the decision. In these instances, approaches with a focus on consensus and a need for participation by a large number of experts (>10) may be best suited, eg, modified Delphi or the SHELF method.

Conclusions

Evidence lies at the heart of explicit HCDM yet, in instances where evidence is insufficient or unsuitable, judgments are required that can materially affect the decision at hand. By using SEE, the judgments of experts can describe knowledge and provide assessments of uncertainty to be usefully integrated in formal HCDM processes. Elicited expert judgment is prone to bias, and hence, SEE and transparent reporting are called for.

The task force examined 5 existing SEE protocols: SHELF, the modified Delphi, Cooke's classical method, the IDEA protocol, and the MRC protocol. All 5 make alternative recommendations on how to define and conduct SEE. We described these protocols and considered their suitability across a range of different constraints that may be apparent in HCDM settings. Among the 5 protocols, certain elements show consistency (eg, in consideration for managing biases) whereas others show fundamental differences (eg, whether the approach to aggregation is via consensus or mathematical). Although there is little to no empirical evidence to support choices over these different methodological approaches, their applicability and suitability may differ across particular HCDM settings.

Within particular decision-making processes, the definition of a protocol for SEE is desired given that protocols provide structured, predefined approaches important for transparency and reproducibility. Most existing SEE studies published in HCDM do not use 1 of the 5 protocols but instead adapt these or define de novo protocols. This suggests that decision makers should primarily consider what methods and processes for SEE are well suited to the context in which they are to be used. This may require them to adapt existing protocols or develop new ones.

Although there are many possible uses of SEE, there are only a few decision makers (predominantly HTA agencies including Canadian Agency for Drugs and Technologies in Health, Zorginstituut Nederland, and NICE) that advocate their use. These agencies have started to navigate some of the practical complexities in using SEE. However, there are many settings in which processes need to be established before informed SEE protocol recommendations can be made. The recommendations made here regarding choice of SEE protocol should help to prepare decision makers in navigating the different steps they must take to allow SEE to be conducted efficiently, effectively, and consistently.

Recommendations for Further Research

This emerging good practice guidance identifies existing SEE protocols that can support HCDM and offers key considerations relevant for the choice of protocols in specified settings. To move toward more prescriptive guidance on which protocols meet the needs of particular settings requires further research, both in terms of applications of protocols across different HCDM processes and through empirical evidence generation necessary to test their performance. This is both in terms of robustness, the extent to which the elicited distributions reflect experts' beliefs, and suitability, to the extent that the protocol methods and results meet the needs of decision makers. Important areas for empirical testing include the use of consensus versus mathematical aggregation, particularly in settings such as early HTA, the selection, and recruitment of experts and methods to minimize bias.

In the absence of definitive guidance on SEE protocol choice, HCDMs are urged to consider the specificities of the setting in which they operate and the applicability of the various protocols given these characteristics. This task force report is intended to support decision makers in making the choice about which protocol to recommend in their particular setting. Consistency within specific decision-making processes is recommended to aid transparency and improve the robustness of the resulting decisions.

Transparency in the reporting of the methods and results of SEE is a requirement that is common across all protocols analyzed by this task force, but has poor implementation in practice.[11,16,17]

Where no specific protocol has been recommended, therefore, analysts should carefully identify and justify methodological and design choices including, where relevant, the choice of the protocol implemented. As experience with SEE in HCDM continues to grow, more specific practical guidance may be possible in the future.

Author Disclosures

Author disclosure forms can be accessed below in the Supplemental Material section.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2024.07.027.

Article and Author Information

Accepted for Publication: July 18, 2024

Published Online: October 14, 2024

doi: https://doi.org/10.1016/j.jval.2024.07.027

Author Affiliations: Centre for Health Economics, University of York, York, England, UK (Soares, Bojke); Department of Management Science, University of Strathclyde, Glasgow, Scotland, UK (Colson); Department of Medical Evaluation, HAS, Paris, France (Ghabri); Roche Diagnostics International AG, Rotkreuz, Switzerland (Garay); Eli Lilly and Company, Indianapolis, IN, USA (Felli); Canadian Agency for Drugs and Technologies in Health (CADTH), Ottawa, ON, Canada (Lee); ISPOR-Professional Society for Health Economics and Outcomes Research, Lawrenceville, NJ, USA (Molsen-David); Delft University of Technology, Delft, The Netherlands (Morales-Napoles); Department of Psychological Sciences, University of Missouri, Columbia, MO, USA (Shaffer); University of Melbourne, School of Population and Global Health, Parkville, Australia (IJzerman); Erasmus School of Health Policy & Management, Rotterdam, The Netherlands (IJzerman).

Correspondence: Maarten J. IJzerman, PhD, University of Melbourne, Cancer Health Services Research, Grattan Street, Parkville, Victoria 3010, Australia. Email: ijzerman@eshpm.eur.nl

Author Contributions: *Concept and design:* Soares, Colson, Bojke, IJzerman
Acquisition of data: Bojke, Garay, Morales-Napoles
Analysis and interpretation of data: Soares, Colson, Bojke, Ghabri, Felli, Lee, IJzerman
Drafting of the manuscript: Soares, Colson, Bojke, Molsen-David, IJzerman
Critical revision of paper for important intellectual content: Soares, Colson, Bojke, Ghabri, Garay, Felli, Lee, Molsen-David, Morales-Napoles, Shaffer, IJzerman
Administrative, technical, or logistic support: Molsen-David
Drs Ghabri and IJzerman are members of the Editorial Advisory Board for *Value in Health* and had no role in the peer-review process of this article.

Funding/Support: The authors received no financial support for this research.

Acknowledgment: We acknowledge the following reviewers for providing useful written comments that improved the manuscript: Nicole Mittman (CADTH), Farah Husein (CADTH), Lesley Dunfield (CADTH), Danielle Riley, Frederik Holmboe, Rachel Fleurence, Sabine Grimm, Niamh Carey, Michael S. Broder, Koen Degeling, Mario Ouwens, Juergen Polifka, Torbjørn Wisløff, Bart Willigers, Isabelle Newell, Ricardo Yajamin-Villamarin, Jonathan Karnon, Naohiro Yonemoto, Amarens Geuzinge (Zorginstituut Nederland), Belinda Mohr, Yoshita Paliwal, James Horscroft, and Carlos Díaz (CENETEC – Mexico HTA agency). The authors also thank those who provided oral feedback during task force presentations at ISPOR conferences.

REFERENCES

1. Jönsson B, Hampson G, Michaels J, Towse A, von der Schulenburg JMG, Wong O. Advanced therapy medicinal products and health technology assessment principles and practices for value-based and sustainable healthcare. *Eur J Health Econ HEPAC Health Econ Prev Care*. 2019;20(3):427–438.
2. Faulkner E, Holtorf AP, Walton S, et al. Being precise about precision medicine: what should value frameworks incorporate to address precision medicine? A report of the personalized precision medicine Special Interest Group. *Value Health*. 2020;23(5):529–539.
3. Wang SV, Schneeweiss S, Gagne JJ, et al. Using real-world data to extrapolate evidence from randomized controlled trials. *Clin Pharmacol Ther*. 2019;105(5):1156–1163.
4. O'Hagan A, Buck CE, Daneshkhah A, et al. *Uncertain Judgements: Eliciting Experts' Probabilities*. Chichester, England: John Wiley Sons; 2006.
5. Jandhyala R. Delphi non-rand modified Delphi, rand/UCLA appropriateness method and a novel group awareness and consensus methodology for consensus measurement: a systematic literature review. *Curr Med Res Opin*. 2020;36(11):1873–1887.
6. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd ed. New York, NY: Chapman & Hall/CRC; 2015.
7. Edwards W, Miles RF, Von Winterfeldt D. *Advances in Decision Analysis: From Foundations to Applications*. Cambridge, England: Cambridge University Press; 2012.
8. Briggs AH, Weinstein MC, Fenwick EAL, et al. Model parameter estimation and uncertainty: a report of the ISPOR-SMDM modeling good research practices task force-6. *Value Health*. 2012;15(6):835–842.
9. Hemming V, Burgman MA, Hanea AM, McBride MF, Wintle BC. A practical guide to structured expert elicitation using the IDEA protocol. *Methods Ecol Evol*. 2018;9(1):169–180.
10. Cooke RM. *Experts in uncertainty: opinion and subjective probability in science*. New York, NY: Oxford University Press; 1991.
11. Soares MO, Sharples L, Morton A, Claxton K, Bojke L. Experiences of structured elicitation for model-based cost-effectiveness analyses. 2018;21(6):715–723.
12. Montibeller G, von Winterfeldt D. *Biases and debiasing in multi-criteria decision analysis. Presented at: 2015 48th Hawaii International Conference on System Sciences, Kauai, HI, USA*. 2015:1218–1226.
13. Croskerry P, Norman G. Overconfidence in clinical decision making. *Am J Med*. 2008;121(5 Suppl):S24–S29.
14. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science*. 1974;185(4157):1124–1131.
15. Tversky A, Kahneman D. The framing of decisions and the psychology of choice. *Science*. 1981;211(4481):453–458.
16. Cadham CJ, Knoll M, Sánchez-Romero LM, et al. The use of expert elicitation among computational modeling studies in Health Research: a systematic review. *Med Decis Mak*. 2022;42(5):684–703.
17. van Hest N, Upton E, Ader J, Woodhouse F, O'Connor ME. PDG85 trust the experts? Acceptance of expert elicitation in the National Institute for Health and Care Excellence (NICE) single technology appraisal (STA) process. *Value Health*. 2019;22:S611.
18. Soares MO, Sculpher MJ, Claxton K. Health opportunity costs: assessing the implications of uncertainty using elicitation methods with experts. *Med Decis Mak*. 2020;40(4):448–459.
19. Dallow N, Best N, Montague TH. Better decision making in drug development through adoption of formal prior elicitation. *Pharm Stat*. 2018;17(4):301–316.
20. Clemen RT, Winkler RL. Combining probability distributions from experts in risk analysis. *Risk Anal*. 1999;19(2):187–203.
21. Center for Health Economics, University of York. Eliciting expert judgements in HTA. https://www.york.ac.uk/media/che/documents/Elicitation-NICE-final-report-York_01042020.pdf; Published 2020. Accessed April 26, 2023.
22. Morris DE, Oakley JE, Crowe JA. A web-based tool for eliciting probability distributions from experts. *Environ Modell Softw*. 2013;52:1–4.
23. Grigore B, Peters J, Hyde C, Stein K. EXPLICIT: a feasibility study of remote expert elicitation in health technology assessment. *BMC Med Inform Decis Mak*. 2017;17(1):131.
24. STEER. Centre for Health Economics, University of York. https://www.york.ac.uk/che/economic-evaluation/steer/. Accessed October 5, 2024.
25. Cooke RM, Solomatine D, Ababei D. Excalibur. http://www.lighttwist.net/wp/excalibur; Published 2013. Accessed October 9, 2024.
26. Rongen G, Hart CMP, Leontaris G, Morales-Nápoles O. Update (1.2) to ANDURIL and ANDURYL: performance improvements and a graphical user interface. *SoftwareX*. 2020;12:100497.
27. Hart CMP, Leontaris G, Morales-Nápoles O, Leontaris G, Morales-Nápoles O. Update (1.1) to ANDURIL — a MATLAB toolbox for ANalysis and Decisions with UnceRtaInty: Learning from expert judgments: ANDURYL. *SoftwareX*. 2019;10:100295.
28. Leontaris G, Morales-Nápoles O. ANDURIL. A MATLAB toolbox for ANalysis and Decisions with UnceRtaInty: learning from expert judgments. *SoftwareX*. 2018;7:313–317.
29. Gosling JP. SHELF: the Sheffield elicitation framework. In: Dias LC, Morton A, Quigley J, eds. *Elicitation: the Science and Art of Structuring Judgement*. Springer International Publishing; 2018:61–93.
30. Oakley JE, O'Hagan A. Sheffield elicitation framework. https://shelf.sites.sheffield.ac.uk/. Accessed June 11, 2024.

31. Ayers D, Cope S, Towle K, Mojebi A, Marshall T, Dhanda D. Structured expert elicitation to inform long-term survival extrapolations using alternative parametric distributions: a case study of CAR T therapy for relapsed/ refractory multiple myeloma. *BMC Med Res Methodol*. 2022;22(1):272.
32. Rowe G, Wright G, Bolger F. Delphi: a reevaluation of research and theory. *Technol Forecast Soc Change*. 1991;39(3):235–251.
33. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. *Int J Forecast*. 1999;15(4):353–375.
34. European Food Safety Authority. Guidance on expert knowledge elicitation in food and feed safety risk assessment. *EFSA J*. 2014;12(6):3734.
35. Dalkey N, Helmer O. An experimental application of the DELPHI method to the use of experts. *Manag Sci*. 1963;9(3):458–467.
36. Woudenberg F. An evaluation of Delphi. *Technol Forecast Soc Change*. 1991;40(2):131–150.
37. Carayanni V, Kitsos C. Model oriented statistical analysis for cancer problems. In: Pilz J, Oliveira TA, Moder K, Kitsos CP, eds. *Mindful Topics on Risk Analysis and Design of Experiments*. Springer International Publishing; 2022:37–53.
38. Cooke RM, Goossens LH. Procedures guide for structured expert judgment. European Commission. https://research.tudelft.nl/en/publications/procedures-guide-for-structured-expert-judgment; Published 2000. Accessed June 11, 2024.
39. Quigley J, Colson A, Aspinall W, Cooke RM. Elicitation in the classical model. In: Dias LC, Morton A, Quigley J, eds. *Elicitation: the Science and Art of Structuring Judgement*. Springer International Publishing; 2018:15–36.
40. Colson AR, Megiddo I, Alvarez-Uria G, et al. Quantifying uncertainty about future antimicrobial resistance: comparing structured expert judgment and statistical forecasting methods. *PLoS One*. 2019;14(7):e0219190.
41. Hanea AM, McBride MF, Burgman MA, et al. Investigate discuss estimate aggregate for structured expert judgement. *Int J Forecast*. 2017;33(1):267–279.
42. Hanea AM, McBride MF, Burgman MA, Wintle BC. Classical meets modern in the IDEA protocol for structured expert judgement. *J Risk Res*. 2018;21(4):417–433.
43. Lan J, Plint AC, Dalziel SR, et al. Remote, real-time expert elicitation to determine the prior probability distribution for Bayesian sample size determination in international randomised controlled trials: bronchiolitis in Infants Placebo versus epinephrine and dexamethasone (BIPED) study. *Trials*. 2022;23(1):279.
44. Bojke L, Soares M, Claxton K, et al. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. *Health Technol Assess*. 2021;25(37):1–124.
45. Bojke L, Soares MO, Claxton K, et al. Reference case methods for expert elicitation in health care decision making. *Med Decis Mak*. 2022;42(2):182–193.
46. Harnan S, Kearns B, Scope A, et al. *Final Report for the Technology Evaluation of Ceftazidime With Avibactam for Treating Severe Aerobic Gram-Negative Bacterial Infections*. University of York; 2022.
47. Chang W, Cheng J, Allaire J, et al. shiny: Web application framework for R. R package version 1.8.1.9001. https://github.com/rstudio/shiny; Published 2024. https://shiny.posit.co/. Accessed October 9, 2024.
48. Bolger F. The selection of experts for (probabilistic) expert knowledge elicitation. In: Dias LC, Morton A, Quigley J, eds. *Elicitation: the Science and Art of Structuring Judgement*. Springer International Publishing; 2018:393–443.