Reporting guidelines for the use of expert judgement in model-based economic evaluations

Cynthia P Iglesias1

Alexander Thompson2

Wolf H Rogowski3,4

Katherine Payne2\*

Authors Institutions

1 Department of Health Sciences, Centre for Health Economics and the Hull and York Medical School, University of York.

2 Manchester Centre for Health Economics, The University of Manchester

3 Helmholtz Zentrum München, German Research Center for Environmental Health, Institute of Health Economics and Health Care Management, Neuherberg, Germany

4 Department of Health Care Management, Institute of Public Health and Nursing Research, Health Sciences, University of Bremen, Germany

\*corresponding author: Manchester Centre for Health Economics, 4th floor, Jean McFarlane Building, The University of Manchester, Oxford Road, Manchester M13 9PL, UK

Katherine.payne@manchester.ac.uk

Keywords: expert judgement, expert elicitation, expert opinion, economic evaluation, model, guidelines

Word count: 4284

**Acknowledgements**

CPI contributed to the design and the conduct of the study, commented on drafts of the manuscript and produced the final version of the manuscript. AT contributed to the design and the conduct of the study, analysed the data, commented on drafts of the manuscript and approved the final version of the manuscript. WR contributed to the design and the conduct of the study, commented on drafts of the manuscript and approved the final version of the manuscript. KP conceived the idea for this study, contributed to the design and the conduct of the study, produced a first draft of the manuscript, commented on subsequent drafts and approved the final version of the manuscript. Also, we want to thank the following members of the Health Economics Elicitation Group for their contribution to this study: Laura Bojke, Qi Cao, Ali Daneshkhah, Christopher Evans, Bogdan Grigore, Anthony O'Hagan, Vincent Levy, Claire Rothery, Fabrizio Ruggeri, Ken Stein, Matt Stevenson, Patricia Vella Bonanno.

**Compliance with Ethical Standards**

(a) No funding was received for this study and (b) Cynthia P Iglesias, Alexander Thompson, Wolf H Rogowski, and Katherine Payne authors of this paper declare that there is no conflict of interest regarding its publication.

**Abstract 246 [250 words max]**

**Introduction:** Expert judgement has a role in model-based economic evaluations (EE) of healthcare interventions. This study aimed to produce reporting criteria for two types of study design to use expert judgement in model-based EE (1) an expert elicitation (quantitative) study and (2) a Delphi study to collate (qualitative) expert opinion.

**Method:** A two-round on-line Delphi process identified the degree of consensus for four core definitions (expert; expert parameter values; expert elicitation study; expert opinion) and two sets of reporting criteria in a purposive sample of experts. The initial set of reporting criteria comprised: 17 statements for reporting a study to elicit parameter values and/or distributions; 11 statements for reporting a Delphi survey to obtain expert opinion. Fifty experts were invited to become members of the Delphi panel by e-mail. Data analysis summarised the extent of agreement (using a pre-defined 75% ‘consensus’ threshold) on the definitions and suggested reporting criteria. Free text comments were analysed using thematic analysis.

**Results:** The final panel comprised 12 experts. Consensus was achieved for the definitions of: expert (88%); expert parameter values (83%); expert elicitation study (83%). Criteria were recommended to use when reporting an expert elicitation study (16 criteria) and a Delphi study to collate expert opinion (11 criteria).

**Conclusion:** This study has produced guidelines for reporting two types of study design to use expert judgement in model-based EE (1) an expert elicitation study requiring 16 reporting criteria and (2) a Delphi study to collate expert opinion requiring 11 reporting criteria.

**Introduction**

Model-based EE, in general, and model-based cost effectiveness analysis (CEA), specifically, now have a core role to inform reimbursement decisions and the production of clinical guidelines [1-5]. Early model-based CEA has emerged as a particularly useful approach [6] in the development phases of new diagnostic or treatment options. The practical application of model-based CEA can be severely hampered when: i) analysis is conducted early in the development phase of a new technology [7]; and ii) there is limited randomised controlled trial or observational data available to populate the model (e.g. devices and diagnostics) [8]. This paucity of available data to populate model-based CEA has stimulated the reliance on the use of expert judgement, a phenomenon that pervades beyond the health context and has attracted attention in ecology, environment and engineering [9, 10].

Expert judgement has many potential roles in the context of model-based CEA. Qualitative judgement expressions can: frame the scope of a model-based CEA; define care pathways; assist conceptualisation of model’s structure; investigate model’s face validity. Quantitative expressions of expert judgement can contribute to defining point estimates of key model parameters and characterise the uncertainty. The process of aggregating (collating or pooling) the views of a group of experts can be performed using: i) consensus methods (e.g. Delphi survey) to identify the extent of consensus in a qualitative sense; ii) mixed methods (e.g. nominal group) require getting experts together to exchange views to draw forth a single quantitative expression of their collective judgement using mathematical elicitation methods (e.g. roulette); and iii) mathematical aggregation to pool individual quantitative expressions of judgements using statistical methods (e.g. linear pooling).

Within the area of expert judgement there are many concepts and terms that are used interchangeably but often with no specific clear definitions offered leading to inconsistent use of terminology. Table 1 summarises dictionary-based definitions of key terms. From the definitions in Table 1 it is possible to develop a potential nomenclature system shown graphically in Figure 1. The proposed nomenclature suggests separating methods to elicit expert judgement in terms of whether the study design is underpinned by primarily a qualitative or quantitative paradigm. By definition “elicitation” implies getting information from somebody in a “particular way”. This suggests that the term “expert elicitation” may be better suited to describe methods aimed at drawing forth experts’ judgements expressed in a quantitative format (i.e. as probability density functions). In contrast, the term ‘expert opinion’ may be better suited to describe studies that aim to draw forth the opinions or beliefs of experts expressed in a qualitative format. This study uses the proposed nomenclature system to characterise studies that draw on expert judgement.

<Table 1 here>

<Figure 1 here>

Anecdotally, the ‘Delphi process’ is emerging as the most widely used consensus method used in the context of model-based EE. It is not possible to formally quantify the extent of the use of the Delphi in practice because of inconsistencies in terminology, description and application of the ‘method’. Generally, using a ‘Delphi process’ is described as a survey approach that involves at least two rounds to allow respondents, collectively grouped into an ‘expert panel’, to formulate and change their opinions on difficult topics [11]. Researchers commonly refer to the Delphi as a consensus method but are often not explicit that the method can only establish consensus if a set of clear decision rules are set when consensus has been reached [12]. In practice, the Delphi process is not a single method and has been adapted to answer different research questions. Sullivan & Payne (2011) [13] specified three types of Delphi defined by their stated purpose and research question to be answered[14]. A ‘classical’ Delphi could be used, for example, to inform a decision-analytic model structure. A ‘policy’ Delphi could be used to identify what value judgements are used by the decision-making body appraising health technologies. A ‘decision’ Delphi could provide a consensus view on the care pathways needed to inform the selection of model comparator(s). Although the potentially useful distinction between these three types of Delphi approach were suggested, this recommendation has not emerged into published examples in the context of model-based economic evaluations potentially as a result of the lack of clear reporting guidelines.

Importantly, Sullivan & Payne (2011) [13] were explicit that the Delphi should not be used as a method of behavioural aggregation to generate parameter values but more appropriate quantitative mathematical aggregation methods should be used. A substantial literature exists on the role and use of quantitative methods to elicit expert values (e.g. roulette, quartile, bisection, tertile, probability; hybrid) [15, 16]. These quantitative methods are common in that they are routed in mathematical and statistical Bayesian frameworks. In 2006, O’Hagan and colleagues produced a seminal textbook describing the rationale and application of quantitative methods to elicit expert’s probability values [15]. Within the collective set of methods to draw forth judgements for use in a model-based CEA –see Fig 1 - there is division amongst researchers on which methods are most suitable and lack of empirical research to support the use of one method over another [16].

While there is no agreement on the appropriate type of study to elicit expert judgement, there is a consensus view on a need for standardised reporting. Grigore et al (2013) reported a systematic review of 14 studies using the quantitative elicitation of probability distributions from experts undertaken to inform model‐based EE in healthcare [17]. The review identified variation in the type of elicitation approach used and a failure to report key aspects of the methods concluding with the need for better reporting. The potential strengths of using expert judgements in the context of eliciting quantitative values for model-based CEA can only be realised if studies are well-designed to minimise bias, conducted appropriately and reported with clarity and transparency. With the provision of a set of key criteria for reporting on quantitative estimates, papers can be quality assessed to assist with peer review and to aid those who may use the expert judgements in their own analysis. O’Hagan and colleagues’ textbook offers a potential starting point for reporting criteria for an expert elicitation study to generate quantitative values but needs adaptation to be practical and feasible in the context of writing up a study for publication in a peer reviewed manuscript [15].

Evans & Crawford (2000) commented on the use of the Delphi - as a consensus generating - method and suggested the need for: clear definition of techniques used; agreement on consensus reaching criteria; and conducting validation exercises [18]. Eleven reporting criteria were offered but these were not underpinned by agreement within the research community and have not been taken forward in practice [19]. More generally, Hasson and colleagues offered a set of reporting criteria for the Delphi but these were not specific to the context of using Delphi methods in a model-based CEA [12]. The aim of this study was to produce reporting criteria for two types of study design used when identifying expert judgements for use in model-based CEA: a ‘consensus’ Delphi study as the most frequently used method in “expert opinion” studies; and an “expert elicitation” study.

#### Method

This study followed published recommendations on how to develop reporting guidelines for health services research [20]. A rapid review of the literature failed to identify existing reporting criteria. In the absence of existing reporting criteria a two-round on-line Delphi process [21] was used to identify the degree of consensus in a purposive sample of experts. The objectives were to understand the degree of consensus on definitions of core terms relevant in the context of obtaining expert judgement for use in model-based EEs; and reporting criteria in this context for (1) an expert elicitation study and (2) a Delphi study to collate expert opinion. Ethical approval was required for this study because the Delphi involved asking respondents for their contact e-mail address. Ethical approval for the study was granted by The University of Manchester Research Ethics committee project reference REF: 15462.

### **The expert panel**

In this study an expert was defined as someone with previous experience of conducting a study to identify expert judgements in the context of EE or had written on this topic. The sampling frame was informed by a published systematic review of 14 expert elicitation studies to inform model‐based EEs in healthcare [17]. This sampling frame was supplemented with hand searching of relevant journals. A sampling frame of 50 potential members of an expert panel representing the views from different healthcare jurisdictions was generated. The study aimed to recruit a sample size of 15 experts representing the views of > 20% of the total available international experts. Contact E-mail addresses were obtained from the published study and updated using a Google search. The sample size was pre‑specified in a dated protocol (available from the corresponding author on request) and based on two criteria (1) the practicalities of using a Delphi method; and (2) the size of the potential pool of experts with knowledge of using expert judgements in the context of model-based economic evaluations. As sample size calculations for Delphi studies are not available it is necessary to rely on pragmatic approaches to define the relevant sample size. A published systematic review identified the potential pool of experts to be ~50, thus a sample size of 15 would represent a substantial proportion of the available pool.

Experts were invited to become members of the Delphi panel by e-mail with a link to the on‑line survey (hosted using SelectSurvey.NET) and an attached study information sheet. Respondents gave ‘assumed’ consent to take part by completing round-one of the survey and indicating if they were willing to be sent a second survey and named as a member of the Health Economics Expert Elicitation Group (HEEEG).

**The Delphi**

Round one of the Delphi survey (Appendix 1) comprised four sections: definitions of four concepts (expert; expert parameter values; expert elicitation study; a study designed to collect expert opinion); reporting criteria for an expert elicitation study (17 criteria); reporting criteria for a Delphi survey (11 criteria); background questions on the expert.

Concept definitions were created by a group of four health economists (authors of this paper) based on their knowledge of the expert judgement literature and the deliberative processes described in the introduction. Respondents were asked to rate their agreement with each definition as described using a five-point scale (see Table 2). A free text section at the end of the ‘definitions section’ and again at the end of the survey asked for comments on the definitions and general comments, respectively.

<Table 2 here>

In round-one, each of the suggested reporting criteria was presented as a statement. Each statement included in sections two and three of the first-round survey was informed by three published sources [15, 12, 21]. Respondents were asked to indicate *‘whether you think each of the stated criteria is required as a minimum standard for reporting the design and conduct’* of a study to identify expert values for use in a model-based EE (section two) or a (Delphi) method used to collate expert opinion (section three). The focus was to establish which reporting criteria the experts thought would be essential for including in a standalone expert opinion or expert elicitation study or in a study reporting an expert opinion or expert elicitation exercise as an element of a wider EE study. A question at the end of sections two and three, respectively, asked respondents to indicate which, if any, criteria could be removed if the study identifying expert judgement was reported as part of the model-based EE paper. Respondents rated each criteria using a five-point scale (see Table 2).

Respondents, who indicated that they were willing to take part, were sent a second-round survey with a summary of their responses to round-one with a summary of the expert panel’s responses. The second-round survey (Appendix 2) comprised three sections including: re‑worked definitions of four concepts (expert; expert parameter values; expert elicitation study; expert opinion); reporting criteria for expert elicitation studies for which no consensus was reached in round‑one; reporting criteria for a Delphi survey to obtain expert opinion for which no consensus was reached in round‑one. Respondents were asked if they had any comments on criteria for which consensus had been reached in round‑one and general comments.

#### Data analysis

Data analysis aimed to summarise the extent of agreement about the appropriateness of the core definitions and requirement to use the suggested reporting criteria. Only round‑two results were used in the final analysis. A pre‑defined criterion was set to define the concept of consensus with at least 75% of panel members agreeing (rating of 4 or 5; see Table 2) on each definition or reporting criteria in terms of ‘required’ (rating of 4 or 5; see Table 2) or ‘not required’ (see Table 2). Free text comments collated in each round of the survey were also analysed using thematic analysis (see Appendix 3).

**Results**

The first and second rounds of the Delphi survey were conducted in November 2015 and January 2016, respectively. In round-one, 17 (34% response rate) of the invited 50 respondents completed the survey. Of these 17 respondents, 13 (26% response rate) agreed to be a member of the expert panel for round-two of the Delphi survey. Appendix 4 shows the level of consensus reached on each definition or statement and the response distribution from each round of the Delphi survey.

***The expert panel***

The final expert panel, completing all questions in both rounds, comprised 12 experts (24% response rate). These experts named their primary role as: Health Economist (n=2); Decision Analyst (n=3); Operations Researcher (n=1); Outcomes Researcher (n=1); Statistician (n=2) Clinician (n=1); HTA researcher (n=1); Decision Maker (n=1). Of these 12 experts, nine (75%) had been working in their primary role for more than 10 years and five (42%) had published more than three studies using methods to identify expert judgement.

***Key definitions***

Box 1 shows the ‘agreed’ definitions for four concepts (expert; expert parameter values; expert elicitation study; expert opinion). In round-one, consensus was only achieved for the definition of expert (n=15; 88%). The analysis of free text comments (see Appendix 3) was used to modify definitions for an expert elicitation study and for expert opinion. In round-two, consensus was achieved on the definitions for expert parameter values (n=10; 83%) and expert elicitation study (n=10; 83%). The expert panel was evenly divided on the appropriateness (n=6; 50%) of the definition offered for ‘expert opinion. The free text comments highlighted the reasons for the lack of consensus that seem to focus on the attempt to use ‘opinion’ to reflect a study underpinned by a ‘qualitative paradigm’ and ‘elicitation’ to reflect a ‘quantitative paradigm’ with respondents suggesting that the use of the word ‘opinion’ should be specific to the context of the study.

<Box 1 here>

***Reporting criteria for an expert elicitation study***

Table 3 shows the agreed 16 reporting criteria recommended for an expert elicitation study seeking to attach a value and/or a distribution to a parameter for use in a model-based EE. The expert panel felt that all 16 criteria should be reported in a standalone paper reporting the expert elicitation study and when an expert elicitation study is included within a paper reporting the EE it informs. In the latter case, this may require appropriate use of supplementary appendices.

<Table 3 here>

In the first round of the Delphi, the expert panel reached the pre-defined threshold for consensus (75% of experts) on the relevance of 15 of the original 17 potential reporting criteria. The two criteria that failed to reach consensus were: data recording (n=7; 41% of experts agreed needed) and ethical issues (n=9: 53% of experts agreed needed). Following the second round, consensus was reached for the inclusion of ethical issues (n=9; 75%) but not for the approach taken for data recording (n=8; 67% ). The expert panel viewed the reporting criteria dealing with data aggregation and the interpretation of results to be most important as 100% of consensus was achieved in round one of the Delphi. Appendix 3 summarises the free text comments made on the reporting criteria.

***Reporting criteria for a Delphi study***

Table 4 shows the final 11 criteria recommended for reporting a study designed to identify the degree of consensus of opinion in an expert panel. All 11 criteria were judged to be appropriate for reporting either a standalone Delphi study and / or one reported within the EE the Delphi study had informed.

<Table 4 here>

After round-one, the expert panel agreed on seven of the 11 criteria (see Appendix 4). No consensus was reached on the need for reporting: the research rationale (n=12; 71% of experts agreed); the literature review (n=15; 73%); the survey (n=11; 67%) and ethical issues (n=9: 53%). After round-two, consensus was reached for 10 criteria. The expert panel did not reach the pre-defined threshold for consensus (75% of experts) on including ethical issues (n=9; 67%). However, the need to report ethical issues is a general requirement for most journals and for this reason was included in the final list of required criteria. Appendix 3 summarises the free text comments offered on these reporting criteria.

**Discussion**

A two-round Delphi survey was used to generate criteria for two types of study designs commonly used to incorporate expert judgement in model-based EEs of healthcare interventions. A literature is emerging, both within and outside a healthcare focus, on the use of quantitative expert judgement. O’Hagan and colleagues substantive textbook takes the reader from the rationale and theoretical underpinning of the use of expert judgement through the range of available elicitation methods and remaining areas where future research is required [15].

Commentators have started to assimilate the range of methods available to identify expert judgement [22] and written their experiences in designing elicitation studies [23]. In addition, there are emerging guidelines on steps to follow when designing an expert elicitation study[24]. Knol et al [25] suggested a seven step procedure [25, 24]. Implementing software is now also readily available (e.g. SHELF, MATCH) [26-28]. Lack of agreement on which methods should be used to elicit parameter values and the uncertainty (i.e. strength of belief) associated with them is a remaining challenge. Probabilistic characterisation of a parameter value is a key process to be able to incorporate expert judgements into model‑based EEs. The most relevant method to elicit a quantitative expression of expert judgement is likely to depend on the type of parameter in question. Selecting the most appropriate method for eliciting a parameter and its distribution is an open area for future empirical research.

Compared with quantitative methods to elicit expert judgements, developing methods to conduct Delphi surveys in the context of model-based EEs has attracted less attention in the literature commonly read by health economists, decision analysts or operations researchers. The community of academic nurses and pharmacists have both embraced this method, writing the majority of published material on the use of Delphi as a survey method in health services research [12, 29]. Developing a standardised toolkit for the design and conduct of a Delphi survey is problematic as Delphi is not really a single method [11]. It is also challenging as it is not clear whether it is a qualitative or quantitative technique. If used to collate opinions (i.e. qualitative expressions of expert judgement) then it can be viewed to be underpinned by a qualitative paradigm. However, if used as a means of behavioural aggregation, as has been done in some applications in the context of generating parameter values for model-based EE, then it could be viewed to be underpinned by a quantitative paradigm [30]. The view taken in this study was in line with the recommendations of Sullivan & Payne (2010) suggesting that a Delphi study is best suited to identify qualitative expressions of expert opinion [13]. It should also be recognised that a Delphi survey is only one of a number of available approaches to use as a consensus method and was chosen as the focus for this study because other approaches, such as the nominal group technique, have attracted less attention and use in the practice of model-based EEs.

There are a number of existing reporting criteria and guidelines published for use when reporting EEs, such as the CHEERS [31], assessing the methodological quality of EEs [32], conducting decision-analytic modelling [33] and stated preference studies [34]. These existing guidelines, and the ones suggested in this study, all work on the premise that standardisation of reporting is desirable. The belief that it is good to standardise reporting can be viewed to be unequivocal. Caution is however necessary. The standardisation of reporting should not serve to stagnate or halt future empirical research to improve the methods or applied use of expert opinion or expert elicitation studies.

This study had a number of limitations. It clearly assumed that Delphi studies have an appropriate role to draw forth qualitative expressions (“expert opinion”) of expert judgement on how to report an “expert opinion” study. The tautological nature of the assumption is acknowledged by the authors. However, in the absence of other methods using a Delphi survey was a practical solution to identify the extent of consensus on how to report a Delphi. Another potential limitation is that it relied on collating the opinion of experts that had published studies using quantitative approaches to elicit probability distributions for model-based economic evaluations. It was not feasible to use the opinions of known experts who had experience in the use of Delphi methods because they could not be identified due to the lack of consistent keywords used in published manuscripts reporting model-based economic evaluations using the Delphi method and lack of reporting guidelines of such studies.

This study achieved a final sample size of 12 experts. Sample size calculations for a Delphi survey are not as established, or as well defined, as that for a randomised controlled trial. Delphi processes in health care have used sample sizes between 4 and 3,000 panel members [29]. Some guidance, has suggested that the optimum size of a panel is between 7 and 12 members, with seven members being the minimum panel size but some researchers have recommended panels of 300 to 500 people [11]. The optimal size of a panel is a subject for empirical research but it has been suggested that the size of a panel should be governed by the purpose of the investigation [29]. Using authorship of peer-reviewed publications as a proxy for expertise, this study identified a potential sampling frame of 50 respondents. Related studies using Delphi surveys that arguably had a larger potential pool of respondents of all health economists achieved a panel membership of 23 experts to develop criteria for the assessment of the methodological quality of EEs [32]. We must acknowledge that the final sample of 12 experts can only represent the views of a select few and their opinions may not be generalisable to a wider audience. The test for the acceptance of these reporting criteria is whether the guidelines are taken up and used in future studies.

The study was not able to achieve a consensus view on the definition of ‘expert opinion’. The aim was to try and distinguish between a study designed to elicit a quantitative expression of expert judgement on parameter values and the uncertainty surrounding them (expert elicitation) compared with a study designed to draw forth qualitative expressions of expert judgement (expert opinion). This distinction was not achieved in the definition we used. However, we maintain that it is important to make this distinction as often, to paraphrase one of the respondents in this study, when using expert judgement in model based EEs we may not be using the study to *collate expert opinion (e.g. to inform the model structure) but might want to generate a value for a parameter and its distribution (to inform a model input parameter).*

The expert panel were also divided on the focus on the Delphi as a consensus method suggesting that other approaches could have been used. Ideally, this study would have generated reporting criteria for all types of consensus methods suitable for use in the context of a model-based EE but this was not practical or feasible. There was further disagreement within the panel around whether it is appropriate to use the Delphi as a behavioural aggregation method to combine expert estimates of a parameter value. We advocate the use of the Delphi in line with previous recommendations [13] as a method useful when collating opinions (qualitative expressions of expert judgement) on, for example care pathways or model structures. However, we recognise that this view is not underpinned by empirical research and future studies should compare and contrast whether mathematical or behavioural aggregation methods are robust and practical when used in model-based EEs of healthcare interventions.

**Conclusion**

This study has produced guidelines for reporting two types of study design to use expert judgements in model-based EEs (1) an expert elicitation study requiring 16 reporting criteria and (2) a Delphi study to collate expert opinion requiring 11 reporting criteria. These two sets of criteria are suggested as guidelines to be used by journals wanting to assess the quality of reporting such studies and researchers designing such studies or critiquing their quality. Further research is required to develop and apply the methods to elicit parameter values and/or their distributions and collate expert opinions. Specifically, it is necessary to conduct empirical research on whether mathematical or behavioural aggregation methods are robust and practical when used in model-based EE of healthcare interventions.

**References**

1. Dakin H, Devlin N, Feng Y, Rice N, O'Neill P, Parkin D. The Influence of Cost‐Effectiveness and Other Factors on Nice Decisions. Health economics. 2015;24(10):1256-71.

2. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. Health economics. 2006;15(12):1295-310.

3. Excellence NIfC. Guide to the methods of technology appraisal. London: NICE, 2004.

4. Sculpher MJ, Claxton K, Drummond M, McCabe C. Whither trial‐based economic evaluation for health care decision making? Health economics. 2006;15(7):677-87.

5. Marsden G, Wonderling D. Cost-effectiveness analysis: role and implications. Phlebology. 2013;28(suppl 1):135-40.

6. Annemans L, Genesté B, Jolain B. Early modelling for assessing health and economic outcomes of drug therapy. Value in Health. 2000;3(6):427-34.

7. IJzerman MJ, Steuten LM. Early assessment of medical technologies to inform product development and market access. Applied health economics and health policy. 2011;9(5):331-47.

8. Iglesias CP. Does assessing the value for money of therapeutic medical devices require a flexible approach? Expert review of pharmacoeconomics & outcomes research. 2015;15(1):21-32.

9. Morgan MG. Use (and abuse) of expert elicitation in support of decision making for public policy. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(20):7176-84. doi:<http://dx.doi.org/10.1073/pnas.1319946111>.

10. Kuhnert PM, Martin TG, Griffiths SP. A guide to eliciting and using expert knowledge in Bayesian ecological models. Ecology letters. 2010;13(7):900-14. doi:<http://dx.doi.org/10.1111/j.1461-0248.2010.01477.x>.

11. Mullen PM. Delphi: myths and reality. Journal of health organization and management. 2003;17(1):37-52.

12. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. Journal of advanced nursing. 2000;32(4):1008-15.

13. Sullivan W, Payne K. The appropriate elicitation of expert opinion in economic models. Pharmacoeconomics. 2011;29(6):455.

14. Rauch W. The decision delphi. Technological Forecasting and Social Change. 1979;15(3):159-69.

15. O'Hagan A. Uncertain judgements eliciting experts' probabilities. Chichester u.a.: Wiley; 2006.

16. Gosling JP. Methods for eliciting expert opinion to inform health technology assessment. In: Vignette Commissioned by the MRC Methodology Advisory Group. Medical Research Council (MRC) and National Institure for Health Research (NIHR). 2014. https://[www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjr7Krwu\_LMAhWGLsAKHbylCNUQFgghMAA&url=https%3A%2F%2Fwww.mrc.ac.uk%2Fdocuments%2Fpdf%2Fmethods-for-eliciting-expert-opinion-gosling-2014%2F&usg=AFQjCNGhWG6GGK0oAPr0IYtYCBttaC3Jnw&sig2=aRuqYgYzvqJ1ibAH8EDoHA&bvm=bv.122676328,d.ZGg](http://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjr7Krwu_LMAhWGLsAKHbylCNUQFgghMAA&url=https%3A%2F%2Fwww.mrc.ac.uk%2Fdocuments%2Fpdf%2Fmethods-for-eliciting-expert-opinion-gosling-2014%2F&usg=AFQjCNGhWG6GGK0oAPr0IYtYCBttaC3Jnw&sig2=aRuqYgYzvqJ1ibAH8EDoHA&bvm=bv.122676328,d.ZGg).

17. Grigore B, Peters J, Hyde C, Stein K. Methods to elicit probability distributions from experts: a systematic review of reported practice in health technology assessment. Pharmacoeconomics. 2013;31(11):991-1003.

18. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. Journal of clinical nursing. 2003;12(1):77-84.

19. Evans C, Crawford B. Expert judgement in pharmacoeconomic studies. Pharmacoeconomics. 2000;17(6):545-53.

20. Moher D, Schulz KF, Simera I, Altman DG. Guidance for developers of health research reporting guidelines. PLoS Med. 2010;7(2):e1000217.

21. Jones J, Hunter D. Consensus methods for medical and health services research. BMJ: British Medical Journal. 1995;311(7001):376.

22. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. Journal of clinical epidemiology. 2010;63(4):355-69.

23. Kadane J, Wolfson LJ. Experiences in elicitation. Journal of the Royal Statistical Society: Series D (The Statistician). 1998;47(1):3-19.

24. Kinnersley N, Day S. Structured approach to the elicitation of expert beliefs for a Bayesian‐designed clinical trial: a case study. Pharmaceutical statistics. 2013;12(2):104-13.

25. Knol AB, Slottje P, van der Sluijs JP, Lebret E. The use of expert elicitation in environmental health impact assessment: a seven step procedure. Environmental Health. 2010;9(1):19.

26. O' Hagan A. SHELF: the Sheffield Elicitation Framework. 2010. <http://www.tonyohagan.co.uk/shelf/>. Accessed 26/02/2016.

27. Morris DE, Oakley JE, Crowe JA. A web-based tool for eliciting probability distributions from experts. Environmental Modelling & Software. 2014;52:1-4.

28. Morris E, Oakley J. MATCH. 2014. <http://optics.eee.nottingham.ac.uk/match/uncertainty.php/>. Accessed 26/02/2016.

29. Cantrill J, Sibbald B, Buetow S. The Delphi and nominal group techniques in health services research. International Journal of Pharmacy Practice. 1996;4(2):67-74.

30. Melton Lr, Thamer M, Ray N, Chan J, Chesnut Cr, Einhorn T et al. Fractures attributable to osteoporosis: report from the National Osteoporosis Foundation. Journal of Bone and Mineral Research. 1997;12(1):16-23.

31. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D et al. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Bmj. 2013;346:f1049.

32. Evers S, Goossens M, De Vet H, Van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. International journal of technology assessment in health care. 2005;21(02):240-5.

33. Ramos MCP, Barton P, Jowett S, Sutton AJ. A Systematic Review of Research Guidelines in Decision-Analytic Modeling. Value in Health. 2015;18(4):512-29.

34. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics. 2008;26(8):661-77.

35. Oxford Dictionaries. Oxford University Press. 2016. <http://www.oxforddictionaries.com/definition/english>. Accessed 26/02/2016.

36. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. Journal of the American Statistical Association. 2005;100(470):680-701.

37. The Free Dictionary. Farlex. 2016. <http://www.thefreedictionary.com/UK>. Accessed 26/02/2016.

38. Macmillan Dictionary. Macmillan Publishers. 2016. <http://www.macmillandictionary.com/>. 26/02/2016.