
Article:
Iglesias Urrutia, Cynthia Paola orcid.org/0000-0002-3426-0930, Thompson, ALexander, Rogowski, Wolf H et al. (1 more author) (2016) Reporting guidelines for the use of expert judgement in model-based economic evaluations. Pharmacoeconomics. 1161–1172. ISSN 1179-2027

https://doi.org/10.1007/s40273-016-0425-9

Appendix 1

Round one of the Delphi survey

Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 1

Introduction

Introduction

Thank you for considering participation in this survey, which is the first part of a two round Delphi survey.

This survey should take no more than 30 minutes to complete.

You are being asked to contribute your expertise and thoughts on the minimum reporting criteria that should be used when reporting the results of studies designed to elicit expert judgement (collate opinion or parameter value elicitation) as inputs into model-based economic evaluations.

By continuing with this survey, you are consenting to participate in the research.

The process of collating these reporting criteria is being led jointly by Katherine Payne (The University of Manchester) and Cynthia Iglesias (University of York) in collaboration with Alex Thompson, The University of Manchester and Wolf Rogowski, Institute of Health Economics and Health Care Management, Munich.

Take care when you are completing this survey to regularly use the save button and DO NOT use the back button on your browser (as your answers will be lost).

If you require any further information, please contact:

Katherine Payne
Professor of Health Economics
Manchester Centre for Health Economics
The University of Manchester
Oxford Road, Manchester M13 9PL
E-mail: katherine.payne@manchester.ac.uk

or

Cynthia Iglesias
Senior Research Fellow in Health Economics
Department of Health Sciences
University of York
Heslington, York, YO10 5DD
E-mail: Cynthia.iglesias@york.ac.uk

Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 2

Your views on key definitions

Your views on key definitions

In this section, we will ask you to indicate your level of (dis) agreement with four key definitions related to expert judgement.

We are interested in the methods that allow us to learn about expert judgement. Broadly, there are two types of methods:
(1) to quantify parameter input values (mainly quantitative);
(2) to collate expert opinions (mainly qualitative).

You will be asked to rate each of the definitions using a five-point scale in which:

1 = strongly disagree: this means that you think that the definition as written does NOT define the term and requires extensive modification.

2 = disagree: this means that you think that the definition as written does NOT define the term but only requires minor modification.

3 = neither agree nor disagree: means that you do not have a strong opinion on the definition as written.

4 = agree: this means that you think that the definition as written DOES define the term but requires some minor modification.

5 = strongly agree: this means that you think that the definition as written DOES define the term and requires no modification.

There is also an option to indicate if you 'do not know' the answer.

Now please read each of the following 4 definitions and indicate your rating using the scale provided.

---

1. An **expert** is: 'someone who has in-depth knowledge of the topic of interest gained through their life experience, education or training'

   Please rate how much you agree or disagree with this definition by choosing the relevant number on the 5-point scale.*

   - ◯ 1. Strongly disagree
   - ◯ 2. Disagree
   - ◯ 3. Neither agree nor disagree
   - ◯ 4. Agree
   - ◯ 5. Strongly agree
   - ◯ Do not know

   Please use this space if you want to explain the reason for your answer

2. A study designed to generate '**expert parameter values**' uses: 'a quantitative elicitation method to: derive point estimates and distributions for model input parameters; and/or pool expert judgement.'

   Please rate how much you agree or disagree with this definition by choosing the relevant number on the 5-point scale.*

   - ◯ 1. Strongly disagree
   - ◯ 2. Disagree
   - ◯ 3. Neither agree nor disagree
   - ◯ 4. Agree

- 5. Strongly agree
- Do not know

Please use this space if you want to explain the reason for your answer

3. The purpose of an **expert elicitation study** is: 'to construct a probability distribution for a parameter value that appropriately represents the knowledge/judgement of the expert and the degree of uncertainty in that knowledge/judgement'.

Please rate how much you agree or disagree with this definition by choosing the relevant number on the 5-point scale.

- 1. Strongly disagree
- 2. Disagree
- 3. Neither agree nor disagree
- 4. Agree
- 5. Strongly agree
- Do not know

Please use this space if you want to explain the reason for your answer

4. A study designed to collate '**expert opinion**' uses: 'a qualitative Delphi panel, or similar consensus method, to collate views from experts to: frame the scope of the model-based economic evaluation; inform model conceptualisation; identify model face validity; quantify point estimates without specifying a distribution; pool elicitation results'.

Please rate how much you agree or disagree with this definition by choosing the relevant number on the 5-point scale.*

- 1. Strongly disagree
- 2. Disagree
- 3. Neither agree nor disagree
- 4. Agree
- 5. Strongly agree
- Do not know

Please use this space if you want to explain the reason for your answer

Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Reporting criteria for a study to elicit expert values

Reporting criteria for a study to elicit expert values

In this section, we will ask you 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 elicit expert values for use in a model-based economic evaluation.

These criteria focus on a study designed to elicit experts' quantitative values about one or more uncertain quantities (parameter input values) in probabilistic form.

The criteria have been informed by the textbook by O'Hagan and colleague: Uncertain judgements: Eliciting experts' probabilities. John Wiley & Sons, Chichester: 2006

You will be asked to rate each of the criteria using a five-point scale in which:

1 = definitely not required: this means that you think that the criteria should NOT be included in the reporting criteria.

2 = possibly not required: this means that you think that the criteria could probably be omitted without any loss of key detail.

3 = no strong opinion: means that you do not have a strong opinion to indicate whether the criteria is, or is not, required.

4 = possibly required: this means that the criteria could be included but it is not vital (it would be 'nice to have')

5 = definitely required: this means you think that the criteria should be included in the reporting criteria otherwise key detail will not be reported.

There is also an option to indicate if you 'do not know' the answer.

Now please read each of the following 17 criteria and indicate your rating using the scale provided.

Please consider the criteria listed below for inclusion in a standalone paper that describes the design and conduct of a study used to elicit expert parameter values.

---

5. Research rationale: The need for using an expert elicitation exercise should be described.

   Note: This should ideally include some reference to the design and conduct of systematic reviews to identify key input parameters for the decision analytic model and a statement confirming that these reviews did not identify data relevant for the model-based economic analysis as specified.*

   - ◯ 1. Definitely not required
   - ◯ 2. Possibly not required
   - ◯ 3. No strong opinion
   - ◯ 4. Possibly required
   - ◯ 5. Definitely required
   - ◯ Do not know

6. Research problem: All uncertain quantities (model input parameters) that will be elicited should be described.

   Note: In some instances, there may be a substantial number of uncertain quantities required, and a degree of 'pre-selection' will have occurred to identify a relevant sub-set. Clear justification for model parameters identified as key for the decision problem needs to be provided.*

   - ◯ 1. Definitely not required
   - ◯ 2. Possibly not required
   - ◯ 3. No strong opinion

- 4. Possibly required
- 5. Definitely required
- Do not know

7. Measurement type of uncertain quantities: The rationale for the measure type of each uncertain quantity elicited should be described.

   Note: The measurement type of uncertain quantities can be (but not limited to): scalar quantities (i.e. numbers); proportions (e.g. probabilities); ratios (e.g. odds, hazard); risk (e.g. relative); rate (e.g. mortality), etc. Some measures are easier to understand and elicit than others thus it is important to fully justify the selection of any measurement type.*

   - 1. Definitely not required
   - 2. Possibly not required
   - 3. No strong opinion
   - 4. Possibly required
   - 5. Definitely required
   - Do not know

8. Definition of an expert: The nature of the expert population should be described to clearly state what topic of expertise they represent and why.

   Note: It is unlikely that a single expert will be sufficient and it is generally necessary to elicit judgement from a group of experts that were selected to represent the views of a larger population.*

   - 1. Definitely not required
   - 2. Possibly not required
   - 3. No strong opinion
   - 4. Possibly required
   - 5. Definitely required
   - Do not know

9. Number of experts: The selection criteria and final number of experts recruited to provide expert judgement should be reported.

   Note: Selection criteria need to be described in detail. There should be clear and specific pre-defined criteria used to identify how experts were selected and if/how their elicited quantities were used.*

   - 1. Definitely not required
   - 2. Possibly not required
   - 3. No strong opinion
   - 4. Possibly required
   - 5. Definitely required
   - Do not know

10. Preparation: There should be a clear reference made to a protocol that describes the design and conduct of the elicitation exercise.*

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

11. Piloting: It should be clearly reported if the elicitation exercise process was piloted and a summary of any modifications made.

    Note: The selection and number of experts used in the piloting process should be reported. Key aspects that may have required modification include: selection of experts; measure type and number of uncertain quantities to be elicited; training exercise; framing of the elicitation question; method of aggregation.*

    - 1. Definitely not required

- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

12. Data collection: The approach to collect the data should be reported.

    Note: Data can be collected from individual experts or a group/s of experts. Collecting data from individual experts means that a mathematical aggregation process may need to be used. Collecting data from a group/s of experts means that behavioural aggregation methods may be used.*

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

13. Administration: The mode of administering the elicitation exercise should be reported.

    Note: Elicitation exercises can be conducted face-to-face or via the telephone and/or computer. In a limited number of situations it may be feasible to collect the data using a self-administered online or postal survey but this is unlikely to be successful in most instances. Both face-to-face and telephone data collection is likely to be supported by using a computer.*

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

14. Training: The use of training materials should be reported and made available.

    Note: This may include background training materials sent to the experts and/or training in the use of probabilities and nature of distributions. This document need to provide explanation of efforts made to prevent influencing experts' knowledge and judgement. In practice, this recommendation will require a copy of the elicitation exercise to be included, which is likely to be presented as electronic supplementary material.*

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

15. The exercise: The number and framing of questions used in the exercise should be reported and made available.

    Note: This will require a copy of the elicitation exercise to be included, which is likely to be presented as electronic supplementary material.*

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

16. Data recording: The format used to record the expert judgment should be reported.

Note: This may use pen and paper but it is more likely to involve the use of an Excel spreadsheet or a bespoke software package.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

17. Data aggregation: The type of aggregation method (mathematical or behavioural) should be reported together with a description of the method or process used to aggregate the data.

Note: Mathematical aggregation (relevant when data were collected from multiple individual experts) can be conducted using a range of methods, for example: Bayesian methods; opinion pooling; Cooke's method.

Behavioural aggregation (relevant when data were collected from group/s of experts) can be conducted using processes such as, for example: Delphi or Nominal Group technique.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

18. Measures of performance for data aggregation: The processes followed to estimate measures of performance (calibration/information) for data aggregation need to be fully described

Note: Calibration is the process of measuring the performance of experts by comparing their judgement with a 'seed parameter' (parameter whose true values are known or can be found within the duration of a study). Calibration scores represent the probability that any differences between expert's probabilities and observed values of 'seed parameters' might have arisen by chance. Information represents the degree to which an expert's distribution is concentrated, relative to some user-selected background measure.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

19. Ethical issues: The ethical issues for the expert sample and research community should be described.

Note: The use of expert elicitation should acknowledge the issues of ethical responsibility, anonymity, reliability, and validity in an ongoing manner throughout the data collection and aggregation process.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

20. Presentation of results: The individual, and aggregated, point estimate(s) and distribution for each uncertain quantity (quantities) should be presented.

Note: The units of measurement should be clear and attention should be paid to the style of presentation that may benefit from the use of figures rather than relying on a tabular format.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion

- 4. Possibly required
- 5. Definitely required
- Do not know

21. Interpretation of results: The interpretation of uncertain quantities elicited should be presented together with a description of how the results will be used in the model-based economic analysis.

    Note: This should include an explanation of how the reader should interpret the results. It should be recognised that the number and type of experts used will affect the results obtained. The interpretation of results should comment on the degree of uncertainty observed.*

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

22. If the study used to elicit expert values were to be reported in the same paper as that reporting the main model based economic evaluation which of the above criteria (1 to 17) would you exclude?*
    Select at least 1 and no more than 15.

    - [ ] 1. Research rationale
    - [ ] 2. Research problem
    - [ ] 3. Measurement type of uncertain quantities
    - [ ] 4. Definition of an expert
    - [ ] 5. Number of experts
    - [ ] 6. Preparation
    - [ ] 7. Piloting
    - [ ] 8. Data collection
    - [ ] 9. Administration
    - [ ] 10. Training
    - [ ] 11. The exercise
    - [ ] 12. Data recording
    - [ ] 13. Data aggregation
    - [ ] 14. Measures of performance for data aggregation
    - [ ] 15. Ethical issues
    - [ ] 16. Presentation of results
    - [ ] 17. Interpretation of results
    - [ ] 18. none

23. Please use this space to indicate if you think we have missed any key criteria and write down a description of any missing criteria.

Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 4

Reporting criteria for a study to collate expert opinion

Reporting criteria for a study to collate expert opinion

In this section, we will ask you to indicate whether you think the stated criteria is required as a minimum standard for reporting the design and conduct of a method used to collate expert opinion.

These criteria focus on how to report a study designed to collate and generate consensus on experts' views to

inform a model-based economic evaluation.

The criteria have been informed by two key publications:

Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000; 32(4):1008-15
Jones J & Hunter D. Consensus methods for medical and health services research. BMJ 1995; 311: 376-80

Now please read each of the following 11 criteria and indicate your rating using the scale provided.

Please consider the criteria listed below for inclusion in a standalone paper that describes a study used to collate expert opinion.

---

24. Research problem: The research problem should be clearly defined and ideally framed explicitly as a research question to be addressed.

Note: When clarifying the research problem, remember the Delphi technique is a group facilitation technique and as such only lends itself to group involvement.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

25. Research rationale: The topic and use of the Delphi method should be justified.

Note: The Delphi is best used when the research requires anonymity to avoid dominance of one opinion. It should also be remembered that the strength of the Delphi method lies in the use of iteration in which the process of gaining opinion occurs in rounds to allow individuals to change their opinion.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

26. Literature review: The rationale for using the Delphi method must be informed by a clear description of the evidence base for the topic of the study.

Note: The focus of using the Delphi method should be where unanimity of opinion does not exist owing to a poor evidence base. This section should also describe the process of determining the most important issues to refer to in the design of the initial round of the Delphi.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

27. Data collection: This should include a clear explanation of the Delphi method employed.

Note: This should be sufficiently detailed for a reader to be able to duplicate the process of conducting the Delphi method. This includes a description of the types of questions used (qualitative or quantitative and ranking, rating or scoring scale used). This section should describe which medium was used to collect the data (electronic or written communication). This section should also describe how results from previous rounds were fed back to the experts and whether feedback is given to the group and/or individual response.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

28. The survey: A copy of each round of the survey used in the Delphi method should be presented.

Note: The use of journal supplementary appendices should be exploited to allow the reader access to a full copy of the survey used for each round of the Delphi.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

29. Rounds: This should state the number of rounds planned and used together with the plans for moving from one round to the next.

Note: The structure of the initial round (either qualitative or quantitative) should be decided from the protocol stage of the study together with the number of rounds to be used.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

30. The sample: The sample or 'expert' panel should be described in terms of the definition of an expert in the context of the study and the selection and composition of the panel including how it was formed from a sampling frame and response rate achieved.

Note: It should be noted that the composition of the panel will affect the results obtained from the Delphi method. Careful thought should be given to the criteria employed to define an expert, the justification of a participant as an 'expert' and the use of non-probability sampling techniques (such as purposive or criterion methods).*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

31. Ethical issues: The ethical issues for the expert sample and research community should be described.

Note: The use of the Delphi method should acknowledge the issues of ethical responsibility, anonymity, reliability, and validity in an ongoing manner throughout the data collection and analysis process.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

32. Data analysis: The management of opinions, analysis and handling of both qualitative and quantitative data should be described.

Note: As with any other survey-based approach, a pre-specified data analysis plan should be prepared. This should include a clear description of the meaning of 'consensus' in relation to the stated aim of the study and how 'agreement' is defined. This should also take account of reliability and validity issues identified.

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

33. Presentation of results: The results for each round, and final round, should be presented clearly while taking account of the audience of the study findings.

Note: The response rate for each round should be stated. Careful consideration should be paid on how to present the interim (between round) and final results in either graphical and/or statistical representations. In round 1, a summary of the total number of issues generated should be presented. In the final round, the strength of overall consensus should be summarised. Reporting data from quantitative questions should acknowledge the limitations associated with eliciting point estimates (e.g. no indication of uncertainty).*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

34. Interpretation of results: The interpretation of consensus (not) gained should be presented together with the meaning of the results and direction of further research needed.

Note: This should include an explanation of how the reader should interpret the results, and how to digest the findings in relation to the emphasis being placed upon them. It should be recognised that the composition of the panel will affect the results obtained. The interpretation of results should state whether 'outliers' to the overall consensus were asked for the reasons for their answers.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

35. If the study used to collate expert opinions were to be reported in the same paper as that reporting the main model based economic evaluation which of the above criteria (1 to 11) would you exclude?*
Select at least 1 and no more than 10.

- 1. Research problem
- 2. Research rationale
- 3. Literature review
- 4. Data collection
- 5. The survey
- 6. Rounds
- 7. The sample
- 8. Ethical issues
- 9. Data analysis
- 10. Presentation of results
- 11. Interpretation of results
- 12. none

36. Please use this space to indicate if you think we have missed any key criteria and write down a description of any missing criteria.

Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 5

Additional comments

Additional comments

37. Please use this space to write any additional comments you have either about the use of reporting criteria or the design of this survey.

Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 6

Your background

Your background

38. How would you describe your primary role?*
- Health economist
- Operations researcher
- Decision analyst
- Other, please specify

39. How many years of experience do you have working in your primary role?*
- Less than 5
- 5 to 10 years
- More than 10 years

40. How many studies in healthcare have you published either as first author or co-author in which you used methods to identify expert judgement?*
- none
- 1
- 2
- 3

- 4
- 5 to 9
- 10 or more

41. Thinking about your most recent relevant study; were you the lead member of the research team who designed the study using methods to identify expert judgement?*
    - Yes
    - No
    - Do not know

42. Thinking about your most recent relevant study; do you think the published paper accurately reflected the design and analysis of the study used to elicit expert judgements?*
    - Yes
    - No
    - Do not know

Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 7

Recent paper analysis

Recent paper analysis

43. Why do you think the published paper did not accurately reflect the design and analysis of the study used to elicit expert judgements?

    Please explain your answer here.*

Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 8

Follow up

Follow up

44. Are you willing to be sent a survey to complete for round two of this Delphi?
    - Yes
    - No

    If yes, please give your email address here.

45. Are you willing to be named as a contributing author to the reporting criteria and named as part of the Health Economics Expert Elicitation Group?

- ◯ Yes
- ◯ No

If yes, please give your email address here.

Appendix 2

Round two of the Delphi survey

2nd Round: Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 1

Introduction

Introduction

Thank you for participating in this survey, which is the second part of a two round Delphi survey.

Respondents to round one of this survey gave some useful qualitative comments which have been taken into account. Consequently we have modified two of the definitions offered in our 'Key Definitions' section.

In addition, we want to address a common theme which was concern about the predominance of the Delphi method in our survey. The need for reporting criteria when using a Delphi survey is not driven by a belief that this method SHOULD be used to combine expert's values (behavioural aggregation). However, the Delphi method IS used in practice. Importantly a leading journal has specifically requested reporting criteria when a Delphi survey has been used in the context of an economic evaluation. A Delphi survey does perhaps have a role when collating 'expert opinion' to inform model conceptualisation and structure. We therefore feel it is useful to collate views on how a Delphi should be reported with sufficient transparency such that someone using the results can understand exactly what was done and why.

Please use the "Feedback" document (which will be found in the email inviting you to this survey) to help you complete this second round of the survey. The feedback document contains your individual responses from the first round as well as the aggregate scores from the other experts.

We will ask you only to update your selection for the questions where no consensus in the first round was achieved but for each question you may provide comments if you so wish.

We anticipate that this survey should take no more than 15 minutes to complete.

If you require any further information, please contact:

Katherine Payne
Professor of Health Economics
Manchester Centre for Health Economics
The University of Manchester
Oxford Road, Manchester M13 9PL
E-mail: katherine.payne@manchester.ac.uk

or

Cynthia Iglesias
Senior Research Fellow in Health Economics
Department of Health Sciences
University of York
Heslington, York, YO10 5DD
E-mail: Cynthia.iglesias@york.ac.uk

The process of collating these reporting criteria is being led jointly by Katherine Payne (The University of

Manchester) and Cynthia Iglesias (University of York) in collaboration with Alex Thompson, The University of Manchester and Wolf Rogowski, Institute of Health Economics and Health Care Management, Munich.

2nd Round: Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 2

Your views on key definitions

Your views on key definitions

1. **1.** An expert is: 'someone who has in-depth knowledge of the topic of interest gained through their life experience, education or training'.

   First round consensus achieved for the inclusion of this criteria.

   If you have further comments you may use the space below.

---

2. **2.** A study designed to generate '**expert parameter values**' uses: 'a quantitative elicitation method to: derive point estimates and distributions for model input parameters; and/or pool expert judgement.'

   Please rate how much you agree or disagree with this definition by choosing the relevant number on the 5-point scale.*

   - 1. Strongly disagree
   - 2. Disagree
   - 3. Neither agree nor disagree
   - 4. Agree
   - 5. Strongly agree
   - Do not know

   Please use this space if you want to explain the reason for your answer

---

In the next two questions we will ask you about two definitions in which we are trying to differentiate between studies that aim to generate a quantitative estimate for a parameter value (expert elicitation study) and studies that aim to collate expert views in a qualitative sense (expert opinion).

Given the qualitative comments provided in the previous round, we have slightly altered these two definitions. However, you may still use the previous round results to guide your selection.

---

3. **3.** The purpose of an **expert elicitation study** is: 'to quantify a parameter value that appropriately represents the knowledge/judgement of the expert and the degree of uncertainty in that knowledge/judgement'.

Please rate how much you agree or disagree with this definition by choosing the relevant number on the 5-point scale.

- 1. Strongly disagree
- 2. Disagree
- 3. Neither agree nor disagree
- 4. Agree
- 5. Strongly agree
- Do not know

Please use this space if you want to explain the reason for your answer

4. **4.** A study designed to collate '**expert opinion**' uses: 'a qualitative consensus method (e.g. a Delphi method or other approach) to collate views from experts to: frame the scope of the model-based economic evaluation; inform model conceptualisation; identify model face validity; quantify point estimates without specifying a distribution; pool elicitation results'.

Please rate how much you agree or disagree with this definition by choosing the relevant number on the 5-point scale.*

- 1. Strongly disagree
- 2. Disagree
- 3. Neither agree nor disagree
- 4. Agree
- 5. Strongly agree
- Do not know

Please use this space if you want to explain the reason for your answer

2nd Round: Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 3

Reporting criteria for a study to elicit expert values

Reporting criteria for a study to elicit expert values: standalone paper

In this section, we will ask you 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 elicit expert values for use in a model-based economic evaluation.

These criteria focus on a study designed to elicit experts' quantitative values about one or more uncertain quantities (parameter input values) in probabilistic form.

Please consider the criteria listed below for inclusion in a standalone paper that describes the design and conduct of a study used to elicit expert parameter values.

---

5. **5.** Research rationale: The need for using an expert elicitation exercise should be described.

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

6. **6.** Research problem: All uncertain quantities (model input parameters) that will be elicited should be described.

   First round consensus achieved for the inclusion of this criteria.

   If you have further comments you may use the space below.

7. **7.** Measurement type of uncertain quantities: The rationale for the measure type of each uncertain quantity elicited should be described.

   First round consensus achieved for the inclusion of this criteria.

   If you have further comments you may use the space below.

8. **8.** Definition of an expert: The nature of the expert population should be described to clearly state what topic of expertise they represent and why.

   First round consensus achieved for the inclusion of this criteria.

   If you have further comments you may use the space below.

9. **9.** Number of experts: The selection criteria and final number of experts recruited to provide expert judgement should be reported.

   First round consensus achieved for the inclusion of this criteria.

   If you have further comments you may use the space below.

10. **10.** Preparation: There should be a clear reference made to a protocol that describes the design and conduct of the elicitation exercise.

    First round consensus achieved for the inclusion of this criteria.

    If you have further comments you may use the space below.

11. **11.** Piloting: It should be clearly reported if the elicitation exercise process was piloted and a summary of any modifications made.

    First round consensus achieved for the inclusion of this criteria.

    If you have further comments you may use the space below.

12. **12.** Data collection: The approach to collect the data should be reported.

    First round consensus achieved for the inclusion of this criteria.

    If you have further comments you may use the space below.

13. **13.** Administration: The mode of administering the elicitation exercise should be reported.

    First round consensus achieved for the inclusion of this criteria.

    If you have further comments you may use the space below.

14. **14.** Training: The use of training materials should be reported and made available.

    First round consensus achieved for the inclusion of this criteria.

    If you have further comments you may use the space below.

15. **15.** The exercise: The number and framing of questions used in the exercise should be reported and made available.

    First round consensus achieved for the inclusion of this criteria.

    If you have further comments you may use the space below.

16. **16.** Data recording: The format used to record the expert judgment should be reported.

    Note: This may use pen and paper but it is more likely to involve the use of an Excel spreadsheet or a bespoke software package.

    First round consensus not achieved.

    Please review your previous answer and those of the other experts. *

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

17. **17.** Data aggregation: The type of aggregation method (mathematical or behavioural) should be reported together with a description of the method or process used to aggregate the data.

    First round consensus achieved for the inclusion of this criteria.

    If you have further comments you may use the space below.

18. **18.** Measures of performance for data aggregation: The processes followed to estimate measures of performance (calibration/information) for data aggregation need to be fully described

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

19. **19.** Ethical issues: The ethical issues for the expert sample and research community should be described.

Note: The use of expert elicitation should acknowledge the issues of ethical responsibility, anonymity, reliability, and validity in an ongoing manner throughout the data collection and aggregation process.

First round consensus not achieved.

Please review your answer and those of the other experts. *

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

20. **20.** Presentation of results: The individual, and aggregated, point estimate(s) and distribution for each uncertain quantity (quantities) should be presented.

First round consensus achieved.

Please enter comments here if you have any.

21. **21.** Interpretation of results: The interpretation of uncertain quantities elicited should be presented together with a description of how the results will be used in the model-based economic analysis.

First round consensus achieved.

Please enter comments here if you have any.

2nd Round: Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 4

Reporting criteria for a study to elicit expert values

Reporting criteria for a study to elicit expert values which is not a standalone paper

The following questions are based on Q22 in the previous survey.

The previous responses identified these criteria as potential options to EXCLUDE from reporting if the study used to expert elicit values (quantitative estimates for a parameter value) were to be reported in the same paper as that reporting the main model based economic evaluation.

Please consider the criteria listed below if the study used to elicit expert values were to be reported in the same paper as that reporting the main model based economic evaluation.

You do not need to refer to the 'Feedback' sheet for the questions on this page.

---

22. **22a.** Data collection: The approach to collect the data should be reported.

Note: Data can be collected from individual experts or a group/s of experts. Collecting data from individual experts means that a mathematical aggregation process may need to be used. Collecting data from a group/s of experts means that behavioural aggregation methods may be used.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

23. **22b.** Measurement type of uncertain quantities: The rationale for the measure type of each uncertain quantity elicited should be described.

Note: The measurement type of uncertain quantities can be (but not limited to): scalar quantities (i.e. numbers); proportions (e.g. probabilities); ratios (e.g. odds, hazard); risk (e.g. relative); rate (e.g. mortality), etc. Some measures are easier to understand and elicit than others thus it is important to fully justify the selection of any measurement type.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

24. **22c.** Preparation: There should be a clear reference made to a protocol that describes the design and conduct of the elicitation exercise.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

25. **22d.** Piloting: It should be clearly reported if the elicitation exercise process was piloted and a summary of any modifications made.

Note: The selection and number of experts used in the piloting process should be reported. Key aspects that may have required modification include: selection of experts; measure type and number of uncertain quantities to be elicited; training exercise; framing of the elicitation question; method of aggregation.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

26. **22e.** Data recording: The format used to record the expert judgment should be reported.

Note: This may use pen and paper but it is more likely to involve the use of an Excel spreadsheet or a bespoke software package.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

27. **22f.** Ethical issues: The ethical issues for the expert sample and research community should be described.

Note: The use of expert elicitation should acknowledge the issues of ethical responsibility, anonymity, reliability, and validity in an ongoing manner throughout the data collection and aggregation process.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

2nd Round: Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 5

Reporting criteria for a study to collate expert opinion

Reporting criteria for a study to collate expert opinion: standalone paper

In this section, we will ask you to indicate whether you think the stated criteria is required as a minimum standard for reporting the design and conduct of a method used to collate expert opinion.

Please consider the criteria listed below for inclusion in a standalone paper that describes a study used to collate expert opinion.

---

28. **24.** Research problem: The research problem should be clearly defined and ideally framed explicitly as a research question to be addressed.

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

29. **25.** Research rationale: The topic and use of the Delphi method should be justified.

Note: The Delphi is best used when the research requires anonymity to avoid dominance of one opinion. It should also be remembered that the strength of the Delphi method lies in the use of iteration in which the

process of gaining opinion occurs in rounds to allow individuals to change their opinion.

First round consensus not achieved.

Please review your answer and those of the other experts. *

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

30. **26.** Literature review: The rationale for using the Delphi method must be informed by a clear description of the evidence base for the topic of the study.

Note: The focus of using the Delphi method should be where unanimity of opinion does not exist owing to a poor evidence base. This section should also describe the process of determining the most important issues to refer to in the design of the initial round of the Delphi.

First round consensus not achieved.

Please review your answer and those of the other experts. *

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

31. **27.** Data collection: This should include a clear explanation of the Delphi method employed.

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

32. **28.** The survey: A copy of each round of the survey used in the Delphi method should be presented.

Note: The use of journal supplementary appendices should be exploited to allow the reader access to a full copy of the survey used for each round of the Delphi.

First round consensus not achieved.

Please review your answer and those of the other experts. *

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

33. **29.** Rounds: This should state the number of rounds planned and used together with the plans for moving from one round to the next.

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

34. **30.** The sample: The sample or 'expert' panel should be described in terms of the definition of an expert in the context of the study and the selection and composition of the panel including how it was formed from a sampling frame and response rate achieved.

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

35. **31.** Ethical issues: The ethical issues for the expert sample and research community should be described.

Note: The use of the Delphi method should acknowledge the issues of ethical responsibility, anonymity, reliability, and validity in an ongoing manner throughout the data collection and analysis process.

First round consensus not achieved.

Please review your answer and those of the other experts.
*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

36. **32.** Data analysis: The management of opinions, analysis and handling of both qualitative and quantitative data should be described.

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

37. **33.** Presentation of results: The results for each round, and final round, should be presented clearly while taking account of the audience of the study findings.

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

38. **34.** Interpretation of results: The interpretation of consensus (not) gained should be presented together with the meaning of the results and direction of further research needed.

First round consensus achieved for the inclusion of this criteria.

If you have further comments you may use the space below.

2nd Round: Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Reporting criteria for a study to collate expert opinion

Reporting criteria for a study to collate expert opinion which is not a standalone paper

The following questions are based on Q35 in the previous survey.

The previous responses identified these criteria as potential options to EXCLUDE from reporting if the study used to elicit expert opinion (expert views in a qualitative sense) were to be reported in the same paper as that reporting the main model based economic evaluation.

Please consider the criteria listed below if the study used to elicit expert values were to be reported in the same paper as that reporting the main model based economic evaluation.

You do not need to refer to the 'Feedback' sheet for the questions on this page.

---

39. **35a.** Research rationale: The topic and use of the Delphi method should be justified.

Note: The Delphi is best used when the research requires anonymity to avoid dominance of one opinion. It should also be remembered that the strength of the Delphi method lies in the use of iteration in which the process of gaining opinion occurs in rounds to allow individuals to change their opinion.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

40. **35b.** Literature review: The rationale for using the Delphi method must be informed by a clear description of the evidence base for the topic of the study.

Note: The focus of using the Delphi method should be where unanimity of opinion does not exist owing to a poor evidence base. This section should also describe the process of determining the most important issues to refer to in the design of the initial round of the Delphi.*

- 1. Definitely not required
- 2. Possibly not required
- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

41. **35c.** Data collection: This should include a clear explanation of the Delphi method employed.

Note: This should be sufficiently detailed for a reader to be able to duplicate the process of conducting the Delphi method. This includes a description of the types of questions used (qualitative or quantitative and ranking, rating or scoring scale used). This section should describe which medium was used to collect the data (electronic or written communication). This section should also describe how results from previous rounds were fed back to the experts and whether feedback is given to the group and/or individual response.*

- 1. Definitely not required
- 2. Possibly not required

- 3. No strong opinion
- 4. Possibly required
- 5. Definitely required
- Do not know

42. **35d.** Rounds: This should state the number of rounds planned and used together with the plans for moving from one round to the next.

    Note: The structure of the initial round (either qualitative or quantitative) should be decided from the protocol stage of the study together with the number of rounds to be used*

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

43. **35e.** Ethical issues: The ethical issues for the expert sample and research community should be described.

    Note: The use of the Delphi method should acknowledge the issues of ethical responsibility, anonymity, reliability, and validity in an ongoing manner throughout the data collection and analysis process.*

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

44. **35f.** Interpretation of results: The interpretation of consensus (not) gained should be presented together with the meaning of the results and direction of further research needed.

    - 1. Definitely not required
    - 2. Possibly not required
    - 3. No strong opinion
    - 4. Possibly required
    - 5. Definitely required
    - Do not know

2nd Round: Reporting criteria for methods to elicit expert judgements or expert opinion in model-based economic evaluation

Page 7

Additional comments

45. Please use this space to write any additional comments you have either about the use of reporting criteria or the design of this survey.

46. Please provide your email address below so we can distribute our findings and recommendations. *

Appendix 3

Analysis of free text comments

Method

All respondents who completed the survey (both rounds) were asked to offer their views, using free text boxes, on the definitions (see Box 1), individual statements and also general comments at the end of both rounds of the survey. These comments were collated in an Excel file for each question and grouped into key themes and summarised for each round of the survey.

Box 1: Definitions of core concepts offered in Round One

> An expert is: 'someone who has in-depth knowledge of the topic of interest gained through their life experience, education or training'.
>
> A study designed to generate '**expert parameter values**' uses: 'a quantitative elicitation method to: derive point estimates and distributions for model input parameters; and/or pool expert judgement.'
>
> The purpose of an **expert elicitation study** is: 'to quantify a parameter value that appropriately represents the knowledge/judgement of the expert and the degree of uncertainty in that knowledge/judgement'.
>
> A study designed to collate '**expert opinion**' uses: 'a qualitative consensus method (e.g. a Delphi method or other approach) to collate views from experts to: frame the scope of the model-based economic evaluation; inform model conceptualisation; identify model face validity; quantify point estimates without specifying a distribution; pool elicitation results'.

Results: round one

Definitions

Of the 17 round one respondents, on average around half provided some comments on each of the offered definitions of key terms (7 comments on 'expert'; 10 comments on 'expert parameter values'; 8 comments on 'expert elicitation study'; 11 comments on 'expert opinion').

Four themes emerged from the free text comments on the definition of 'an expert' from seven respondents. Two respondents, agreed with the definition, but pointed out that 'an expert' was often someone who is available for your study and their participation may not always be reliant on their level of expertise. Two respondents (one who agreed with the definition; and one who disagreed with the definition) stated that it is often difficult to define 'in-depth' knowledge. Related to this comment, one respondent (who strongly agreed with the definition) added that 'life experience' is also relevant. One respondent stated that they

agreed with the definition offered but it would only be relevant to 'an expert' taking part in an elicitation study. One respondent neither agreed or disagreed with the definition and was concerned that it only considered 'substantiveness and fails to consider other basic components of expertise such as normativeness'.

Five themes emerged from the comments on the definition of 'expert parameter values'. The most common theme, suggested by three respondents (one who 'agreed; two who neither agreed nor disagreed), was that it was not always necessary to include 'derive…distributions' and respondents suggested adding 'or' in the definition to indicate this. One respondent, who agreed with the definition, wanted to stress that sometimes it is the uncertainty and not the value that is generated. One respondent disagreed with the definition as they did not like the implied use of confidence intervals suggesting the use of one-way sensitivity analysis. Two respondents offered their own definitions suggesting 'elicitation methodology is used to generate quantitative expression of parameters of interest' or the use of the term 'hyper-parameter values'. One respondent, although agreeing with the definition, did not like the phrase 'generate expert parameter values'.

Two themes emerged from the free text comments on the definition of 'expert elicitation study'. Five respondents (one strongly agreed; one agreed; one neither agreed nor disagreed; one disagreed; one strongly disagreed) queried how 'probability distribution' appeared in the definition. Three suggesting that the term was restrictive but one respondent felt the term 'probability distribution' was redundant as that is what 'degree of uncertainty' means. One respondent wanted the definition to be clearer in distinguishing between deterministic elicitation of a point estimate and probabilistic elicitation of a distribution. Three respondents (one agreed; one disagreed; one neither agreed nor disagreed) suggested that there are other uses of an elicitation study.

The definition of 'expert opinion' generated the most free text comments. Six themes emerged. One respondent strongly disagreed to the suggested distinction between eliciting expert opinion compared with identifying a parameter value '*If this is the definition, then I have never wanted to collated expert opinion. I guess I'm not interested in opinion per se, unless it is quantified as judgement*'. Three respondents (one did not know; two neither agreed nor disagreed) did not like the omission of specifying a distribution from the definition. Two respondents (one agreed; one disagreed) did not like the restriction to a

'qualitative' method in the definition. Two respondents (one agreed; one disagreed) questioned the appropriateness of referring to consensus methods as consensus is not always required. Two respondents (one disagreed; one strongly disagreed) did not like the reference to the use of the Delphi as the suggested method. One respondent (disagreed) said you could still use the method to generate a range of values but in an extreme scenario analysis rather than in a probabilistic sensitivity analysis.

Individual statements and general comments

Of the 17 round one respondents, seven of them provided some free text comments about individual statements or offered general comments at the end of the survey. There were two themes raised about the statements for reporting expert elicitation study and relevant to reporting a Delphi survey, respectively.

One respondent suggested that external factors affecting the practical application of the expert elicitation study should be reported explicitly to acknowledge when the design was modified to take into account experts who failed to complete the exercise or time constraints and availability of relevant experts. Another respondent pointed out that it is important to report whether there was feedback of the elicitation results to experts to provide some validation of the data.

One respondent pointed out that it is also important to provide feedback to respondents as part of the Delphi process. Another respondent stressed the importance of describing any changes made to the survey field in each round of the Delphi if the data from each round are not presented in the paper.

The general comments centred on the use of the Delphi approach. Six respondents raised serious concerns about the use of the Delphi approach as a method in the context of economic evaluation, in general, and for behavioural aggregation specifically. Respondents stressed that the Delphi is not the only approach to behavioural aggregation and other methods should be considered. Some respondents strongly objected to the use of behavioural aggregation methods, such as the Delphi, in the context of economic evaluation, in general, and also in collating quantitative estimates for parameter values and/or generating measures of uncertainty. One respondent added that more stringent reporting criteria should be used when a Delphi has been used because the dynamics of the process can hide important factors that

may influence the synthesis of data. Another respondent added that the results of the Delphi will be strongly influenced by the skills of the facilitator, which should be acknowledged when reporting the methods.

Three additional comments were made about the role of reporting criteria, describing the experts and the difference between the reporting criteria required when the expert elicitation or expert opinion study is reported as a standalone paper rather than within a paper reporting the economic evaluation. One respondent pointed out that reporting criteria also have a role in terms of contributing to a better understanding of the potential and limitations of using expert elicitation. Another respondent stated that it is important to know if the experts included in the study had previous roles in clinical trials of the intervention under evaluation. One respondent pointed out that although an understanding of what could be left out when reporting an expert elicitation or expert opinion study within an paper reporting the economic evaluation, this would be largely dependent on how critical the evaluation results is on the elicitation findings and how much journal space is available.

Results: round two

The majority of the free text comments offered by respondents when completing round two of the Delphi survey focussed on the definitions (expert parameter values; expert elicitation study; expert opinion).

Four respondents made comments on the definition ‘expert parameter values’ which can be grouped under four distinct themes. One respondent did not agree with the use of the word ’quantitative’ as expert opinion is ‘qualitative’ but aggregation is ‘quantitative’. This is what we were trying to capture in the definition. One respondent wanted the definition to be more explicit in describing what judgements were to be pooled. One respondent wanted the definition to be clearer in explaining that the elicitation method is used to define distributions not point estimates. One respondent felt it was difficult to agree a standardised definition of this concept as it would always be context specific.

Four respondents made comments on the definition ‘expert elicitation study’ using four distinct themes. One respondent did not agree with being explicit that the Delphi should not be used in the context of an expert elicitation study and felt the definitions were trying to be too specific. One respondent reinforced that the expert elicitation study should take a

Bayesian perspective, and therefore, identify the distribution and not the point estimate of a parameter value. One respondent felt the definition was badly phrased but did not offer suggestions on improvements. One respondent felt it was difficult to agree a standardised definition of this concept as it would always be context specific.

Three respondents offered comments on the definition of 'expert opinion' which was fewer respondents than had disagreed with the definition. Two respondents that did give comments focussed on the difficulty in trying to agree a standard definition when it depends on the context in which the concept is used. One respondent did not like the distinction being made between elicitation versus opinion and suggested the concepts could be made distinct by using the terms quantitative and qualitative.

Individual statements and general comments

Two respondents wrote comments about the criteria for which consensus had been agreed in round one. One respondent wanted clarification on whether the criterion referring to the measure of performance for data aggregation for an expert elicitation study referred to *'fitting of distributions to data or accuracy of experts (calibration)?'* One respondent was not sure what was being referred to in the criterion for data collection for a Delphi survey and questioned *'is this the recruitment criteria? Or face to face versus online?'*

Three respondents also offered general comments at the end of the survey. One respondent pointed out that the survey would not allow free text comments to be given without clicking a radio button. This was beyond the control of the research team as it was an annoying feature of the on-line survey software. One respondent found it difficult to conceptualise the difference between reporting criteria for a standalone compared with combined expert judgement and economic evaluation paper. One respondent pointed out that it can be difficult to be prescriptive on which methods are the appropriate ones to be used but it is important to be clear in the reporting of whatever method is used.

Summary

Collating free text comments during the completion of both rounds of the Delphi survey allowed further insights on the views of the experts completing the closed end questions that the degree of agreement with four core definitions and the suggested reporting criteria. The comments on the definitions from round one were used to modify two of the definitions to try

to work towards a consensus view on the key concepts. Consensus in the panel was achieved for three of the definitions but not for the definition of the term ‘expert opinion’.

The majority of the negative comments made focussed on the Delphi survey as a method. Some panel members felt that we were being overly restrictive by referring to the Delphi as the consensus method. Some members of the expert panel questioned the use of the Delphi and focussed on whether it is appropriate to use this method to produce quantitative estimates rather than collate opinions on care pathways or model structures.

Appendix 4

Level of consensus achieved and response distribution from each round of the Delphi survey

Criteria	1st round consensus (%)	Status after first round	2nd round consensus (%)	Status after second round	1st round median (30th-70th percentile)	2nd round median (30th-70th percentile)
Definitions						
1. Definition of an 'expert'	88%	Consensus	N/A	Consensus	4 (4 - 4.6)	
2. Definition of 'expert parameter' values'	71%	No consensus	83%	Consensus	4 (3.4 - 4)	4 (4 - 4)
3. Definition of an 'expert elicitation study'	65%	No consensus	83%	Consensus	4 (3 - 5)	4 (4 - 5)
4. Definition of 'expert opinion'	24%	No consensus	50%	No consensus	3 (2 - 3.2)	4 (2.6 - 4)
Eliciting expert values						
5. Research rationale	88%	Consensus	N/A	Consensus	5 (4 - 5)	
6. Research problem	88%	Consensus	N/A	Consensus	5 (4 - 5)	
7. Measurement type of uncertain quantities	94%	Consensus	N/A	Consensus	5 (4 - 5)	
8. Definition of an expert	82%	Consensus	N/A	Consensus	5 (4 - 5)	
9.Number of experts	88%	Consensus	N/A	Consensus	5 (5 - 5)	
10. Preparation	82%	Consensus	N/A	Consensus	5 (4 - 5)	
11. Piloting	94%	Consensus	N/A	Consensus	4 (4 - 5)	
12. Data collection	94%	Consensus	N/A	Consensus	5 (5 - 5)	
13. Administration	94%	Consensus	N/A	Consensus	5 (4 - 5)	
14. Training	82%	Consensus	N/A	Consensus	4 (4 - 5)	
15. The exercise	82%	Consensus	N/A	Consensus	5 (4 - 5)	
16. Data recording	41%	No consensus	67%	No consensus	3 (3 - 5)	4.5 (3 - 5)
17. Data aggregation	100%	Consensus	N/A	Consensus	5 (5 - 5)	
18. Measures of performance for data aggregation	82%	Consensus	N/A	Consensus	4.5 (4 - 5)	
19. Ethical issues	53%	No consensus	75%	Consensus	4 (3 - 5)	4.5 (3.9 - 5)
20. Presentation of results	94%	Consensus	N/A	Consensus	5 (4.4 - 5)	
21. Interpretation of results	100%	Consensus	N/A	Consensus	5 (5 - 5)	

Criteria	1st round consensus (%)	Status after first round	2nd round consensus (%)	Status after second round	1st round median (30th-70th percentile)	2nd round median (30th-70th percentile)
Collating expert opinion						
24. Research problem	93%	Consensus	N/A	Consensus	5 (4.8 - 5)	
25. Research rationale	71%	No consensus	75%	Consensus	4.5 (3.5 - 5)	5 (3.9 - 5)
26. Literature review	73%	No consensus	75%	Consensus	4 (3.8 - 5)	4 (3.9 - 4.1)
27. Data collection	87%	Consensus	N/A	Consensus	5 (4.8 - 5)	
28. The survey	67%	No consensus	83%	Consensus	4 (2.8 - 5)	4 (4 - 5)
29. Rounds	87%	Consensus	N/A	Consensus	5 (4.8 - 5)	
30. The sample	80%	Consensus	N/A	Consensus	5 (4 - 5)	
31. Ethical issues	53%	No consensus	67%	No consensus	4 (3 - 4.2)	4 (3 - 4.1)
32. Data analysis	80%	Consensus	N/A	Consensus	4 (4 - 5)	
33. Presentation of results	80%	Consensus	N/A	Consensus	5 (4 - 5)	
34. Interpretation of results	80%	Consensus	N/A	Consensus	5 (4.8 - 5)	

Key: 1. Definitely not required; 2. Possibly not required; 3. No strong opinion; 4. Possibly required; 5. Definitely required.