Using consensus research methods with multi-stakeholder groups: does one size fit all?

Joanne Coster

Research Fellow

CURE, ScHARR

- Widely used in health research where decision making or prioritising is required
- Aim to determine the extent of agreement between experts/lay people and attempt to resolve disagreement
- Useful where evidence is contradictory, low quality
- Useful alternative to a systematic review in synthesising information
- Representative – allow involvement of multiple perspectives
- Inclusive, allow a wide range of evidence to be considered

Types of consensus methods

- Most commonly used consensus methods in health research are the nominal group technique, consensus conference and the Delphi method

- These can be used alone, with other research methods or sequentially where there are numerous stages of consensus research

- Using formal consensus methods helps to ensure research is robust, systematic, reportable and accountable.

- Formal consensus methods add weight to the outcome of the consensus process.

- Consensus methods have been developing for the past 80 years, with early consensus research using the Delphi method and more interactive consensus methods, such as consensus conferences introduced in the late 1980s.

Processes in consensus methods

Consensus method	Question-naire	Private decisions elicited	Formal feed-back of group choices	Face-to-face contact	Interac-tion structured	Aggreg-ation method
Delphi	Y	Y	Y	N	Y	Explicit
NGT	N	Y	Y	Y	Y	Explicit
Consensus conference	N	N	N	Y	N	Implicit

Different stages and process to each type of consensus method, summarised in the table taken from a HTA monograph

Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J,et al. Consensus development methods, and their use in clinical guideline development.Health Technol Assessment 1998;2(3)

- Development of clinical guidelines
  - Where evidence is of low quality or there is very little evidence, guideline development can be based on expert views

- Prioritising important areas for health research

- Prioritising and developing indicators and quality and performance measures

- To identify a set of potential performance measures for systems of emergency and urgent care, (Delphi study)
- The PhOEBE project, which used consensus method to develop better ways of measuring the quality and performance of ambulance care

PhOEBE study

- Aimed to develop better ways of measuring the performance, quality and impact of ambulance service care

- Used consensus methods with a range of patients, public and professional stakeholders to develop a set of prehospital performance indicators

Pre-hospital Outcomes for Evidence Based Evaluation

Stages in the PhOEBE study

Systematic reviews to identify measures
- Measures used in research
- Aspirational measures

Consensus event (n =42)
- Refine and prioritise the measures
- Including online time measures survey

Delphi study (n =20)
- Further refine and prioritise the measures

PPI (n = 18) consensus conference

Small group expert panel reviewed measures against good outcome measure criteria

- No consensus about what should be measured
- Iterative approach to refine a long list of measures into a smaller list
- To include and reflect a range of participants and perspectives
- For pragmatic reasons, due to the number of measures identified it was useful to have sifting phase
- To address PPI concerns about the suitability of questionnaire based consensus methods. PPI preferred a face-to-face approach
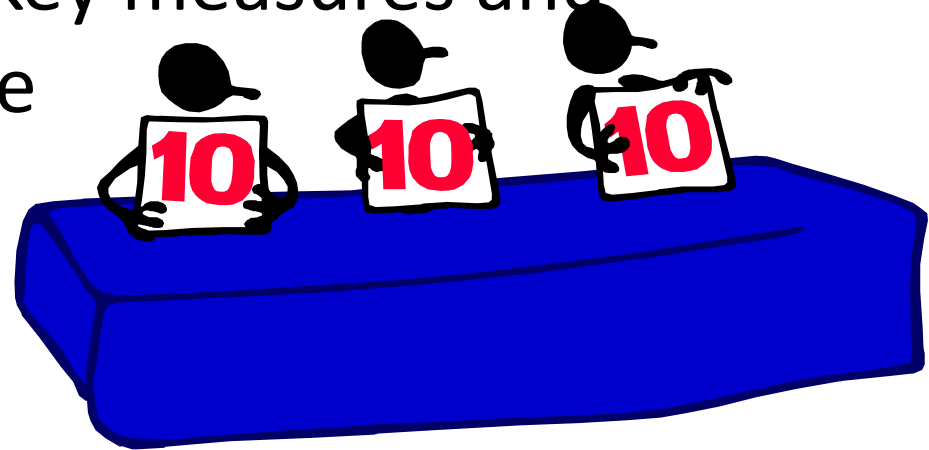
- Systematic review of current measures
- 405 measures/151 papers
  - Patient outcomes 13%
  - Survival and time 60%
- Why measure time and survival?
  - Easy to measure
  - Easy to record
- Issues
  - Only applicable to a small patient group
  - No information about quality of care or patient views What about clinical need, patient experience or effectiveness?
- Policy review aspirational measures
  - More patient outcome based measures
  - Balanced score card approach
    - a suite of measures rather than single measures

- Reviews informed the consensus process

- Large number of time measures – prioritised using an online form

- Consensus conference – used modified nominal group technique with small group discussion and vote of key measures and concepts from literature

Time measures

- Most commonly collected and reported measure

- Pragmatic decision due to large number of time measures

- 23 different time interval measures

- Most common is call to scene

- Recognised as having little relevance or value

- Online survey: which time measures are most useful?

- Highest ranking time measures were taken forward into the Delphi survey

Aim – to prioritise potential measures for measuring ambulance service quality and performance

- 1 day event, small group discussions, voting

  – Ambulance service

  – Patient and public

  – Commissioners

  – Policy makers

  – Academic research

- NGT is a structured group meeting of experts with the process led by a moderator
- Allows face-to-face interaction and discussion between participants
- Modified to incorporate electronic voting and to include our identified candidate measures as a starting point for group discussions
- Small group discussions for each group of measures (whole service measures, clinical management measures, patient outcomes)
- Discussions facilitated by members of the research team
- Participants were encouraged to think of additional measures to share with the group and each participant had an opportunity to contribute

- Discussion sessions were immediately followed by voting to rank the importance of each measure

- Participants voted using an anonymous audience response voting system

- Asked to decide whether each measure was essential, desirable or irrelevant by pressing a button on a handset

- 42 measures/principles, of which 20 were highly prioritised by >50% participants and of these 9 were highly prioritised by >75%

- Measures that were rated highly were included in the next phase of the consensus process (Delphi study)

- Beneficial for identifying what to exclude from subsequent stages

- Paramedics and operational staff, Clinical Directors, research and audit staff, members of NARSG and NASCQG, ambulance commissioners, emergency care physicians

- 67 measures included

- More measures than in the NGT consensus event!
  - Time measures included
  - Broader measures or concepts, such as accuracy of dispatch decisions were included as multiple specific measures

Delphi study method

- Used a RAND based Delphi approach
- "a group of experts who anonymously reply to questionnaires and subsequently receive feedback in the form of a statistical representation of the 'group response', after which the process repeats itself"
- Two rounds. A third was unnecessary due to little change between the rounds
- Round 1: participants scored each measure, gave text comments, suggested additional measures or revisions to existing measures
- Round 2: provided each participant with their individual score, the group median score for each measure, all the text comments from the previous round and a small number of revised measures, based on the round 1 comments.

- Is this measure a good reflection of the quality of care/service provided by the ambulance service and is likely to be a good indicator of the quality of the 999 ambulance service care pathway

- Each measure scored on a 1 – 9 scale (strongly disagree – strongly agree)

- To analyse, we ranked measures by their median scores to classify whether measures achieved a "good," "moderate" or "poor" level of consensus

The University Of Sheffield.

School Of Health And Related Research

Results

Measures reaching consensus as good measures of quality of care

Measure description
Proportion of patients who report pain who are given pain relief
Proportion of all 999 calls referred for telephone advice only re-contacting the ambulance service within 24 hours
Time of call to time of definitive care
Proportion of category A calls correctly identified as category A
Number of patient safety incidents reported as a proportion of all requests for 999 ambulance care
Proportion of all cases with a specific condition who meet the established criteria for transfer, who are transported to an appropriate specialist facility, for example a heart attack, stroke or major trauma centre
Proportion of cases that comply with end of life care plans where these are available
Proportion of all cases with a specific condition who are treated in accordance with established protocols and guidelines, for example stroke, heart attack, diabetes, falls

- Large number of measures

- Lots of very technical language that did not easily translate to a lay audience

- Attempted two versions of the Delphi, 1 for clinical and policy, and a simplified one for PPI

- Difficult because the meaning could be lost or misinterpreted

- New approach needed!

The
University
Of
Sheffield.

School Of
Health
And
Related
Research

How best to include PPI?

- Problems with the Delphi method: Research jargon, difficult concepts requiring a face-to-face explanation and discussion, inclusivity, no opportunity to ask questions

- How were these resolved?
  - Lots of discussion and meetings with our PPI colleagues
  - PPI led consensus event to complement the Delphi survey. PPI tailored information sessions, small discussion groups, electronic voting
  - meaningful PPI engagement with highly technical, complex and often little known aspects of ambulance performance

PPI consensus event

- Our specific objectives for the event were for participants to:

1. Understand the work undertaken by the PhOEBE project so far
2. Have an opportunity to discuss performance measures and why they were needed
3. Choose measures which they considered most important
4. Feel they had been involved and their views listened to
5. Understand how the event contributed to the process of selecting ambulance service measures
6. Understand how the measures selected would be used in the next steps of the PhOEBE project

- "The idea was for the research team to present how ambulance services work from the initial call; problems they face; what the PhOEBE project is and progress so far; presentation of the measures for consideration by lay people; discussion of measures in small groups; voting individually on preferences; conclusion; feedback on the day and results. Each section was to be about 15 minutes long; using video clips where appropriate and giving time for questions and answers before moving on to the next section. A glossary of technical language in plain English I also considered necessary and wrote it with the help of the research team"
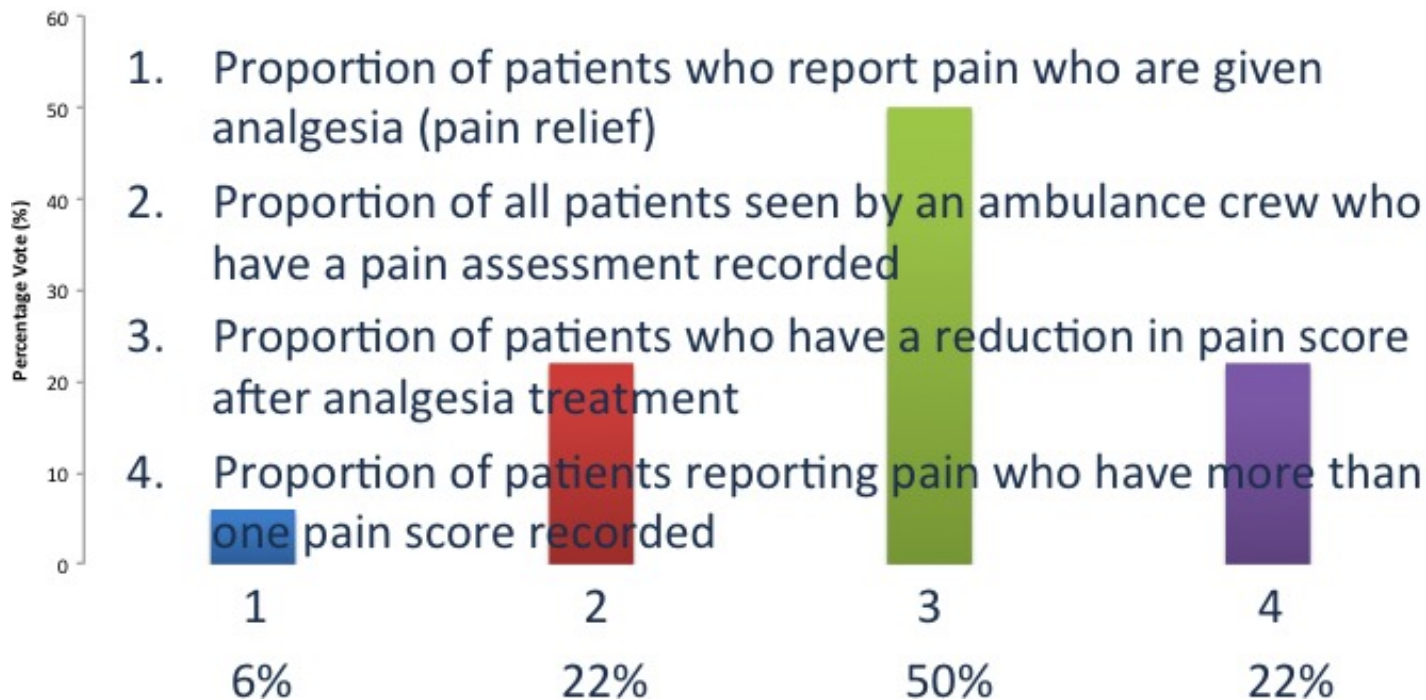
(Maggie Marsh, PPI rep)

Pain measures

1. Proportion of patients who report pain who are given analgesia (pain relief)
2. Proportion of all patients seen by an ambulance crew who have a pain assessment recorded
3. Proportion of patients who have a reduction in pain score after analgesia treatment
4. Proportion of patients reporting pain who have more than one pain score recorded

1	2	3	4
6%	22%	50%	22%

What did new approach this achieve?

- Better understanding of the issues and importance of the topic
- Wider PPI involvement than initially planned and a more meaningful contribution
- Patient and public feedback on complex, little known aspects of ambulance service performance measurement
- PPI felt valued as collaborators
- Proactive and pragmatic PPI research model, which found ways to adapt research methods so they are meaningful to lay people
- The PPI event produced a clear indication of which measures were most highly favoured by participants
- Poster at the Involve conference: Revolution in PPI….or Revolt? A meeting of minds or a clash of opinions?

- Not possible to include all (61) measures from the Delphi survey in the PPI workshop

- Included measures based on the following criteria: those with the highest Delphi scores or related to categories of measures rated highly by Delphi participants, identified as important to PPI in our preliminary work or that our PPI reference group felt should be included.

- For example, only one pain measure was highly scored by Delphi participants. But pain was identified as important by PPI representatives in all our preliminary work, and our PPI reference group thought it was important

- We were able to obtain PPI views on a complex subject for the majority of measures from the Delphi study (35/61, 57%).

- Practical limitations regarding how many measures the PPI participants could feasibly consider during a 1 day event, given that each measure required substantial explanation and group discussion.

- Also limitations on the amount of time PPI were able to contribute to the day, as well as travelling distances and potentially complex health problems to consider for participants.

- Takes time, patience and teamwork

- How did we arrive at a final list of measures?
- Convened a small expert group to consider which measures should be further developed as part of the PhOEBE research programme
- Aimed to select a set of measures that represented and assessed the quality of a service
- Measures were considered against good indicator criteria, Delphi study rating; PPI consensus conference rating
- A shortlist of eight measures was selected for development

Conclusions

- We used a multiple stage consensus process to prioritise ambulance outcomes and performance measures

- We've since tested and further developed the measures using a routine linked ambulance and other health services dataset

- The measures that we developed reflected current pre-hospital ambulance care and services and were important to ambulance clinicians, service providers, commissioners, clinical academics and PPI

- We had to be pragmatic and flexible in our approach in order to involve all stakeholders in a meaningful way

- Developing alternative methods, such as the PPI consensus event, the online survey for time measures and modifying the NGT conference to include voting so that the results could be more easily interpreted

- Challenging process, which involved listening and responding, particularly to PPI, so that there was meaningful involvement

Publications

- Coster JE, Irving AD, Turner JK, Phung VH, Siriwardena AN. Prioritizing novel and existing ambulance performance measures through expert and lay consensus: A three-stage multimethod consensus study. *Health Expectations*, 2017. https://doi.org/10.1111/hex.12610

- Irving AD, Turner J, Marsh M, Broadway-Parkinson A, Fall D, Coster J & Siriwardena AN A coproduced patient and public event: An approach to developing and prioritizing ambulance performance measures. *Health Expectations,* 2017.

- https://doi.org/10.1111/hex.12606