The use of an e-Delphi method to identify core maternity outcome measures

Colm O’Boyle, Dell Horey, Mike Clarke, Cecily Begley, Declan Devane.


Background
Variation in maternity care outcome measures, both in the variety of outcomes reported and disparity in the definition of terms, makes comparison between studies and meta-analyses highly problematic. The aim of this Delphi study was to identify a minimum data set of core outcome measures that could be used to evaluate models of maternity care and provide the basis for comparisons between models.

Methodology
A three round, electronic Delphi study design was used to reach consensus among key maternity care stakeholders internationally. The use of electronic, on-line resources allowed the participation of a wide variety of stakeholders and greatly facilitated international collaboration. Maternity service users, midwives, obstetricians, general practitioners, policy makers, and maternity care researchers from 28 countries (n=320) expressed willingness to take part. Of these, 218 (68%) completed round 1, of whom 173 (79%) completed round 2 and 152 (88%) of these completed round 3. For breakdown of participants see below.

Development of the instrument for Round 1
Systematic review of randomised trials comparing models of maternity care identified 263 outcome measures. These were grouped under five broad headings (Antenatal, Intrapartum, Postnatal, Fetal/Neonatal and Additional). A draft instrument was circulated to the research team before being line resources allowed the participation

320 participants were asked to rate the importance of each of the identified outcomes for inclusion in a minimum dataset using a five-point Likert scale. Participants also asked to report up to two ‘new’ outcomes under each heading. A short, online demographic questionnaire included rating of self-assessment of level of expertise in evaluation of maternity care models.

Data collection-Round 2
The instrument for Round 2 was developed from analysis of responses to round 1 and was re-sent to all participants who had responded to round 1. Outcomes were included in round 2 if the group mean was greater than the mean score for all the outcomes combined, OR if participants who self-assessed their level of expertise in the evaluation of maternity care models (6 or 7 in a 7 point Likert scale) had a mean for that outcome greater than the mean score for all the outcomes combined. ALSO, all newly identified outcomes suggested by two or more participants in Round 1 (73 items) were included in Round 2. Participants were asked to rate / re-rate the importance of each outcome in the Round 2 instrument.

Data collection-Round 3
The instrument for Round 3 was developed in a similar fashion to Round 2 except that outcomes were also included if 70% or more gave an importance rating of ‘4’ or ‘5’ on the 5-point Likert scale in Round 2. Again the overall group’s mean rating and standard deviation for each included outcome were presented, with the participant’s individual rating. Participants were re-assessed to re-rate the importance of each outcome in the Round 3 instrument.

Results - Round 3 presented in next column
Outcomes were retained from Round 3 when the mean score given to the outcome was greater than the overall average score for all outcomes combined (black), OR if 70% or more of the participants gave an importance rating of ‘4’ or ‘5’ on the 5-point Likert scale in Round 3 (red).

Conclusion
Delphi as a technique to elicit and condense opinions towards consensus has the benefits of anonymity and low cost; but ambiguities regarding defining consensus and ‘expertise’ remain problematic. In this study, self-identified expertise as a marker for retaining items in round one, and a combination of means and 70% or more agreement, were attempts to ameliorate those ambiguities. The admittedly rather long data set presented could be shortened by further iterations or tighter selection criteria, but whatever the number of items, no set should be presented as prescriptive or restrictive. People involved in the planning and conduct of research remain free, of course, to supplement this core set with other outcomes that they deem important to their evaluation. If adopted however, this core set could be useful to identify primary outcome measures for multicentre trials and would allow more fruitful comparisons of models of maternity care both nationally and internationally.

Black = Included on basis of being greater than overall mean of all outcome means of 4.18 (rounded down to two decimal places) and where ≥70% of participants rated outcome a 4 or 5 (where 5 is max rating) but where mean for specific outcome is < 4.18

Red = Outcomes for which ≥ 70% of participants rated outcome a 4 or 5 (where 5 is max rating)