Evaluating Maternity Care: A Core Set of Outcome Measures

Declan Devane, Cecily M. Begley, Mike Clarke, Dell Horey, Colm OBoyle

ABSTRACT:

Background: Comparing the relative effectiveness of interventions on specific outcomes across trials can be problematic due to differences in the choice and definitions of outcome measures used by researchers. We sought to identify a minimum set of outcome measures for evaluating models of maternity care from the perspective of key stakeholders.

Methods: A 3-round, electronic Delphi survey design was used. Setting was multinational, comprising a range of key stakeholders. Participants consisted of a single heterogeneous panel of maternity service users, midwives, obstetricians, paediatricians/neonatologists, family physicians/general practitioners, policy-makers, service practitioners, and researchers of maternity care. Members of the panel self-assessed their expertise in evaluating models of maternity care.

Results: A total of 320 people from 28 countries expressed willingness to take part in this survey. Round 1 was completed by 218 (68.1%) participants, of whom 173 (79.4%) completed round 2 and 152 (87.9%) of these completed round 3. Fifty outcomes were identified, with both a mean value greater than the overall group mean for all outcomes combined (x = 4.18) and rated 4 or more on a 5-point Likert-type scale for importance of inclusion in a minimum data set of outcome measures by at least 70 per cent of respondents. Three outcomes were collapsed into a single outcome so that the final minimum set includes 48 outcomes.

Conclusions: Given the inconsistencies in the choice of outcome measures routinely collected and reported in randomised evaluations of maternity care, it is hoped that use of the data set will increase the potential for national and international comparisons of models for maternity care. Although not intended to be prescriptive or to inhibit the collection of other outcomes, we hope that the core set will make it easier to assess the care of women and their babies during pregnancy and childbirth.

The use of randomised trials to evaluate models of care is well established in many areas of health. Although they can provide evidence about the effectiveness of interventions on specific outcomes, comparing the relative effects across trials can be problematic due to differences in the choice and definitions of outcome measures used by researchers.

In maternity care, the assessment of relative safety will often be an important, if not the principal, rationale for study. This characteristic is reflected in the use of outcome measures such as perinatal mortality. However, 2 problems arise with this approach. First, safety is an ambiguous concept that lacks a universally agreed definition and whose meaning often depends on professional interests (1). For example, some studies have related safety to outcome measures, such as mode of birth (2),

whereas others have focused on perinatal mortality (3). Second, the precise definitions of important outcome measures, such as perinatal mortality, vary among countries, making international comparisons difficult (4). Furthermore, the use of relatively rare maternity care outcomes, especially in some populations (e.g., women at low risk of complications), may require trials that are too large to be practical. This factor means that studies evaluating models of maternity care often need to use other outcome measures that are sensitive enough to detect significant differences. Outcomes such as rates of maternal and infant morbidity and maternal satisfaction have been used, although they are not common to all studies, and variable definitions can be used.

A systematic review of randomised trials of models of maternity care (to be reported separately) has identified wide heterogeneity in the outcome measures collected and reported. The use of a consistent set of outcome measures would be helpful to evaluate or compare models of maternity care, which might best be achieved through the development of a minimum core set of outcome measures, ideally taking account of the views and needs of key stakeholders. A core data set would be useful within multi-centre trials and for comparisons between trials. It might also be a significant step in facilitating useful meta-analyses of similar studies. To meet this need, we sought to identify a minimum set of outcome measures to evaluate models of maternity care through the use of an international, electronic Delphi (e-Delphi) survey.

Methods

Design

The Delphi method is iterative and uses a series of rounds of data collection and analysis to condense the opinions of individuals into group consensus. Typically, it involves the use of sequential rounds of postal questionnaires that are designed to elicit participants' opinions on a particular topic. Responses to each round are collated, analysed, and redistributed to participants for further comment in successive rounds.

We chose to conduct our e-Delphi survey online, so as to facilitate (a) international participation without the time lag between successive rounds associated with traditional postal surveys, (b) a relatively low cost structure, (c) data collection efficiencies, and (d) the potential for a higher response rate through rapid communication with participants. The survey was conducted using the online survey software QuestionPro (http://www.questionpro.com).

Participants

Participation was sought from people within the following broad groups: women as maternity service users, midwives, obstetricians, paediatricians/neonatologists, policy-makers, and service providers and researchers with expertise in maternity care research.

We sent an e-mail inviting participation to the following electronic discussion lists, which we had identified as relevant to the broad area of maternity care: Midwifery-Research, Normal-Birth-

Research, Irish-midwifery-eGroup, Cochrane Pregnancy & Childbirth Reviewers' Group, Cochrane Pregnancy & Childbirth consumer networks, International Confederation of Midwives network of research advisors, European Perinatal Epidemiology Network, and Birth Centre Network, United Kingdom. We also used purposeful sampling to approach people with known expertise in the evaluation of maternity care. Snowball sampling was then used, whereby participants from the above groups were asked to forward the invitation to others whom they regarded as having the required expertise.

The invitation e-mail informed potential participants of the purpose of the study and invited those who wished to participate to respond with their name and individual e-mail address.

Instrumentation

Pilot study

An initial instrument was developed containing 299 outcomes identified from the aforementioned systematic review of randomised trials comparing 2 or more models of maternity care (to be reported separately). Outcome measures addressing similar dimensions or events were discussed by the team and collapsed where possible. For example, various modes of delivery/birth were presented as "mode of birth (e.g., spontaneous vaginal, forceps, vaginal breech, caesarean section, vacuum extraction)." This pilot tool was tested for clarity, with a sample of 12 participants, including 3 maternity care consumers, and subsequently refined.

Round 1

After piloting, the round 1 instrument contained 263 outcomes, categorized as antenatal (*n* = 80), intrapartum (*n* = 64), postnatal (*n* = 59), foetal/neonatal (*n* = 49), and additional (*n* = 11).
We chose to group the outcomes within these 5 broad categories rather than presenting them randomly to participants. Furthermore, related outcomes were presented alongside each other (e.g., neonatal anthropometric outcomes of birthweight, length, and head circumference). The same format was presented to all participants.

Participants were asked to rate the importance of each outcome listed using a 5-point Likert-type scale rating their importance for inclusion in a minimum set as: 1 = of no importance, 2 = of some importance, 3 = of moderate importance, 4 = very important, and 5 = extremely important. Participants were also asked to identify up to 2 "new" outcomes under each of the 5 broad headings, which they judged to be relevant or important.

The round 1 instrument included a short online demographic questionnaire and also invited participants to rate their perceived level of expertise in evaluating models of maternity care on a 7-point Likert-type scale (1 = low level of perceived expertise, 7 = high level of perceived expertise).

A copy of each participant's response to every round was e-mailed to them within 48 hours of completion of that round online. Each round had a response closing date 10 days after the date of invitation. An e-mail reminder was sent to anyone who had not responded by day 7.

Round 2

In round 2, participants who responded to round 1 were presented with outcomes retained after analysis of responses from round 1, where (a) the overall mean score for inclusion for that outcome was greater than the mean score for all the outcomes combined and (b) the mean score for inclusion for that outcome was greater than the mean score for all the outcomes combined for those participants who had rated their perceived level of expertise in evaluating models of maternity care as high (i.e., 6 or 7 on the Likert scale). Additional outcomes identified by participants in round 1 were included if suggested by at least 2 participants.

For each outcome retained from round 1, the overall group's mean rating and standard deviation were presented. Participants were asked to re-rate the importance of each outcome with knowledge of their individual and the group's previous ratings. In addition, participants were asked to rate the newly identified outcomes from round 1. All ratings used the same Likert-type scale that was used in round 1.

Round 3

In round 3, participants who responded to round 2 were presented with outcomes retained after analysis of responses from round 2, where (a) the overall mean score for inclusion for that outcome was greater than the mean score for all the outcomes combined and (b) 70 per cent or more of study participants rated their importance for inclusion as a "4" or "5" on the 5-point Likert-type scale used in round 2.

Each of the outcomes in the round 3 instrument was again presented together with the mean rating and standard deviation for the whole group, and participants were asked to re-rate the importance of each item for inclusion in a minimum data set using the same Likert-type scale used in round 2.

Outcome measures were retained from round 3 if the group mean for that outcome was greater than the mean score for all the outcomes combined and where 70 per cent or more of the participants gave an importance rating of 4 or 5 on the 5-point Likert-type scale.

Data analysis

All data were analysed using SPSS (5). Mean and standard deviations are rounded to 2 decimal places.

Ethics

Approval to conduct this study was received from the Faculty of Health Sciences Ethics Committee, Trinity College Dublin, Ireland. The online survey software system used maintains data behind a firewall. Only the researchers could access the data through use of a password and user identifier. A detailed "Study Information and Consent Agreement" form was sent to all potential participants, who had to click an "I Agree" button on the page to indicate their consent to participate.

Results

Participants

In all, 320 people replied to the e-mail invitation, indicating their willingness to participate; 218 (68.1%) completed round 1. Of these, 173 (79.4%) completed round 2 and 152 (87.9%) of these completed round 3. Table 1 contains a summary of response patterns in all 3 rounds.

Table 1. Response Patterns

Response Patterns	Number of Respondents	Per cent of Initial Participants
To all 3 rounds	152	47.5
To rounds 1 and 2 only	173	54.1
To round 1 only	218	68.1

Respondents to round 1 were based in 28 countries. Six countries accounted for 84.4 per cent (179) of respondents (the United Kingdom, Canada, Australia, Ireland, United States, and the Netherlands) (Table 2). Respondents to round 3 were based in 22 countries (Table 2).

Table 2. Countries Represented

Country	Round 1 Respondents No. (%)	Round 3 Respondents No. (%)
United Kingdom	69 (31.7)	48 (31.6)
Canada	38 (17.4)	27 (17.8)
Australia	28 (12.8)	23 (15.1)
Ireland	23 (10.6)	15 (9.9)
United States	21 (9.6)	14 (9.2)
The Netherlands	5 (2.3)	2 (1.3)
Peru	4 (1.8)	2 (1.3)
Sweden	4 (1.8)	3 (2.0)
Germany	3 (1.4)	3 (2.0)
Northern Ireland	3 (1.4)	2 (1.3)
Singapore	2 (0.9)	1 (0.7)
Switzerland	2 (0.9)	2 (1.3)
Other	15* (7.4)	10† (6.6)

* The following countries were represented by 1 respondent: Thailand, Zimbabwe, Saudi Arabia, Argentina, Georgia, New Zealand, Albania, Austria, India, France, Afghanistan, Brazil, Nigeria, Iran,

and Denmark.

† The following countries were represented by 1 respondent: Thailand, Zimbabwe, Saudi Arabia, Argentina, Georgia, New Zealand, Albania, Austria, Afghanistan, and Nigeria.

Most respondents to round 1 were midwives (45%, *n* = 98), with obstetricians representing 11 per cent (*n* = 24) of participants. Representatives from women's organizations made up 6.9 per cent (*n* = 15) of the total, whereas managers and nurses were equally represented (6.4%, *n* = 14 each). Nine women (4.1%) identified their primary interest in maternity care evaluation as being the perspective of women who had or who planned to give birth. Nine epidemiologists (4.1%), 8 neonatologists (3.7%), and 3 family physicians or general practitioners (1.4%) responded. The remaining 11 per cent (*n* = 24) of respondents selected "Other" to identify their primary interest in maternity care and reported being, for example, an "anaesthetist," "social scientist," and "lactation specialist" (Table 3).

Table 3. Respondents' Primary Interest in Maternity Care

Primary Interest	Round 1 Respondents No. (%)	Round 3 Respondents No. (%)
Midwife	98 (45.0)	71 (46.7)
Obstetrician	24 (11.0)	18 (11.8)
Nurse	14 (6.4)	9 (5.9)
Representative from women's organization(s)	15 (6.9)	13 (8.6)
Manager	14 (6.4)	6 (3.9)
Woman who has given birth or plans to	9 (4.1)	5 (3.3)
Epidemiologist	9 (4.1)	5 (3.3)
Neonatologist	8 (3.7)	5 (3.3)
Family physician/general practitioner	3 (1.4)	3 (2.0)
Other	24 (11.0)	17 (11.2)

In round 3, most respondents again were midwives (46.7%, *n* = 71), with obstetricians representing 11.8 per cent (*n* = 18) of participants. In this round, nurses accounted for 5.9 per cent (*n* = 9) of participants, whereas representatives from women's organizations made up 8.6 per cent (*n* = 13) of the total. Managers represented 3.9 per cent of respondents (*n* = 6), whereas 5 women (3.3%) identified their primary interest in maternity care evaluation as being the perspective of women who had or who planned to give birth. Five epidemiologists (3.3%), 5 neonatologists (3.3%), and 3 family physicians or general practitioners (2.0%) responded. The remaining 11.2 per cent (*n* = 17) of respondents selected Other to identify their primary interest in maternity care and reported being, for example, an anaesthetist, social scientist, and lactation specialist.

Most participants (76.6%, *n* = 167) identified either clinical care (40.8%, *n* = 89) or research (35.8%, *n* = 78) as the aspect of their work most related to the evaluation of maternity care, with similar numbers of participants identifying health care policy (8.7%, *n* = 19) and advocacy (8.3%, *n* =

18). Education was identified by 2.3 per cent (*n* = 5). Nine (4.1%) chose the category Other and provided responses such as “maternity care planning” and “assessment of standards of care.”

Round 1

The mean time taken to complete the round 1 survey was 34 minutes. The mean score for the 263 items included in round 1 was 3.69. The mean score for 151 of these items was higher than this score, making them eligible for round 2. An additional 7 outcomes with mean scores below 3.69 were retained because of the high score among participants who rated their perceived level of expertise in evaluating models of maternity care as a 6 or 7.

In addition, 73 outcomes were not included in the round 1 instrument but were suggested by 2 or more participants, resulting in 231 outcomes being included in the round 2 instrument.

Round 2

The round 2 instrument was completed by 173 people who had completed round 1 (79.4%). The mean time taken to complete the round 2 instrument was 30 minutes. The mean score for the 231 items was 3.90, and 105 outcomes had a mean score for inclusion in the data set that was higher than this score and had been rated as a 4 or 5 on the Likert-type scale (where 5 was the maximum rating) by at least 70 per cent of all respondents. These 105 outcomes were included in the round 3 instrument.

Round 3

The round 3 instrument was completed by 152 participants who had completed round 2 (87.9%). The mean time taken to complete round 3 was 23 minutes. The mean score for the 105 items was 4.18, and 50 outcomes had a mean score that was higher than this score and had been rated as a 4 or 5 on the Likert-type scale by at least 70 per cent of the respondents.

Three outcomes (maternal gestation at birth, gestational age of infant at birth, and preterm birth) were considered to be sufficiently similar to allow collapsing to a single outcome, that is, gestational age at birth. This action resulted in the final minimum set of 48 (i.e., 50 − 3 + 1) outcomes shown in Table 4 in descending order of mean score in round 3.

Table 4. Outcomes Where Mean Value is Greater than Overall Mean of all Outcome Means of 4.18 (Rounded Down to 2 Decimal Places) and Where Greater Than or Less Than 70 Per cent of Participants Rated Outcome as 4 or 5 (Where 5 is Maximum Rating)

Rank	*Outcome*	*Mean*	*SD*
1	Maternal death (the death of a woman while pregnant or within 42 days of termination of pregnancy)	4.91	0.33
2	Mode of birth (e.g., spontaneous vaginal, forceps, vaginal breech, caesarean section, vacuum extraction)	4.89	0.40
3	Neonatal death (death before the age of 28 completed days after live birth)	4.86	0.43
4	Stillbirth (a foetal death in late pregnancy)	4.83	0.47
5	Type of labour onset (manner in which labour started, i.e., induced, spontaneous, planned caesarean section)	4.68	0.66
6	Neonatal admission to special care and/or intensive care unit	4.62	0.59
7	Birth injury to infant	4.58	0.68
8	Ruptured uterus	4.56	0.75
9	Postpartum haemorrhage (excess blood loss from the birth canal after childbirth)	4.56	0.66
10	Mother requires admission to intensive care	4.55	0.63
11	Maternal postnatal readmission to hospital	4.50	0.67
12	Method of infant feeding	4.49	0.78
13	Vaginal birth after previous caesarean section (VBAC)	4.47	0.70
14	Gestational age at birth	4.47	0.72
15	Postnatal depression	4.46	0.75
16	Place of birth	4.45	0.80
17	Neonatal resuscitation required	4.44	0.72
18	Normal (i.e., physiological) birth without intervention (vaginal birth without induction, episiotomy, or epidural)	4.41	0.77
19	Oxytocin augmentation of labour (drug used to assist progress of labour)	4.39	0.76
20	Anal sphincter damage	4.37	0.74
21	Hypoxic ischemic encephalopathy (a condition of injury to the brain)	4.35	0.84
22	Intrapartum hypertensive disorders of pregnancy (a group of diseases characterized by high blood pressure with or without proteinuria; this group includes pre-eclampsia, eclampsia, and the syndrome of HELLP)	4.34	0.89
23	Hypertensive disorders of pregnancy (a group of diseases characterized by high blood pressure with or without proteinuria; this group includes pre-eclampsia, eclampsia, and the syndrome of HELLP)	4.33	0.91
24	Puerperal psychosis (a mood disorder often accompanied by features such as loss of contact with reality, hallucinations, severe thought disturbance, and	4.32	0.94

	abnormal behaviour)		
25	Maternal faecal incontinence	4.32	0.80
26	Birth asphyxia (occurs when a baby does not receive enough oxygen before, during, or just after birth)	4.32	0.88
27	Breastfeeding at discharge	4.31	0.93
28	Neonatal readmission to hospital	4.31	0.75
29	Apgar score at 5 min	4.31	0.86
30	Trial of labour after previous caesarean delivery	4.31	0.79
31	Breastfeeding at 3 months	4.30	0.84
32	Maternal satisfaction (postnatal)	4.30	0.94
33	Infant birthweight	4.30	0.83
34	Neonatal fitting/seizures	4.29	0.89
35	Infant requiring intubation	4.27	0.77
36	Congenital anomaly (chromosomal, genetic, and/or structural)	4.27	0.88
37	Use of pharmacological analgesia/anaesthesia (e.g., Entonox, epidural, pethidine)	4.25	0.80
38	Maternal satisfaction (antenatal)	4.25	0.92
39	Postnatal hypertensive disorders of pregnancy (a group of diseases characterized by high blood pressure with or without proteinuria; this group includes pre-eclampsia, eclampsia, and the syndrome of HELLP)	4.24	0.98
40	Maternal satisfaction (intrapartum)	4.24	0.95
41	Caesarean section wound infection	4.24	0.83
42	Pulmonary embolism (a condition in which a blood clot that has formed elsewhere in the body travels to the lungs)	4.22	0.90
43	Intrauterine growth restriction (commonly used when the birthweight is at or below the 10th per centile for gestational age and sex)	4.22	0.76
44	Preterm labour (onset of labour before 37 completed weeks of pregnancy)	4.20	0.87
45	Meconium aspiration (means the newborn inhales a mixture of meconium and amniotic fluid, either in the uterus or just after delivery)	4.19	0.89
46	Intrapartum haemorrhage (excessive blood loss from the birth canal during labour)	4.19	0.76
47	Neonatal infection	4.19	0.88
48	Shoulder dystocia	4.19	0.81
HELLP = haemolysis, elevated liver enzymes, and low platelets.			

Three additional tables are presented, showing (a) those outcomes for which at least 70 per cent of participants rated the outcome as 4 or 5 and where the mean score for inclusion for the specific outcome was greater than or equal to 4.00 (Table 5), (b) those outcomes for which at least 70 per cent of participants rated the outcome as 4 or 5 but where the mean score for inclusion in the data

set for the specific outcome was less than or equal to 4.00 (Table 6), and (c) those additional outcomes where the mean score for inclusion for the specific outcome was greater than or equal to 4.00, but for which less than 70 per cent of participants rated the outcome as 4 or 5 (Table 7).

Table 5. Additional Outcomes for Which Greater Than or Equal to 70 per cent of Participants Rated Outcome as 4 or 5 (Where 5 is Maximum Rating) and Where Mean for Specific Outcome is Greater Than or Equal to 4.00

Rank	*Outcome*	n	*Mean*	*SD*
1	Maternal cerebral infarction (death/damage of part of the brain caused by a sudden insufficiency of blood supply)	149	4.17	0.96
2	Deep vein thrombosis (blood clotting in the veins of the inner thigh or leg)	149	4.17	0.77
3	Overall obstetric intervention score (list of interventions with scores to indicate overall degree of invasive procedures)	149	4.17	0.97
4	Foetal and/or neonatal haemorrhage	151	4.17	0.89
5	Breastfeeding at 6 weeks	150	4.17	1.01
6	Episiotomy	150	4.17	0.81
7	Multiple pregnancy (pregnancy with 2 or more foetuses)	150	4.15	1.02
8	Infant respiratory distress syndrome (lung disease occurring most commonly in premature newborns)	151	4.15	0.91
9	Inverted uterus (a turning of the uterus inside out, whereby the fundus is forced through the cervix and protrudes into or outside the vagina)	151	4.15	0.96
10	Lead professional at birth	149	4.14	0.89
11	Transfer/referral to medical-led care during labour	151	4.13	0.81
12	Urinary incontinence (postnatal)	148	4.12	0.76
13	Cord prolapse (where the umbilical cord lies in front of or beside the presenting part in the presence of ruptured membranes)	149	4.12	0.91
14	Maternal preferences for future care	150	4.09	0.93
15	Foetal acidosis (a disturbance in the normal acid-base balance of the body in which the blood and body tissues are more acidic than normal)	151	4.09	0.88
16	Prolonged rupture of membranes (usually refers to rupture of membranes for more than 24 hr)	152	4.09	0.85
17	Transfer/referral to medical-led care during pregnancy	152	4.08	0.94
18	Need for blood transfusion	149	4.07	0.90
19	Retained products of conception (where all or part of the placenta or membranes is left behind in the uterus during the third stage of labour)	150	4.07	0.90
20	Rhesus isoimmunisation (incompatibility between the blood types of the baby and mother where maternal antibodies destroy the baby's red blood cells during pregnancy and after birth)	149	4.06	0.99

21	Manual removal of placenta	149	4.05	0.84
22	Amniotomy (artificial rupture of the foetal membranes as a means of inducing or expediting labour or assessing foetal well-being)	150	4.05	0.80
23	Intact perineum	151	4.05	0.90
24	Maternal attitudes toward routines and practices within model of care	150	4.04	0.94
25	Time from decision to birth by emergency caesarean section	151	4.03	0.85
26	Postmaturity (usually refers to any baby born after 42 wk gestation)	152	4.03	0.91
27	Breastfeeding at 6 mo	150	4.03	1.00
28	Malpresentation (abnormal position of the foetus)	151	4.02	0.86
29	Retained placenta	150	4.01	0.88
30	Length of infant hospital stay	150	4.01	0.84
31	Long-term infant/child neurodevelopmental outcome	149	4.00	0.87
32	Intrauterine hypoxia (lack of oxygen to the foetus)	151	4.00	0.97

Table 6. Additional Outcomes for Which Greater Than or Equal to 70 Per cent of Participants Rated Outcome a 4 or 5 (Where 5 is Maximum Rating) but Where Mean for Specific Outcome is Less Than or Equal to 4.00

Rank	***Outcome***	**n**	***Mean***	***SD***
1	Long-term infant/child neurodevelopmental outcome	149	4.00	0.87
2	Intrauterine hypoxia (lack of oxygen to the foetus)	151	4.00	0.97
3	Perineal and vaginal tears	149	3.99	0.92
4	Foetal distress (an ill-defined term when the health of the foetus is threatened for reasons such as an abnormal foetal heart rate patterns and/or a lack of oxygen)	150	3.97	0.93
5	Infant feeding problems	150	3.97	0.87
6	Method of foetal heart rate monitoring	151	3.95	0.93
7	Maternal perception of midwifery support during labour	150	3.94	0.99
8	Length of second stage of labour	151	3.88	0.96

Table 7. Additional Outcome Where Mean for Specific Outcome is Greater Than or Equal to 4.00 but for Which Less Than or Equal to 70 Per cent of Participants Rated Outcome a 4 or 5 (Where 5 is Maximum Rating)

Rank	***Outcome***	**n**	***Mean***	***SD***
1	Total cost per birth	149	4.00	0.89

Discussion

From the perspective of key stakeholders with self-identified expertise in the evaluation of different models of maternity care, this e-Delphi study identified a minimum data set consisting of 48 outcomes. Consideration should be given to routinely reporting and using these outcomes in evaluations of models of maternity care. The core set is not intended to be prescriptive or restrictive, and additional outcomes will be necessary to meet the needs of different settings.

Our study has led to the development of an original core set of outcomes for the care of women and their babies during pregnancy and childbirth, but the concept of core sets of outcomes is not novel in health care more generally. Examples include an initiative by the World Health Organization in the late 1970s relating to cancer trials. Two meetings held on the Standardization of Reporting Results of Cancer Treatment in Turin (1977) and Brussels (1979) brought together more than 30 representatives from many of the major cooperative groups performing randomised trials of cancer treatments. These discussions led to a World Health Organization handbook of guidelines on the minimum requirements for data collection in cancer trials (6). These requirements included baseline variables such as sex, age, height, weight, and performance status and data about the cancer, such as tumour size, histopathology, and stage of disease (7). In addition to this core set, investigators in cancer trials are able to collect other information variables that they deem necessary.

Use of the Delphi process in this study incurs similar benefits and limitations to those considered in the literature about this technique. The Delphi process is able to elicit and condense the opinions of many toward consensus, allows anonymity of response among participants, and is relatively low cost (8,9). The main limitation of the method is the ambiguities that can exist around the issue of defining consensus (10,11) and expertise. Although the selection of participants is crucial to the credibility of any Delphi process, it remains one of the most challenging areas each time that the Delphi method is used. Commonly, participants in a Delphi survey have been referred to as “experts” in the area under investigation (12). However, the subjectivity of the concept of “expertise” and difficulties in defining the level of expertise required contribute to criticisms of the rigor of the method (10) and have led to suggestions about the inappropriateness of the term “expert”(8,9). In an attempt to address the limitations inherent within the method, we used the concept of self-identified “expertise in the evaluation of maternity care.” In addition, we chose to elicit consensus through a combination of mean values, self-identified level of expertise, and percentage of agreement (i.e., >70%) rather than any of these factors alone. We recognize that others may choose to elicit consensus by different methods.

Of those consenting to participate, only 47 per cent completed all 3 rounds. Although respectable in the arena of Delphi surveys, the potential for nonresponse bias is present. We are unable to determine the effect of such bias since we did not have the means to identify non-responders other than by e-mail address; thus, it is not known if, or to what extent, they differed systematically from the study population.

We also had to choose between a single panel consisting of a mix of key stakeholders and multiple panels of key stakeholder groups. In line with conventional Delphi techniques (12) and in the context of the multidimensional nature of maternity care, we opted for a single panel consisting of a mix of key stakeholders. It is possible that the use of multiple panels would result in different sets of outcomes that might be thought to be more relevant within each key stakeholder group and that the single panel, with a predominantly midwifery representation, has masked the priorities of groups that had relatively lower representation within the panel, for example, women and obstetricians. In addition, most respondents to all rounds were from European countries. Therefore, this data set is, by virtue of country of origin and primary interest in maternity care of most respondents, inclined toward European countries and midwifery practitioners. Further research is required to increase the representativeness of a single heterogeneous panel or to conduct separate consensus development exercises with homogeneous panels of key stakeholder groups. Until then, this data set may be of international value to multi-professional practitioners.

Most items in this data set are adverse outcomes; that is, they would provide information on something that one would wish to avoid. This focus on adverse outcomes coincides with calls for a reorientation toward outcomes of "optimal" events (13) and a growing emphasis on the normalcy of childbirth within midwifery (14). The emphasis on adverse outcomes in the final data set might be a reflection of their dominance in the round 1 instrument, which was derived from the systematic review of randomised trials comparing models of maternity care, or it could reflect the importance of these events in people's lives and experiences. Nevertheless, participants in the Delphi process, most of whom were midwives, had the opportunity to include additional outcomes to the round 1 instrument; these were subsequently included if they were suggested by 2 or more participants. It would be daunting for trialists to try to power their trial to detect clinically and statistically significant differences for all the outcomes in the core data set. For some outcomes, especially those for which the differences between treatments are likely to be small, it may be impossible to do a sufficiently large trial. However, we suggest that trialists might use the core list of outcomes to identify the primary outcome on which to calculate the sample size for their trial. If they wish, they could then mention the effect sizes that they might be able to detect reliably for some of the other core outcomes, if they achieve this sample size.

In addition, this core data set would be useful within multi-centre trials and for comparisons between trials. It might also be a significant step in facilitating useful meta-analyses of similar studies.

Conclusions

Comparisons between models of maternity care and, in particular, comparisons of the relative effectiveness of interventions in randomised trials are difficult because of inconsistencies in the choice of outcome measures routinely collected and reported. It is hoped that use of the data set

will increase the potential for national and international comparisons of models for maternity care and that any subsequent work will assist in the ongoing development of this data set. In this way, we hope that the core set will make it easier to assess the care of women and their babies during pregnancy and childbirth.

Acknowledgments

We gratefully acknowledge the participation and cooperation of all participants in the pilot and main e-Delphi study. Without their commitment and enthusiasm, this study would not have been possible. Details of participants who contributed to the development of the dataset and who consented for their details to be shared can be found at http://www.nursing-midwifery.tcd.ie/research/maternity_women/res_mat_wmn_index.php.

References

1. Devane D, Begley CM. Childbirth: How safe is safe enough? *Br J Midwifery* 2004;**12**(7):416–417.
2. Ghaffari Z, Bener A, Ahmed B. Safety of vaginal birth after caesarean delivery. *Int J Gynaecol Obstet* 2006;**92**(1):38–42.
3. Gottvall K, Grunewald C, Waldenström U. Safety of birth centre care: Perinatal mortality over a 10-year period. *BJOG*2004;**111**(1):71–78.
4. Buitendijk S, Zeitlin J, Cuttini M, *et al*. Indicators of foetal and infant health outcomes. *Eur J Obstet Gynecol Reprod Biol*2003;**111**(Suppl. 1):S66–S77.
5. SPSS *for Windows, Rel. 14.0*. Chicago, Illinois: SPSS Inc., 2005.
6. World Health Organization. *WHO Handbook for Reporting Results of Cancer Treatment*. WHO Offset publication No. 48. Geneva, Switzerland: Author, 1979.
7. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;**47**(1):207–214.
8. McKenna HP. The Delphi technique: A worthwhile research approach for nursing? *J Adv Nurs* 1994;**19**(6):1221–1225.
9. Goodman CM. The Delphi technique: A critique. *J Adv Nurs* 1987;**12**(6):729–734.
10. Sackman H. *Delphi Critique: Expert Opinions, Forecasting, and Group Process*. Lexington, Massachusetts: Lexington Books,1975.
11. Beretta R. A critical review of the Delphi technique. *Nurse Res* 1996;**3**(4):79–89.