TITLE
Adding value to core outcome set development using multi-method systematic reviews

AUTHORS
Ginny Brunton* [1,2], James Webbe [3], Sandy Oliver [2,4], Chris Gale [3]

* Corresponding Author
Email: ginny.brunton@ontariotechu.ca

AFFILIATIONS
[1] Faculty of Health Sciences, OntarioTech University, Oshawa, Canada
[2] Evidence for Policy and Practice Information and Coordinating (EPPI-) Centre, UCL Institute of Education, University College London, WC1H 0NR, United Kingdom
[3] Neonatal Medicine, School of Public Health, Faculty of Medicine, Imperial College London, Chelsea and Westminster Hospital campus, London, SW10 9NH, United Kingdom
[4] Africa Centre for Evidence, Faculty of the Humanities, University of Johannesburg, Auckland Park, Johannesburg

Contact information
Ginny Brunton
Faculty of Health Sciences, OntarioTech University
2000 Simcoe Street North, Oshawa, Ontario
Canada L1G 0C5
(phone) 905-721-8668 ext. 2434

INTRODUCTION

Evidence-based care is the cornerstone of high-quality health service provision; and clinical trials of specific interventions are essential in establishing a robust evidence base. In many fields multiple trials are conducted that evaluate the same intervention, examining its effect amongst different populations and in varied settings. Perhaps unsurprisingly, these often report conflicting findings. To make sense of these findings and evaluate the overall effectiveness of the intervention, trials are combined using systematic review and meta-analytic methods[1,2]. Such systematic reviews aim to pull together the findings from multiple trials of the same intervention, to provide a more comprehensive and robust answer to the question of effectiveness. Overall effectiveness is accomplished by combining the effect estimates of trial outcomes. Their aggregated sample size provides additional power that can establish whether an intervention is statistically and clinically or socially significant[3,4]. The usefulness of systematic reviews and meta-analyses is commonly limited however, because trials examining the same intervention often report different outcomes, or where the same outcomes are reported they are measured in different ways. This has been noted across multiple areas of health care research[5-8]. Conflicting trial outcomes therefore limit the evidence base and create confusion

for practitioners, decision-makers and the public when deciding on the most effective treatment for a condition[4].

The development of a set of agreed or 'core' outcomes, a Core Outcome Set (COS), to be recorded in all trials in a clinical field, has been proposed to strengthen research and improve evidence-informed decision-making[9]. Core outcome set development involves a sequence of processes: initially the scope, context and setting of trials are determined, followed by identification of a comprehensive inventory of relevant outcome domains[10], and finally this inventory is used to reach a consensus among multiple stakeholders about which outcome domains should form a COS.

Identification of a comprehensive inventory of outcomes that have been reported in research is increasingly achieved by systematically reviewing potentially relevant trials in the clinical field of interest[11]. However, such reviews may simply identify outcomes important to researchers that are relatively easy to undertake. In addition, syntheses of quantitative trials may choose easily measurable rather than clinically relevant outcomes[12], focusing on outcomes that may not speak to the experiences of public stakeholders[13].

Input from public stakeholders, including caregivers, parents and patients is key to determining relevant outcome domains[13]; public stakeholder perspectives are sought through a variety of methods, including focus groups, consensus conferences, Delphi or Nominal group techniques[11,14,15]. These are often conducted in stakeholder groups that are transdisciplinary in nature, comprised of members of the public, doctors and other health care professionals, and academic researchers[11]. Consultations are then initiated to prioritize the domains most important to stakeholders (the Core Outcome Set) before deciding on which outcome measures to use.

However, power imbalances may limit the extent to which public stakeholders are able to contribute where 'experts' are also part of the group[16,17]. Although some COS teams do not actively engage public stakeholders in the identification of outcomes for the inventory, many will try to identify outcomes from parent/patient/public for the inventory. However, this tends to be done using standalone, original qualitative research that can be expensive and time consuming, often only sampling a small and very specific population[18-20]. Research that explores the perspectives, opinions or experiences of the public should inform the range of outcome domains identified for core outcome set development[15]. The aim of the study was to examine the potential added value of conducting a synthesis of multiple qualitative studies alongside a quantitative review of outcomes. These were conducted to inform consensus group meetings where core outcome domains were identified as part of core outcome set development in neonatal care. Trials in neonatology report a wide range of outcome measures, making synthesis difficult, at risk of bias, or even impossible[21,22]. We hypothesised that conducting a review of qualitative research in addition to a review of trials would identify more outcome domains relevant to public as well as to professional stakeholders.

METHODS

Design

Two separate systematic reviews were conducted[23,24], using standard methods which included searching multiple sources, inclusion screening, data extraction and synthesis[2]. One systematic review examined outcomes reported in neonatal trials ('quantitative review'); the other considered qualitative research on stakeholders' perspectives of important outcomes during neonatal care ('qualitative review'). In order to examine what the quantitative and qualitative reviews of neonatal outcomes could contribute to the consensus process, a convergent mixed methods synthesis was conducted[25]. The methods and findings from each review were compared and contrasted, utilising joint displays that array the findings from each review in order to determine the extent that quantitative review results were confirmed by qualitative review results and vice versa[25]. This process is illustrated in **Figure 1**.

[ADD Figure 1 HERE. Convergent Mixed Methods Synthesis Design (from Creswell 2015 p.37)]

Once the methods and findings of each review were merged, two researchers (GB&JW) discussed and interpreted commonalities and differences between reviews. Interpretations were sense-checked by a third researcher (CG).

RESULTS

The quantitative review identified 24,214 citations, of which 76 trials of interventions were included. The qualitative review identified 1,130 citations and included 62 studies that accessed the views of parents, carers or patients. **Figure 2** and **Figure 3** shows the flow of studies through each review[23,24].

[ADD FIGURE 2 HERE – QUANT REVIEW FLOW OF STUDIES]

[ADD FIGURE 3 HERE – QUAL REVIEW FLOW OF STUDIES]

Comparison of methods

Searching

For the quantitative review, we sought references from MEDLINE, CINAHL, EMBASE, and CENTRAL, using a combination of free-text and thesaurus terms. To identify studies of stakeholders' perspectives for the qualitative review, we searched MEDLINE, CINAHL, EMBASE, PSYCINFO and ASSIA, citation checking of included studies, and Google Scholar. Again, a combination of free-text and thesaurus terms were used. Further details of the search strings are provided In Appendix 1.

Eligibility screening
To be included in the quantitative review, studies had to:

- be a randomised controlled trial (RCT) or cluster RCT;
- involve babies who required neonatal care in a high-income country;
- include over 100 babies in each arm of the trial (to identify the largest trials with potentially more than one outcome of interest);
- have been published in the past five years (to identify the most current outcomes of interest); and
- be available in English or provide a comprehensible English translation.

Studies included in the qualitative review had to:

- be a peer review journal article or funders report;
- be a primary study seeking parent, clinicians or ex-patient views of care;
- be concerned with babies who required neonatal care in a neonatal unit;
- contain qualitative data of neonatal outcomes from the perspectives of parents, clinicians or ex-patients;
- be published within past 20 years; and
- be available in English or provide a comprehensible English translation

Data extraction and quality assessment
For the quantitative review, to extract data from trials a previously identified framework of neonatal outcomes ordered by biological health systems was utilised[26]. As per best practice, risk of bias assessment was conducted using a previously validated tool[27]. For the qualitative review, as we were seeking to simply identify outcome domains rather than consider critically the depth of the findings, quality assessment of stakeholder views studies was not undertaken. Further details of data extraction for both reviews are provided elsewhere[23,24].

Analysis
Using the previously identified framework of neonatal outcome domains[26], the number and characteristics of different outcomes were categorised by biological health systems within the quantitative review. This framework of outcome domains then became the initial conceptual framework for the qualitative review. Where studies referred to outcomes that were related to neonatal care, textual data were analysed by two researchers to derive common themes and new outcome domains developed, using framework synthesis methods[17,28]. The research team approached the study design and analysis using a scientific realist critical perspective[29].

Quality assurance
For the quantitative review, screening and coding were tested by three researchers, to both consolidate criteria definitions and to establish a high degree of agreement on studies (90% or over). During the qualitative review, after double-screening a sample of papers and agreeing criteria all screening was completed by one researcher. For quality assurance, a second research screened a random 10% sample of abstracts and titles. Agreement between reviewers was assessed by Cohen's kappa coefficient[30].

For both reviews, all full text references were screened, coded and (where used) assessed for quality by two researchers, with disagreements on inclusion resolved by discussion and a third researcher where needed. Reference management, screening and coding was managed using EPPI-Reviewer research synthesis software[31]. Codes and relevant quotes were analysed using Microsoft Excel.

In summary, the methods used in each systematic review were very similar. For example, the search sources used were appropriate for the types of literature being sought for each review. Qualitative research is more often indexed on social sciences databases and more likely to be located through searching non-commercial electronic sources[32,33]. Inclusion criteria were also similar in both reviews, with specific study types (i.e. trials or research on stakeholders' views) differentiating each review. Data were collected using the same coding framework, and the use of framework synthesis in the qualitative review allowed the framework to be adapted as new thematic categories were identified from the qualitative research.

Reviews differed in their approach to synthesis. In both reviews, studies were coded into the same analytical framework. However, those studies with data that did not 'fit' the framework were coded as 'other' outcome domains and listed textually by either the name of the outcome measure used (quantitative review), or the text indicating that this was considered an effect or outcome of care by stakeholders (qualitative review). The textual data were then thematically synthesised to infer outcomes, which became the new outcome domains added to the pre-existing framework.

Comparison of identified outcomes

Conducting complementary syntheses of quantitative intervention trials and qualitative studies of stakeholder's views allowed us to identify more outcome domains than would have been revealed by either synthesis alone. Across the two reviews (N=138 studies), a total of 26 outcome domains were identified. **Figure 4** illustrates the range and proportions of domains identified across the studies identified across both quantitative and qualitative reviews.

[ADD FIGURE 4 HERE]

The most frequently identified domains across both reviews were focused on gastrointestinal (62%, 85/138 studies), survival (59%, 81/138 studies), or respiratory outcomes (51%, 71/138 studies). In each case, these domains were populated by a much higher proportion of outcomes from quantitative studies (i.e. trials), as shown in **Figure 5**.

[ADD FIGURE 5 HERE]

The quantitative review identified outcomes classified into a total of 17 different health systems outcome domains, compared to 24 outcome domains identified in the qualitative review. Most

studies in the quantitative review focused on domains related to biological health systems (e.g. respiratory outcomes, gastrointestinal outcomes, or infant survival). In contrast, studies included in the qualitative review reported outcome domains that were focused on wider psychological and social issues such as parent support, interactions with healthcare professionals, infant normality and infant suffering. Developmental outcomes were identified similarly in both quantitative and qualitative reviews (34 studies in the quantitative review versus 32 studies in the qualitative review).

Comparison of the two sets of outcome domains revealed that over half (58%, 15/26 outcome domains) were common to both quantitative and qualitative reviews. Outcome domains were also identified which were unique to each type of research: some domains were identified in qualitative studies that were not found in trials, and vice versa. These are shown in **Figure 6**.

[ADD FIGURE 6 HERE]

The qualitative review revealed nine outcome domains not identified in the quantitative reviews. These focused most often on themes of parent support (n=30 studies), healthcare worker communication (n=30 studies), healthcare worker knowledge and competence (n=23 studies), infant normality (n=22 studies) and infant suffering (n=15 studies). However, the quantitative reviews also produced two unique outcome domains: health care costs (n=7) and biochemical variables (n=11).

Even when outcome domains were common to both quantitative and qualitative reviews, the nature of information from qualitative and quantitative studies differed. One example of this is found in trials utilising the gastrointestinal health system domain: while quantitative studies (n=61) tended to use a standardised outcome tool such as Bell staging for necrotising enterocolitis to measure a specific outcome, the qualitative studies (n=24) focused on the duration and meaning of feeding with such quotes as "*I fully breastfed for four months, 100%, and I am so proud of that*" [34](p.234) [34].

In another example, while quantitative studies employed disparate outcomes such as weight, head circumference or developmental measures, it was evident from qualitative studies that study participants were concerned with growth in a more holistic manner, where weight may or may not be a crucial factor:

"*…nurses are aware that increased noise levels in a NICU have detrimental physiological effects for the newborn," [which authors suggested lead to] "problems with appropriate growth and development.* (p.10)[35]

"The neonatologist came in and said that it was to their benefit to grow quicker, they could tolerate it more, their immune system—that's what is best for them. That's all she needed to say, it was decided by sundown." (p.695)[36]

"But I don't look at her as a premature. I don't know why I don't look at her as a premature baby. I just look at her as a little baby… The doctor says she's 3 lb underweight. So that's like 20% body weight. But to me I look at her and I don't see it all. I see this little happy little thing!" (p.166)[37]

A last example illustrates how outcome domains were very similar but appeared to approach the outcome from different underlying principles. For example, while several trials assessed physiological markers such as blood pressure, heart rate and desaturation events to indicate deterioration or improvement in an infant's condition[38-41], nurses in one qualitative study identified that infants' physiological stability was also dependent on low noise levels in the NICU[35], suggesting more of a focus on wellness. These findings highlight that outcomes may also be influenced by factors beyond the treatment under study, and outcome domains should include assessment of wider factors beyond the infant's medical condition.

In summary, systematic reviews of both quantitative and qualitative studies identified 26 outcome domains to be considered by the COS consensus group. Of these, nine were unique to qualitative reviews and, after consensus meetings and the Delphi process exercise, were incorporated into the 'Quality of Life' outcome domain, forming part of the final core outcome set[42].

DISCUSSION

Strengths

These findings demonstrate that a review of qualitative research on key stakeholders' views about important outcomes complemented those outcome domains identified though a quantitative review of trials. A qualitative review identified additional unique outcome domains for consideration by core outcome set development consensus groups. They added depth of understanding to potential outcome domains that were identified across trials that will assist subsequent efforts to identify suitable outcome measures. This is an innovation in the application of research synthesis methods. It addresses the call for core outcome set development methods that make findings more granular and wider in scope to encompass health-related quality of life issues, whilst ensuring that the voices of patients, parents and healthcare professionals' are brought into the consensus development process[43]. The additional diversity and depth of outcome domains identified across both reviews appeared to foster discussions between core outcome consensus group members and in subsequent Delphi consultation exercises. Some outcome domains identified through qualitative review, that would not have been otherwise known, were subsumed into a final core outcome domain.

This is an early example of mixed method research synthesis employed to inform the core outcome domain stage of core outcome set development. Just as outcome domains from trials can be identified using systematic review methods, outcome domains from qualitative research evidence can also be sought, systematically reviewed, and synthesised[44]. Conducting reviews of quantitative and qualitative literature has provided a wider and more nuanced evidence base

of potentially important outcomes in neonatal care than would have been identified by a quantitative review alone. The findings from this comparison of review methods and findings suggest that this type of mixed method synthesis enriched subsequent discussions with key stakeholders in COS development, although more evaluation of this method should be explored across COS development efforts. Such qualitative research evidence of stakeholder perspectives has much to offer the consensus process but has not yet been widely utilised[14,19,45]. This is an efficient way to identify a large number of potential useful outcome domains for consensus group discussion[20].

Outcome domains identified through a qualitative review of research may reflect outcomes more visible to members of the public and health professionals taking part in COS consensus development. Derived from primary research with patients, parents and healthcare professionals, these identified outcome domains, in turn, may stimulate more discussion amongst transdisciplinary COS consensus groups. Consensus group members, and particularly members of the public, may feel too intimidated to provide their opinion when faced with multiple outcomes from a range of body systems, of which they may or may not have experience[17,46]. Outcomes related to the respiratory system, for example, may not be meaningful to public stakeholders who do not have direct experience of them, resulting in less discussion amongst consensus group members. However, outcome domains derived from research on parent's, patients' and professionals' views identified universal issues more widely experienced by public stakeholders and understood by professional stakeholders. For example, research citing parents' concerns about their child developing 'normally' and 'fitting into' school were commonly experienced and understood by both public and professional stakeholders, which led to more discussion between consensus group members. These particular findings have additional relevance to core outcome set development, given that preterm birth has a considerable impact over the lifespan and outcome measures should reflect these longer-term outcomes[47-49]. The expansion of conversations between consensus group members constitute more shared knowledge of all stakeholder perspectives, which may influence the groups' choices in outcome domain selection for subsequent Delphi consensus processes[50].

Dataset limitations

The qualitative review synthesised outcome domains from studies that did not originally aim to prioritise important outcomes, an issue that has been noted elsewhere in COS development[19]. Only three of the 58 included qualitative studies asked participants specifically about priorities[51-53]. These focused on neonatal nurses' priorities in terms of training and providing care. The remaining studies asked participants more generally about their neonatal care experiences, and textual data suggested important outcome domains.

It can be argued that the identified outcomes were inferred as important, simply because participants in the included qualitative studies identified them. However, without directly asking people, this remains a proxy for understanding more directly what is important to people

(parents and healthcare professionals) involved in neonatal care. This is not unusual in research synthesis, as the aims of primary studies included in reviews do not always match exactly with the aims of the reviews themselves[54]. This results in findings that are tangentially relevant to the review, but of use because they discuss some aspect of the review topic that answers the review question. It is noted that many outcome domains were identified across multiple qualitative studies. These comprised several thousand participants from varied settings and across different countries. They are also similar to priorities identified in preterm birth research priority-setting exercises[55-57]. This commonality and magnitude suggest that these outcome domains are widely considered important.

Methodological limitations

As a predominantly aggregative and deductive exercise, categorising qualitative data can be challenging, due to the different meanings and wordings adopted across studies. This scoping review of the insights gained from narrative text from patients, parents and health professionals required additional time for thoughtful sifting and grouping into outcome domains[58,45]. Outcome domain development is further challenged by the necessary removal of broad concepts and consolidation of similar concepts prior to the derivation of a final outcome domain list for stakeholder consideration[58]. Here, researcher discussion and agreement were crucial to ensure that the meanings of concepts that were subsumed into broader outcome domains were retained. The qualitative concepts underpinning many of these outcome domains may assist the future selection of appropriate outcome measures.

The use of framework synthesis methods in the qualitative review further addressed some of the challenges in the meaning and ordering of qualitative concepts to outcome domains. This method offers a very flexible approach to research synthesis. Framework synthesis is suitable for mixed methods research synthesis that deals with complex concepts[59], because it can fractionate and manage heterogenous data to generate, explore and/or test theory. The extent of the development of theory depends on the contextual, theoretical and process characteristics and on review team challenges (Brunton et al. in press). Examples of framework synthesis utilised to scope or map research for policy and research development are increasingly evident[60-62]. It is recommended in similar standard-setting exercises, including both clinical and policy guideline development[59,63]. Used in the context of core outcome set development, framework synthesis was employed in the qualitative review to combine and deductively summarise individual study findings[63]. Our use of framework synthesis in this instance is aligned more with meta-aggregation methods that seek to summarise concepts in an integrative manner, rather than to generate or explore theory[64-67].

One limitation relates to the inclusion criteria applied to the quantitative review. Although this is the largest review to map outcomes in neonatal randomised controlled trials, it only contains a subset of all trials published in this field. The initial searches identified over 900 neonatal randomised trials over the five-year period, which was not feasible to review. While including more trials may have led to the identification of additional outcomes, the review was limited

to the largest trials, in order to focus the review on outcomes that were current, most prominently researched and thus most relevant to clinical practice.

Including more trials in our analysis may not have added depth of information. We also found differences in the outcomes reported between the two reviews at both outcome domain and individual outcome level. Even outcomes common to both reviews reported differences in how the outcomes were discussed. This suggests that while a larger quantitative review may have identified more outcomes, it is unlikely to have identified the novel outcomes seen in the qualitative synthesis.

Qualitative data is not uniformly available across all healthcare topics. Where suitable data exists, the use of qualitative evidence synthesis improves the efficiency of COS development. In the field of neonatology there have been a considerable number of qualitative studies undertaken, and we demonstrate their use here. This may be a less useful method in other areas of core outcome set development where there is less qualitative research available. However it is not always easily located; searching in social science databases in addition to medical sources is recommended[32,33]. Future research could also consider the benefits of using this method to derive public stakeholders' views in comparison to focus groups and interviews which are more commonly used in core outcome set development[11,19].

Comparing the methods across two systematic reviews also raised questions about the use of risk of bias and quality assessment procedures in systematic reviews used to inform core outcome domain identification. Our findings suggest that, if the purpose of such reviews is to 'list' potential outcome domains for discussion rather than to synthesise effect sizes, they should be more usefully considered scoping reviews which aim to 'map' the research[65,68]. As a starting point for discussion of important outcomes in consensus group meetings, these outcome domains were then further critically considered for their credibility, transferability, dependability and confirmability[69].

CONCLUSIONS

Findings from comparing quantitative and qualitative reviews suggest that qualitative synthesis is a useful way of adding value in identifying outcomes identified by parents, patients and professionals, which were subsequently prioritized in COS development. Qualitative synthesis identified a broad range of outcome domains and more depth to outcomes, complementing the standard quantitative review undertaken. This represents a novel application of research synthesis methods. Outcomes identified through the qualitative review were independently prioritized by the COS development group and emerged in the final COS. Future research could compare more directly the benefits of using this method to derive public stakeholders' views in comparison to primary qualitative research currently in use. Such reviews conducted for COS development should also more usefully be considered scoping reviews, which could improve the efficiency of COS development.

HIGHLIGHTS

- Core outcome set (COS) development is increasingly used to agree sets of outcomes to be used across trials, to allow knowledge to be combined across trials in future meta-analyses.
- Outcome domains are identified through a systematic review of trials and through qualitative research with key stakeholders. This latter exercise is often time consuming with unclear impact.
- Instead, conducting a qualitative systematic review of research on patients', parents' and professional caregivers' experiences alongside a quantitative systematic review of trial outcomes to inform COS development in neonatal care identified a wider range and greater depth of health and social outcome domains.
- These were incorporated into the subsequent Delphi process and informed the final set of core outcome domains.
- This qualitative scoping of participant perspectives research, used with systematic review of trials, could identify more outcome domains for consideration and provide greater depth of understanding to inform stakeholder group discussions in COS development.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. Egger M, Davey-Smith G, Altman D. *Systematic reviews in health care: meta-analysis in context.* John Wiley & Sons; 2008.
2. Gough D, Oliver S, Thomas J. *An introduction to systematic reviews.* Sage; 2017.
3. Mulrow CD. Systematic reviews: rationale for systematic reviews. *Bmj.* 1994;309(6954):597-599.
4. Petticrew M, Roberts H. Systematic reviews in the social sciences: a practical guide. 2006. *Malden USA: Blackwell Publishing CrossRef Google Scholar.*
5. Duffy JM, Hirsch M, Kawsar A, et al. Outcome reporting across randomised controlled trials evaluating therapeutic interventions for pre-eclampsia. *BJOG: An International Journal of Obstetrics & Gynaecology.* 2017;124(12):1829-1839.
6. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. *Trials.* 2017;18(1):122.
7. Miyar J, Adams CE. Content and quality of 10 000 controlled trials in schizophrenia over 60 years. *Schizophrenia bulletin.* 2012;39(1):226-229.
8. Willhelm C, Girisch W, Gottschling S, Gräber S, Wahl H, Meyer S. Systematic Cochrane reviews in neonatology: a critical appraisal. *Pediatrics & Neonatology.* 2013;54(4):261-266.
9. Williamson PR, Altman DG, Blazeby JM, Clarke M, Gargon E. The COMET (core outcome measures in effectiveness trials) initiative. *Trials.* 2011;12(Suppl 1):A70.
10. Williamson PR, Altman DG, Bagley H, et al. The COMET handbook: version 1.0. *Trials.* 2017;18(3):280.
11. Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and user survey. *PLoS One.* 2016;11(1):e0146444.
12. Bursztajn HJ. Clinical trials and effectiveness research. *American Journal of Psychiatry.* 2000;157(1):152-a-152.
13. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. In: SAGE Publications Sage UK: London, England; 2012.
14. Gargon E, Gorst SL, Harman NL, Smith V, Matvienko-Sikar K, Williamson PR. Choosing important health outcomes for comparative effectiveness research: 4th annual update to a systematic review of core outcome sets for research. *PloS one.* 2018;13(12):e0209869.
15. Sinha IP, Smyth RL, Williamson PR. Using the Delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies. *PLoS medicine.* 2011;8(1):e1000393.
16. Abelson J, Forest P-G, Eyles J, Smith P, Martin E, Gauvin F-P. Deliberations about deliberative methods: issues in the design and evaluation of public participation processes. *Social science & medicine.* 2003;57(2):239-251.
17. Oliver SR, Rees RW, Clarke-Jones L, et al. A multidimensional conceptual framework for analysing public involvement in health services research. *Health Expectations.* 2008;11(1):72-84.
18. Brett J, Staniszewska S, Mockford C, et al. A systematic review of the impact of patient and public involvement on service users, researchers and communities. *The Patient-Patient-Centered Outcomes Research.* 2014;7(4):387-395.
19. Jones JE, Jones LL, Keeley TJ, Calvert MJ, Mathers J. A review of patient and carer participation and the use of qualitative research in the development of core outcome sets. *PloS one.* 2017;12(3):e0172937.

20. Keeley T, Williamson P, Callery P, et al. The use of qualitative methods to inform Delphi surveys in core outcome set development. *Trials.* 2016;17(1):230.
21. Clarke M. Standardising outcomes in paediatric clinical trials. *PLoS Medicine.* 2008;5(4):e102.
22. Sinclair JC, Haughton DE, Bracken MB, Horbar JD, Soll RF. Cochrane neonatal systematic reviews: a survey of the evidence for neonatal therapies. *Clinics in perinatology.* 2003;30(2):285-304.
23. Webbe J, Brunton G, Ali S, Duffy JM, Modi N, Gale C. Developing, implementing and disseminating a core outcome set for neonatal medicine. *BMJ paediatrics open.* 2017;1(1).
24. Webbe J, Brunton G, Ali S, Longford N, Modi N, Gale C. Parent, patient and clinician perceptions of outcomes during and following neonatal care: a systematic review of qualitative research. *BMJ paediatrics open.* 2018;2(1).
25. Creswell JW. *A concise introduction to mixed methods research.* Sage Publications; 2015.
26. NHS UK. NHS Neonatal Data Set Overview. National Health Service (NHS) UK. http://www.datadictionary.nhs.uk/data_dictionary/messages/clinical_data_sets/data_sets/national_neonatal_data_set/national_neonatal_data_set_-_episodic_and_daily_care_fr.asp?shownav=1. Published 2016. Accessed 17 January 2019, 2019.
27. Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Controlled clinical trials.* 1996;17(1):1-12.
28. Brunton G, Oliver S, Thomas J. Innovations in framework synthesis as a systematic review method. *Research Synthesis Methods.* in press.
29. Chakravartty A. Scientific realism. . Stanford University. Stanford encyclopedia of philosophy [Online] Web site. http://plato.stanford.edu/entries/scientific-realism/. Published 2011. Updated 2011 Apr 27. Accessed 17 January 2019, 2019.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *biometrics.* 1977:159-174.
31. Thomas J, Brunton J, Graziosi S. EPPI-Reviewer 4.0: software for research synthesis. EPPI-Centre Software 2010. London: Social Science Research Unit. *Institute of Education, University of London.* 2010.
32. Stansfield C, Brunton G, Rees R. Search wide, dig deep: literature searching for qualitative research. An analysis of the publication formats and information sources used for four systematic reviews in public health. *Research synthesis methods.* 2014;5(2):142-151.
33. Stansfield C, Kavanagh J, Rees R, Gomersall A, Thomas J. The selection of search sources influences the findings of a systematic review of people's views: a case study in public health. *BMC medical research methodology.* 2012;12(1):55.
34. Brødsgaard A, Zimmermann R, Petersen M. A preterm lifeline: Early discharge programme based on family-centred care. *Journal for Specialists in Pediatric Nursing.* 2015;20(4):232-243.
35. Darcy AE, Hancock LE, Ware EJ. A descriptive study of noise in the neonatal intensive care unit: ambient levels and perceptions of contributing factors. *Advances in Neonatal Care.* 2008;8(5):S16-S26.
36. Miracle DJ, Meier PP, Bennett PA. Mothers' Decisions to Change From Formula to Mothers' Milk for Very-Low-Birth-Weight Infants. *Journal of Obstetric, Gynecologic & Neonatal Nursing.* 2004;33(6):692-703.

37. Holditch-Davis D, Bartlett TR, Blickman AL, Miles MS. Posttraumatic stress symptoms in mothers of premature infants. *Journal of Obstetric, Gynecologic & Neonatal Nursing.* 2003;32(2):161-171.
38. Baud O, Maury L, Lebail F, et al. Effect of early low-dose hydrocortisone on survival without bronchopulmonary dysplasia in extremely preterm infants (PREMILOC): a double-blind, placebo-controlled, multicentre, randomised trial. *The Lancet.* 2016;387(10030):1827-1836.
39. Kribs A, Roll C, Göpel W, et al. Nonintubated surfactant application vs conventional therapy in extremely preterm infants: a randomized clinical trial. *JAMA pediatrics.* 2015;169(8):723-730.
40. Kamlin COF, O'Connell LA, Morley CJ, et al. A randomized trial of stylets for intubating newborn infants. *Pediatrics.* 2013:peds. 2012-0802.
41. Loewy J, Stewart K, Dassler A-M, Telsey A, Homel P. The effects of music therapy on vital signs, feeding, and sleep in premature infants. *Pediatrics.* 2013:peds. 2012-1367.
42. Webbe J, Duffy JM, Afonso E, et al. Core outcomes in neonatology: Development of a core outcome set for neonatal research. . *JAMA Paediatrics.* 2019 submitted.
43. Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012;13(1):132.
44. Thomas J, O'Mara-Eves A, Harden A, Newman M. Synthesis methods for combining and configuring textual or mixed methods data. In: Gough D, Oliver S, Thomas J, eds. *An introduction to systematic reviews.* 2 ed. London: Sage; 2017:211-250.
45. Gorst SL, Young B, Williamson PR, Wilding JP, Harman NL. Incorporating patients' perspectives into the initial stages of core outcome set development: a rapid review of qualitative studies of type 2 diabetes. *BMJ Open Diabetes Research and Care.* 2019;7(1):e000615.
46. Martin GP. Public deliberation in action: emotion, inclusion and exclusion in participatory decision making. *Critical Social Policy.* 2012;32(2):163-183.
47. Hall EO. A double concern: grandmothers' experiences when a small grandchild is critically ill. *Journal of Pediatric Nursing.* 2004;19(1):61-69.
48. Muller-Nix C, Forcada-Guex M, Pierrehumbert B, Jaunin L, Borghini A, Ansermet F. Prematurity, maternal stress and mother–child interactions. *Early human development.* 2004;79(2):145-158.
49. Saigal S. Functional outcomes of very premature infants into adulthood. Paper presented at: Seminars in Fetal and Neonatal Medicine2014.
50. Macefield R, Blencowe N, Brookes S, et al. Core outcome set development: the effect of Delphi panel composition and feedback on prioritisation of outcomes. *Trials.* 2013;14(S1):P77.
51. Ahern K. What neonatal intensive care nurses need to know about neonatal palliative care. *Advances in Neonatal Care.* 2013;13(2):108-114.
52. Turrill S. A focus of care for neonatal nursing: The relationship between neonatal nursing practice and outcomes. Part 1. *Paediatric nursing.* 2003;15(4):13.
53. Wielenga JM, Tume LN, Latour JM, van den Hoogen A. European neonatal intensive care nursing research priorities: an e-Delphi study. *Archives of Disease in Childhood-Fetal and Neonatal Edition.* 2015;100(1):F66-F71.
54. Gough D. Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research papers in education.* 2007;22(2):213-228.
55. Duley L, Uhm S, Oliver S. Top 15 UK research priorities for preterm birth. *The Lancet.* 2014;383(9934):2041-2042.

56. Heazell A, Whitworth M, Whitcombe J, et al. Research priorities for stillbirth: process overview and results from UK Stillbirth Priority Setting Partnership. *Ultrasound in Obstetrics & Gynecology.* 2015;46(6):641-647.
57. James Lind Alliance. Preterm birth Top 10 priorities. James Lind Alliance. http://www.jla.nihr.ac.uk/priority-setting-partnerships/preterm-birth/top-10-priorities/ Published 2014. Accessed 17 January 2019, 2019.
58. Smith H, Horobin A, Fackrell K, Colley V, Thacker B, Hall DA. Defining and evaluating novel procedures for involving patients in Core Outcome Set research: creating a meaningful long list of candidate outcome domains. *Research involvement and engagement.* 2018;4(1):8.
59. Flemming K, Booth A, Garside R, Tunçalp Ö, Noyes J. Qualitative evidence synthesis for complex interventions and guideline development: clarification of the purpose, designs and relevant methods. *BMJ global health.* 2019;4(Suppl 1):e000882.
60. Gabler G, Coenen M, Lycett D, Stamm T. Towards a standardized nutrition and dietetics terminology for clinical practice: An Austrian multicenter clinical documentation analysis based on the International Classification of Functioning, Disability and Health (ICF)-Dietetics. *Clinical nutrition (Edinburgh, Scotland).* 2019;38(2):791-799.
61. Hamzeh J, Pluye P, Bush PL, Ruchon C, Vedel I, Hudon C. Towards an assessment for organizational participatory research health partnerships: A systematic mixed studies review with framework synthesis. *Evaluation and program planning.* 2019;73:116-128.
62. Pomare C, Churruca K, Ellis LA, Long JC, Braithwaite J. A revised model of uncertainty in complex healthcare settings: A scoping review. *Journal of evaluation in clinical practice.* 2019;25(2):176-182.
63. Carroll C. Qualitative evidence synthesis to improve implementation of clinical guidelines. *Bmj.* 2017;356:j80.
64. Campbell R, Pound P, Morgan M, et al. Evaluating meta ethnography: systematic analysis and synthesis of qualitative research. 2012.
65. Gough D, Thomas J, Oliver S. Clarifying differences between reviews within evidence ecosystems. *Systematic reviews.* 2019;8(1):170.
66. Lee RP, Hart RI, Watson RM, Rapley T. Qualitative synthesis in practice: some pragmatics of meta-ethnography. *Qualitative Research.* 2015;15(3):334-350.
67. Lockwood C, Munn Z, Porritt K. Qualitative research synthesis: methodological guidance for systematic reviewers utilizing meta-aggregation. *International journal of evidence-based healthcare.* 2015;13(3):179-187.
68. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International journal of social research methodology.* 2005;8(1):19-32.
69. Garside R. Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how? *Innovation: The European Journal of Social Science Research.* 2014;27(1):67-79.