

BMJ Open Reaching consensus on reporting patient and public involvement (PPI) in research: methods and lessons learned from the development of reporting guidelines

Jo Brett,¹ Sophie Staniszewska,² Iveta Simera,³ Kate Seers,⁴ Carole Mockford,² Susan Goodlad,⁵ Doug Altman,^{3,6} David Moher,⁷ Rosemary Barber,⁸ Simon Denegri,⁹ Andrew Robert Entwistle,¹⁰ Peter Littlejohns,¹¹ Christopher Morris,¹² Rashida Suleman,¹⁰ Victoria Thomas,¹³ Colin Tysall¹⁰

To cite: Brett J, Staniszewska S, Simera I, *et al.* Reaching consensus on reporting patient and public involvement (PPI) in research: methods and lessons learned from the development of reporting guidelines. *BMJ Open* 2017;**7**:e016948. doi:10.1136/bmjopen-2017-016948

► Prepublication history for this paper is available online. To view these files please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2017-016948>).

Received 21 March 2017

Revised 18 July 2017

Accepted 9 August 2017

ABSTRACT

Introduction Patient and public involvement (PPI) is inconsistently reported in health and social care research. Improving the quality of how PPI is reported is critical in developing a higher quality evidence base to gain a better insight into the methods and impact of PPI. This paper describes the methods used to develop and gain consensus on guidelines for reporting PPI in research studies (updated version of the Guidance for Reporting Patient and Public Involvement (GRIPP2)).

Methods There were three key stages in the development of GRIPP2: identification of key items for the guideline from systematic review evidence of the impact of PPI on health research and health services, a three-phase online Delphi survey with a diverse sample of experts in PPI to gain consensus on included items and a face-to-face consensus meeting to finalise and reach definitive agreement on GRIPP2. Challenges and lessons learnt during the development of the reporting guidelines are reported.

Discussion The process of reaching consensus is vital within the development of guidelines and policy directions, although debate around how best to reach consensus is still needed. This paper discusses the critical stages of consensus development as applied to the development of consensus for GRIPP2 and discusses the benefits and challenges of consensus development.

BACKGROUND

Patient and public involvement (PPI) has become an embedded part of health research nationally and internationally and has the potential to improve the quality, relevance and impact of health research while also improving the transparency of the process and the accountability to the wider community of the researchers themselves. INVOLVE defines public involvement in research as research being carried out ‘with’ or ‘by’

Strengths and limitations of this study

- This study describes the methods for the development of the first international guidance for the reporting of patient and public involvement (PPI) in health and social care research.
- Guidance for Reporting Patient and Public Involvement (GRIPP2) long form and GRIPP2 short form have been developed using the EQUATOR Network robust methods for the development of reporting guidelines.
- Lack of Medical Subject Heading terms for PPI, inconsistency of indexing between databases, large number of titles in searches and difficulty locating evidence of PPI in the papers lead to time consuming and costly systematic review.
- While the online Delphi survey provided a pragmatic and anonymous process for consensus, challenges were encountered with selection bias in the sample, avoiding response fatigue and decision making around the presentation of data.
- Success of the consensus meeting was due to care planning and the critical role of the facilitator.

members of the public rather than ‘to’, ‘about’ or ‘for’ them. This includes, for example, working with research funders to prioritise research, offering advice as members of a project steering group, commenting on and developing research materials, undertaking interviews with research participants, identifying themes in the data collected and aiding dissemination through advocacy.

However, a number of reviews have identified the inconsistency of reporting PPI within papers.^{1–3} This may be attributable to a range of reasons including weaknesses in the way the studies were conducted, undervaluing the importance of reporting the results of the PPI or not recognising the importance of



CrossMark

For numbered affiliations see end of article.

Correspondence to

Dr Jo Brett;
jbrett@brookes.ac.uk

contributing to the PPI evidence base. Poorly reported PPI can lead to a weaker understanding of the evidence base of what works, for whom, in what context and why. This weaker understanding means it is more difficult to implement the findings of studies in terms of best PPI practice and enhancing future PPI.

The challenges in relation to inconsistent reporting in health research more generally led to the development of the EQUATOR Network, which promotes transparent and accurate reporting of research studies and has enhanced the quality of research reporting through the promotion of guidelines such as Consolidated Standards for Reporting Trials (CONSORT)⁴ and Strengthening the Reporting of Observational Studies in Epidemiology (STROBE).⁵ These are now widely used by researchers and journals.^{6,7} While the rate of published PPI studies has increased, there has been a lack of guidance for researchers reporting PPI, which prompted the development of the Guidance for Reporting Patient and Public Involvement (GRIPP) guideline.³ While the original GRIPP checklist represented an important starting point in creating high-quality PPI reporting, its development drew on systematic review evidence, without broader input from the worldwide PPI research community. Achieving consensus is now acknowledged as a crucial step in producing a reporting guideline.⁶ GRIPP2 addresses this gap by developing consensus within the international PPI research networks. This paper reports the methods used to develop an updated version of the GRIPP (GRIPP2) through rigorous systematic reviews and the development of consensus using the method proposed by the EQUATOR Network.⁸ The final checklist, structured in a short form (GRIPP-SF) and in a long form (GRIPP-LF), is presented in a companion paper.⁹ GRIPP2-SF is a short checklist for studies where PPI is a secondary or tertiary focus such as in an randomised controlled trial, and GRIPP2-LF is a longer checklist for studies where PPI forms the primary focus of a study, such as a paper primarily reporting the impact of PPI on the study. For GRIPP2-LF, the entire paper can be shaped by the guidance, with researchers selecting the items of relevance. With GRIPP2-SF, researchers could present all the information in a short section or in a separate box.

METHODS

Three stages recommended by EQUATOR in the development of reporting guidelines were followed for the development of GRIPP2⁸: first, a systematic review of the current evidence of the impact of patient and public involvement in international health and social care research and on National Health Service (NHS) services in the UK; second, a three-phase online Delphi study to gain consensus on the items included in the GRIPP checklist identified from the evidence; and third, a face-to-face consensus meeting with an expert panel to resolve divergences and any remaining uncertainties following the Delphi study and to improve the content and clarity

of the checklist. This paper describes the methods used and highlights the challenges and lessons learnt from developing GRIPP2.

Systematic reviews

GRIPP2 was informed by two systematic reviews.^{1,2} Brett and colleagues aimed to assess the international evidence of the impact of PPI on health and social care research and on the patients, researchers and communities involved (Patient and Public Involvement in Research Impact, Conceptualisation, Outcomes and Measurement -

PIRICOM), while Mockford and colleagues aimed to assess the impact of PPI on the NHS in the UK. The strength of a systematic review lies in the ability to employ a robust and effective search strategy to efficiently integrate existing information and provide data for rational decision making.¹⁰ For each of the systematic reviews, an advisory group, including two lay members and other stakeholders with expertise in PPI and systematic reviews, was established to oversee the systematic review process. The advisory group were consulted at each stage of the review process through regular meetings and by email.

Search strategies combining title and abstract words and database headings relating to PPI were used to locate the evidence of the impact of PPI. Searches were conducted by experienced systematic reviewers (JB and CM) in the electronic databases: MEDLINE, CINAHL, PsychINFO, Health Management Information Consortium, British Nursing Index, Social Science Citation Index, Conference Papers Index, the Cochrane Library, Embase and Web of Science. Hand searching of reference lists of papers and of key journals was also conducted.¹¹ Grey literature (unpublished reports) was identified through searches in InvoNet, NHS Evidence, the Kings Fund Library, National Library for Health and Joseph Rowntree Fund and obtained by contact with experts in the field.¹²⁻¹⁶

Title and abstract searches were conducted to narrow down the number of papers ordered. A set of inclusion and exclusion criteria was used to select papers for the review. A proportion of the papers (10%) were independently assessed by two researchers, to improve the reliability of the inclusion process. . As agreement between the reviewers on included papers was high (94%), and because of the large

number of papers involved in this process, the rest of the papers were reviewed by one reviewer and checked by the second reviewer.

The papers obtained were checked against the inclusion and exclusion criteria,^{1,2} and then quality assessed using the Critical Appraisal Skills Programme.¹⁷ Grey literature was assessed using the checklist developed by Dixon-Woods as used by Hubbard *et al* to review grey literature from cancer studies.^{18,19} Descriptive tables were developed to summarise the evidence.

In order to identify the items for the original GRIPP guideline from the evidence reported, the research team carefully considered each issue in relation to several

criteria: (1) whether the information was important to report within a paper that included some level of PPI, (2) whether it would contribute to enhancing the evidence base of PPI reporting more generally and (3) where the information should be reported to create greater transparency and so enhance the ease of future synthesis.³ Criteria were used to consider where an aspect of PPI should be reported within the structure of a paper, which helped to structure the guideline according to the key sections usually expected within a paper. This process was repeated with updated literature gained for GRIPP2. The aim was to create a guideline that was logically structured and could be easily used by authors when writing PPI papers and reports, editors and peer reviewers when reviewing manuscripts for publication and also for readers in critically appraising published articles and reports.³

Delphi survey

The Delphi technique was used to gain consensus on the updated reporting guideline. The Delphi process sought expert opinion on the included items in the guideline checklist through several rounds of feedback and revisions to develop consensus. As the evidence base from which the reporting guideline items were identified was suboptimal in terms of consistency of reporting, this step was essential to harness subjective judgements from key stakeholder groups in a systematic way and to comment on the suitability and comprehensiveness of the items selected. Three Delphi rounds were selected to gain consensus on the GRIPP2 guideline, reflecting previous EQUATOR guidance development methods.

Ethical approval for the study was secured by the Centre for Education and Industry (CEI) at the University of Warwick who gained generic approval from the University Ethics Committee for all of its online surveys. The committee reviewed the rigorous survey procedures CEI had in place and granted generic ethical approval for its robust process and procedures. The GRIPP2 Delphi survey was assessed by CEI as being covered by the generic approval.

Identification of experts for the Delphi panel was based on the following criteria: individuals with knowledge and/or experience of PPI, individuals working in the field identified through key networks such as INVOLVE and individuals identified through key PPI citations. Individuals were also identified by using a snowball recruitment method where participants forwarded information about the study to other eligible organisations and individuals. This resulted in a diverse sample of stakeholders, including academics, health professional individuals and organisations, patients, carers, patient charities, patient support groups, funders, editors of health-related journals, international organisation networks such as the Health Technology International Citizen and Patient Involvement Group and other European representatives. Our final sample of panellists agreeing to participate in the Delphi process was 143. This composed of 56 researchers, 42 patient and carers, 18 charities or support

groups, 9 members of INVOLVE, the UK NHS Advisory Group for public involvement in research, 5 representatives of the UK National Institute for Health Research and 2 editors of health and social care Research journals. Eleven of the Delphi participants were from international PPI networks. While there are no guidelines that recommended the number of participants that should be included in Delphi panels,²⁰ this was sufficiently large to gain diverse views while providing a manageable amount of data to analyse between the Delphi rounds. This sample size is similar to those involved in the development of previous EQUATOR guidelines.^{8 21}

An electronic online survey was chosen as a practical form of administration. An electronic survey programme called Snapchat²² was used, which offered diverse and flexible function options that were adapted and tested in-house by the Centre for Education and Industry, Warwick University. A pilot study with 10 participants was conducted to check comprehension and acceptability of the questionnaire, and these participants were not involved in the final Delphi survey. The pilot study highlighted a number of issues of importance, such as providing clear instructions of how to access, complete and submit the questionnaire and providing a clear deadline for submission. These issues were addressed before the first round in the main study. The pilot study also facilitated a link to a glossary of terms used, a bibliography of the evidence and to a lay language option on the study research webpage.

A telephone helpline was available for those who had problems completing the questionnaire online, and in three cases a paper questionnaire was sent and completed by participants and manually entered by the researcher.

The Delphi study was conducted over 5 months (September 2013–January 2014). Each round of the Delphi process was open for 4 weeks, with a turnaround of 2–3 weeks for researchers between each stage. Detailed information about the project was sent to participants 2 weeks before the Delphi survey started to ensure respondents had informed decision making around participating and to optimise response rates in each round of the Delphi. Consent was gained prior to participation in the survey. Reminder emails were sent 2 weeks after the start of each Delphi survey if experts had not responded. Unique identifiers were used to enable personalised emails containing a survey link to be sent to participants to aid survey administration and to allow monitoring of responses and issue of timely reminders to non-respondents. Confidentiality was maintained and all questionnaires were only identifiable by a code, with all data kept on a file protected computer system. Only amalgamated results were reported rather than individual responses.

Round 1

In round 1, participants were asked to rate each of the checklist items from 1 to 10, or no judgement. A rating of 1 meant that the respondent considered the item to be unimportant and should be dropped from checklist.

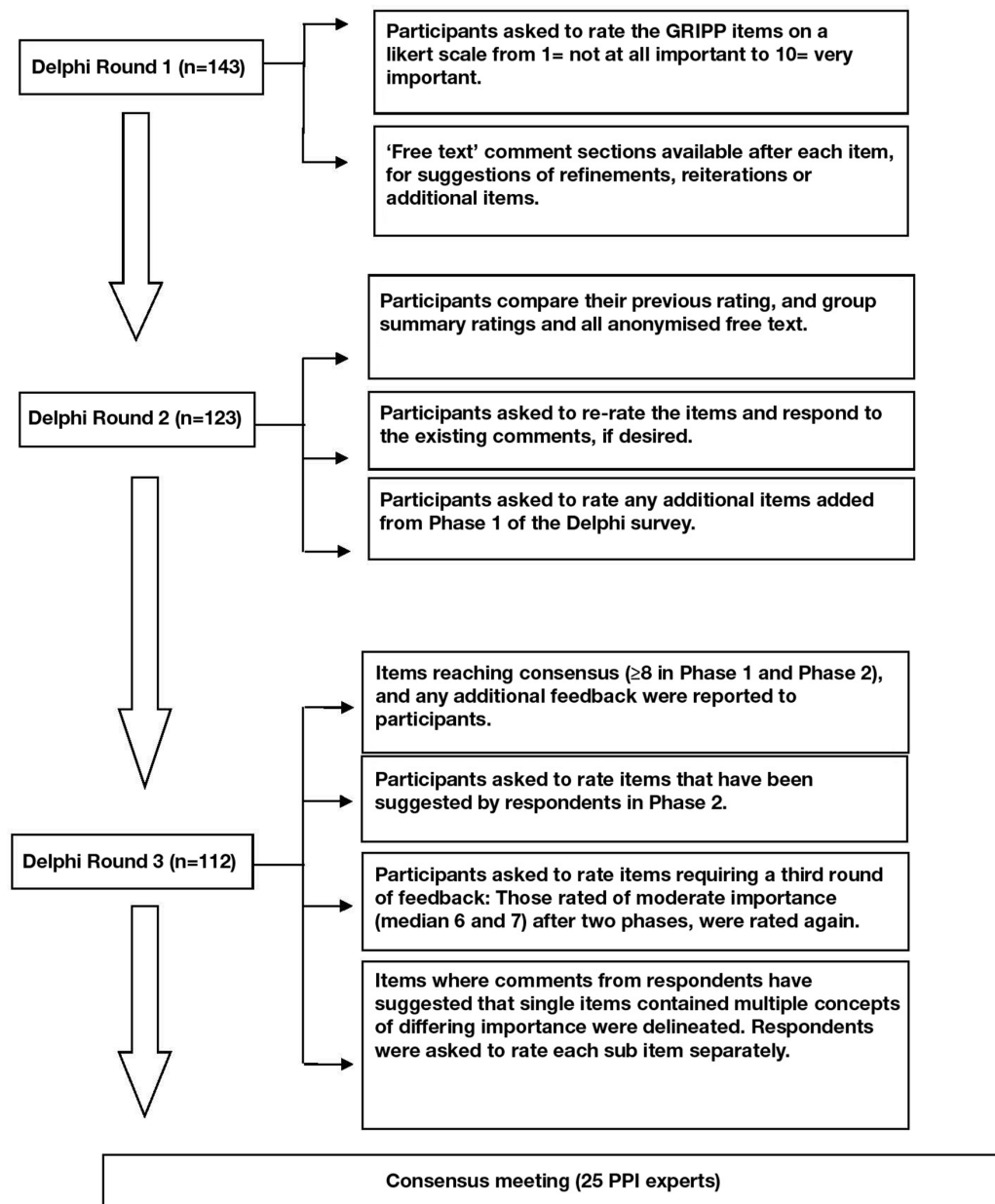


Figure 1 The consensus delphi exercise for GRIPP2.

A rating of 10 meant that the respondent considered the item to be very important and must be included. Each point on the scale had a descriptor. Space was provided against each item for free-text comments to suggest refinements and reiterations or to suggest additional items.

One hundred and forty-three experts took part in this first round. Two researchers analysed the results of round 1, and free-text comments were examined to inform any potential additional items. Consensus was defined by the consistency of median scores (median ≥ 8 =high importance, median 6 or 7=moderate importance and median ≤ 5 =low importance), interquartile ranges (IQR), and the absence of significant issues noted in the text comments. Items reaching high importance or moderate importance were selected for round 2 of the Delphi.

Free-text comments were analysed thematically to identify additional items for round 2.

Round 2

In round 2, participants were asked to rate the GRIPP items again, including any additional items suggested in free-text data in round 1. For each item, panellists were given their previous rating and were also presented with group summary ratings (medians and IQRs), along with all anonymised free-text comments from round 1.

One hundred and twenty-three experts took part in the second round. Two researchers analysed the results of round 2 and agreed items to be included in the third round. The panellists were asked to re-rate the items and add further comments, if desired. As with the first Delphi round, consensus was defined by the consistency

of median scores between rounds (median ≥ 8 = high importance, median 6 or 7=moderate importance and median ≤ 5 =low importance) and the absence of significant issues noted in the text comments. If items scored a median score of ≥ 8 in round 1 and round 2, with low IQRs, they had reached consensus for inclusion.

Round 3

For round 3, the results of items reaching consensus (from rounds 1 and 2) and any additional feedback were presented. Round 3 also included new items introduced in round 2 and items rated of moderate importance (median score 6 or 7) in either or both round 1 and round 2. Additionally, phase 3 included items where comments suggested that single items contained multiple concepts of differing importance. For these, concepts were delineated, and respondents were asked to rate each subitem separately. Items scoring ≤ 5 in both round 1 and round 2 were excluded.

The Delphi process is summarised in [figure 1](#).

The face-to-face consensus meeting

The final stage of the project was a 1-day consensus meeting with 25 experts representing a range of key stakeholders, including lay representatives (n=8), health and social care organisations (n=6), journal editors (n=2) and academics working in the field of PPI (n=9). The agenda for the day included scene-setting for this event through presentations on relevant background topics, including details of the systematic review evidence and results of the Delphi exercise. Materials were sent to participants 2 weeks before the meeting, including: the agenda, participant list, one or two key papers and the results of the Delphi exercise. Consent was gained on the day of the meeting.

The face-to-face consensus meeting followed an approach similar to the Nominal Group Technique^{23 24} by using small group discussion, sharing of ideas and voting techniques. The detailed discussions at the meeting focused on those items that only reached moderate consensus (n=7). Participants were divided into four round-table groups, with a diverse group of stakeholders in each group. Each group had 20 min to discuss each of the items. Participants were encouraged to voice their opinions, with a prerequisite that 'all were equal and every contribution is valid'.²⁵ Opinions arising during their conversations were captured through the use of different media, for example, colour-coded cards, Post-it notes and large pieces of paper placed on the tables. This method is intended to enhance creative thinking, expression and communication.^{25 26} While a professional facilitator hosted the meeting, each table also nominated a 'table host' whose role it was to keep a focus and to encourage all participants to contribute to the discussion.

After the discussion, each table was asked to feedback their comments on each item. After all tables had fed back, each table had a further 5 min to discuss their decision on whether to include the item, and one vote

from each table was recorded. Consensus on whether to include or exclude the item was achieved if three or more tables were in agreement. If two tables voted to include and two tables voted to exclude, further discussion as a group was conducted, with the facilitator recording each viewpoint on a flip chart using the words spoken by the participants. Consensus was then gained through individual votes.

The second half of the consensus meeting discussed issues arising in respect of content and face validity of the checklist. Participants were asked to check the wording and make any suggested changes to wording directly onto each item. Comments around the comprehensibility of the item were also sought. Suggested modifications were made by the research team following the consensus meeting.

A key session towards the end of the face-to-face meeting was held to discuss the 'knowledge translation' strategy to assist with the translation of the reporting guideline into practice. A publication strategy was also developed, and discussions included how the guideline can be implemented by journals.

Evidence of the methods used for developing reporting guidelines have been reported elsewhere.^{27 28}

DISCUSSION AND CONCLUSION

This paper has detailed the development of consensus on the items that researchers should consider when reporting PPI. The lessons learnt are described below.

Important aspects and lessons learnt from the development process

Systematic reviews

While the evidence identified in the systematic review was sufficient to identify key areas of importance to structure specific criteria for the reporting guidelines, several pitfalls can be highlighted in conducting systematic reviews around PPI.

Searching databases to identify potential papers for these reviews provided a number of challenges. As there are no Medical Subject Heading (MeSH) for 'PPI', a combination of search terms was used for the electronic databases. A lack of specific search terms led to a lack of sensitivity in the initial searches, resulting in a large number of papers identified initially.^{1 2} Databases are not consistent in their indexing of studies relating to PPI, which poses many challenges for developing search strategies that aim to locate these papers. Databases also vary the search terms used, which means the search strategies need to vary by database, increasing the complexity of searching and the potential for error.

This phase of the study was therefore cumbersome and costly. Standardising the terms used for PPI would improve search strategies for future reviews of PPI evidence and improve comprehension around PPI among health researchers. Standardised terms could then be adopted by electronic databases improving the MeSH search terms.

Table 1 Summary of lessons learnt**Systematic review**

- ▶ lack of MeSH terms for PPI in databases made searches broad
- ▶ no consistency in indexing between databases leads to development of separate search strategies for each database
- ▶ difficulty identifying PPI data as written up in methods or discussion sections of papers
- ▶ pragmatics of all titles reviewed by two researchers led to 10% reviewed twice.

Online Delphi study**Advantages:**

- ▶ pragmatic consensus approach to gain large, geographically dispersed group
- ▶ participants complete the questionnaire at their leisure
- ▶ participants remain anonymous allowing greater honesty of comments.

Challenges:

- ▶ selection bias—as self-selected. Tendency to recruit participants who are more favourable or more disapproving.
- ▶ excludes those who do not have a computer and those who are not computer literate
- ▶ decision making on number of Delphi rounds to gain consensus. Need enough to gain consensus but not too many that you have low response rates for subsequent rounds
- ▶ difficulty of choosing off-the-shelf data collection packages—importance of customising software
- ▶ issues of response fatigue with too many categories in the Likert scale—the higher the number, the greater the response fatigue
- ▶ need for clear instructions, glossary of terms, bibliography of evidence and other information in lay language
- ▶ use of median score and IQRs versus mean (SD) and modes in presentation of data to overcome extreme values in dataset.

Consensus meeting

- ▶ need for openness and informal approach to encourage discussion and plurality of ideas
- ▶ critical to have expert facilitator to guide the consensus meeting to ensure key and timely contributions from all members
- ▶ importance of planning for dissemination, implementation and adoption of outcomes.

Care was taken with decisions about where and how to search. For example, a dearth of peer-reviewed evidence on PPI studies indicated the importance of searching the grey literature. Restricting the searches to electronic databases, which consist mainly of references to published peer-reviewed journal articles, could have excluded many PPI papers, leading to publication bias.

While it is recommended that all abstracts are reviewed by two researchers in systematic review methodology,¹¹ this greatly increases the cost and time of the process.²⁹ Evidence shows only an 8% improvement in identification of relevant papers when all abstracts are reviewed by two researchers.¹¹ A large number of abstracts were retrieved in the searches, therefore for pragmatic reasons, 10% of abstracts were reviewed by two researchers. A high level of consistency was reported, indicating the reliability of the searches.

The quality of the evidence was also very difficult to assess as the PPI evidence was often inconsistently reported and difficult to locate. For example, PPI studies were sometimes reported in full in either the methods or the discussion of the paper.^{1 2} Quality assessments were therefore conducted to assess inclusion but not used to weigh the papers. If studies were fatally flawed in terms of their quality, they were excluded. Care should always be taken when interpreting the results of critical appraisals as they can be biased because of the subjective nature in which decisions are made by researchers. One study compared Critical Appraisal Skills Programme with the quality framework and intuitive judgement by expert opinion³⁰ and found that no difference could be detected between the different critical appraisal tools and intuitive judgement by expert opinion.

Delphi survey

While a three-phase online Delphi survey was chosen to gain consensus on the reporting guideline, other methods were considered including nominal group technique,³¹ analytic hierarchy process technique³² and use of separate working groups gathering consensus through focus groups.³³

From a pragmatic view point, Delphi methods allowed a large number of geographically dispersed stakeholders to be involved in the consensus process that may not have been possible through alternative consensus processes due to time and cost limitations.³⁴ Respondents could complete the questionnaire at their leisure, and this reduced time pressures and may have allowed for more reflection and contemplation of responses.³⁵

Additionally, the advantage of this method is that members remain anonymous in responding to individual questions and this is likely to encourage opinions that are free of influences from others and more likely to be 'true'.^{36 37} It also provides an iterative process with controlled feedback, and average score responses from a group of experts, providing stability of responses. Furthermore, it recognises and acknowledges the contribution of each participant. This method can therefore facilitate

consensus where there is contradictory or insufficient information to make effective decisions.³⁵⁻³⁸

The disadvantages of Delphi methods, which are perhaps true for any consensus method, are that there is a purposeful selection of 'experts' chosen because the respondents' reputations are known to the researcher. These experts meet a minimal number of criteria of familiarity with the research field, self-rating their expertise.³⁹ Furthermore, the self-selected sample may be biased in that they are willing to take part and therefore either more favourable and more inclined to agree with items included in the GRIPP2 checklist or disapproving and more likely to disagree with the items included.³⁹

Delphi methods traditionally use two or more rounds to gain consensus.⁴⁰ The number of rounds depends on the level of initial consensus gained but may also be controlled by time and cost limitations. Studies focusing on the number of rounds needed in a Delphi survey to achieve consensus suggest that most changes occur in the transition from the first to the second round.⁴¹ When the number of rounds exceeds four, the response rates can be very low due to the response burden on participants.⁴¹

The possibility that participants may alter their estimates in order to conform to the group (conformance), without actually changing their opinions (consensus), was considered. However, evidence suggests that the influence of expert knowledge helps move towards consensus rather than conformance with the 'median score moving towards true value'.⁴²

An electronic survey was chosen over postal surveys. This allowed a fast deployment of surveys and a quick return time. It was relatively low cost, removing the cost and time of printing, posting and data entry. It also saved the participants from the inconvenience of posting a survey back to us. The disadvantage of this method is that the participants had to have access to a computer and be computer literate. This may have biased the sample, particularly excluding some hard-to-reach patients and carers.

An online survey software, Snapchat, was adapted by in-house expertise to record the survey responses.²¹ Other electronic packages were considered such as Survey Monkey, Zoomerang, GoogleForms and Survey-Gizmo,⁴³⁻⁴⁶ but Snapchat was internally available and proved a reliable and flexible survey software for the study's requirements. Furthermore, using in-house expertise allowed us to tailor and customise the online form to our needs more effectively thus allowing us to refine the design and functionality after piloting. Survey responses were monitored, and the survey data were cleaned and analysed to produce the results in the required format.

The electronic software offered a save and return function within the survey, allowing participants to reflect and return to the survey. This also would have minimised the risk of incomplete responses where participants sometimes underestimate the time the survey will take and then run out of time to complete the questionnaire in full.

Consideration was given to the number of categories in the Likert scale for rating the items within the GRIPP2 Delphi survey. Streiner and Norman⁴⁷ have argued that the benefit from large numbers of options is subject to the law of diminishing returns and that from the 7-item scale and upwards, the scales become too cumbersome to use. Any additional benefits are cancelled out by respondent fatigue and reliability plummets.⁴⁷ However, in reality, we are often asked to rate issues on a scale of 1–10; it provides a better opportunity to detect and discriminate when responses are skewed at one end of the scale, and it felt more ‘natural’ to patient advisors on the research team.⁴⁸ This is also how previous Delphi surveys had been scored in the development of reporting guidelines through EQUATOR.

Different descriptive statistics can be used when feeding back data in each round. While the mean, mode and median scores provide three forms of averages,⁴⁹ the range, IQR and SD are all measures that indicate the amount of variability within a dataset.⁴⁹ As there was the potential for a range of scores for each item, the median was selected as a better score of the average than the mean or the mode. The range is the simplest measure of variability to calculate but can be misleading if the dataset contains extreme values. The IQR reduces this problem by considering the variability within the middle 50% of the dataset. The SD is used to take into account a measure of how every value in the dataset varies from the mean. The IQR was therefore used to report the variation from the median scores.⁵⁰

The validity and reliability of a Delphi study may be questioned due to the subjective nature of the feedback and potential instability of responses. However, this approach engages a wide range of expertise more effectively than any other group consensus method and provides a ‘fair’ representation of the views of each participant because each participant has an equal opportunity to have their views taken into account. Furthermore, Delphi methods clearly state the rationale for inclusion or exclusion in the final checklist, whereas other methods of gaining consensus rarely provide such a transparent decision trail for each item. The quality of the Delphi survey was further increased by the quick turnaround between Delphi rounds.⁵¹ The Delphi study for the reporting guidelines provided opinions from a representative sample of all PPI stakeholders, carefully and rigorously collated over 3 rounds of voting. In this study the Delphi methods used brought agreement from a diverse group of stakeholders whose commitment to the project resulted in good response rates at each stage of the Delphi project. The validity and reliability of the process was therefore deemed satisfactory.

Finally, ethical issues were also examined in this Delphi survey. The main ethical issues related to consent to participate, privacy and confidentiality of the data provided, all of which were addressed in the methods used.

Consensus face-to-face meeting

The consensus face-to-face meeting was important to finalise consensus on the reporting guideline. An informal approach to voting was adopted where consensus was gained through the round-robin process with a gradual move towards synthesis and building consensus rather than through anonymous voting.⁵¹ This approach encourages the sharing and discussion of the reasons for the choices made by each group member, thereby identifying common ground and a plurality of ideas and approaches.⁵¹ The meeting also ensured face and content validity through small group discussion.

The success of the consensus group was enhanced by careful planning and commitment of the stakeholders. Furthermore, a critical role in the success of the consensus group was that of the facilitator. Key aspects of this role were ensuring that participants of the group understand their roles and adopting the listening stance so participants all felt listened to. This important role helped to ensure that balanced views were recorded through individual and group work. Facilitated groups develop greater consensus than user-driven groups.^{52 53} However, highly structured facilitation can have an adverse effect on the consensus process, and an element of flexibility in the process is recommended.^{52 53} A highly skilled facilitator was therefore used to mediate the group process and to ensure key and timely contributions from all members. The facilitator presented items where consensus decisions were needed and guided the participants to reach agreement, ensuring all have an opportunity to participate. For the items related to economic assessment and testing conceptual or theoretical models, where decisions on whether to include or not were difficult, the facilitator used problem-solving techniques to finalise the decision. This involved the group drawing up a list of the pros and cons of the item and asking them to review and evaluate the list and then to re-evaluate their initial decision.

An important aspect of the consensus meeting was also agreeing a plan of action for dissemination, implementation and adoption of GRIPP2. Evaluation of GRIPP2 will be conducted through ongoing feedback from authors.

Table 1 below summarises the lessons learnt during this study:

Contribution of the patient partners

The patient partners contributed to the development of GRIPP in a number of ways. Throughout the initial stages of collating the evidence and identifying items for the GRIPP checklist, the patient partners highlighted the importance of including items referring to the context and processes of PPI, suggesting that this affected the impact that PPI had on research. The patient partners, along with other patient organisations and charities, recruited nearly half of all participants for the Delphi survey. The patient partners helped other patients with the technical aspects of completing the online survey, improving the response rate in each Delphi survey round. The patient partners checked the comprehension of the changed

items and comments from the lay perspective between rounds and were integral to helping the researchers keep to the scheduled time of the Delphi survey. Throughout the write-up phase for both the results paper and the methods paper, the patient partners contributed to the lay sections and contributed to edits of the paper.

CONCLUSION

In conclusion, this paper details the methods used in the development of EQUATOR-recognised guidelines for the reporting of PPI in research (GRIPP2).⁹

GRIPP2 has been developed using the robust methods used in the development of other EQUATOR guidelines such as CONSORT and STROBE. The development process involved identification of relevant evidence through systematic reviews, consensus of included items through an online Delphi survey with PPI experts and a face-to-face meeting of PPI experts to finalise consensus.

Systematic reviews informed the development of items for the reporting guideline. However, the systematic searches were time-consuming due to the lack of MeSH terms for PPI in the electronic databases, the inconsistency of indexing between electronic databases, the large number of titles identified in the searches and the poor reporting of PPI in the papers. An online Delphi survey provided a pragmatic and anonymous process for consensus, although challenges were encountered with selection bias in the sample, avoiding response fatigue and decision making over the presentation of data. The success of the consensus meeting was due to careful planning and the critical role of the facilitator.

GRIPP2 has been developed to improve the accuracy and consistency of PPI reporting in health research to improve interpretation and better application in future research. With a growing evidence base, we expect greater discussion around the conceptualisation and theoretical underpinning of PPI to provide a greater insight into practices and processes.^{54–56} As the field of PPI develops, we expect further development of GRIPP2 to refine and update the guidance.

Author affiliations

¹Department of Midwifery, Community and Public Health, Faculty of Health and Life Sciences, Oxford Brookes University, Oxford, Oxfordshire, UK

²Division of Health Sciences, RCN Research Institute, Warwick Medical School, University of Warwick, Coventry, UK

³Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, EQUATOR Network, Centre for Statistics in Medicine, University of Oxford, Oxford, Oxfordshire, UK

⁴Division of Health and Social Care Research, RCN Research Institute, Warwick Medical School, University of Warwick, Coventry, UK

⁵Centre for Research in Psychology, Behaviour and Achievement, University of Coventry, Coventry, UK

⁶Nuffield Department of Orthopaedics, Rheumatology & Musculoskeletal Sciences, Centre for Statistics in Medicine, Oxford, UK

⁷Ottawa Hospital Research Institute, School of Epidemiology, Public Health, and Preventive Medicine, University of Ottawa, Ottawa, Ontario, Canada

⁸School of Health and Related Research, University of Sheffield, Sheffield, UK

⁹INVOLVE, National Institute of Health Research (NIHR), University of Southampton, Southampton, UK

¹⁰UNTRAP, University of Warwick, Coventry, Warwick, UK

¹¹Faculty of Life Sciences, King's College London, London, UK

¹²Peninsula Cerebra Childhood Disability Research Unit (PenCRU), University of Exeter Medical School, Exeter, UK

¹³Patient and Public Involvement Unit, Public Involvement Programme, National Institute for Health and Clinical Excellence, London, UK

Acknowledgements We are very grateful to everyone who participated in the Delphi survey and attended the consensus event. We are grateful for Sally Crowe who facilitated the consensus workshop. SS is part funded by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care West Midlands. PL is supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King's College Hospital NHS Foundation Trust.

Contributors SS, JB, IS, KS, CM, SG, CM, SD, RB, PL, VT, RS, AE and CT made substantial contributions to conception and design. JB, SS, IS, DA, and DM made substantial contributions to developing the protocol. SS, JB, SG made substantial contributions to the acquisition of data, analysis, and interpretation of data. All authors have been involved in drafting the manuscript or revising it critically for important intellectual content; given final approval of the version to be published.

Funding This study was funded by RCNRI, Warwick Medical School, University of Warwick.

Disclaimer This paper presents independent research and the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests None declared.

Ethics approval University of Warwick Research Ethics Committee.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Deidentified data will be shared through university accessible databases or repositories at Warwick University and Oxford Brookes University. Please contact Dr JB if additional information is required: email: Jbrett@brookes.ac.uk.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Brett J, Staniszewska S, Mockford C, *et al.* Mapping the impact of patient and public involvement on health and social care research: a systematic review. *Health Expect* 2014;17:637–50.
2. Mockford C, Staniszewska S, Griffiths F, *et al.* The impact of patient and public involvement on UK NHS health care: a systematic review. *J Qual Health Care* 2012;24:28–38.
3. Staniszewska S, Brett J, Mockford C, *et al.* The GRIPP checklist: strengthening the quality of patient and public involvement reporting in research. *Int J Technol Assess Health Care* 2011;27:391–9.
4. Schulz KF, Altman DG, Moher D. CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;23:c332.
5. von Elm E, Altman DG, Egger M, *et al.* The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet* 2007;370:1453–7.
6. Moher D, Altman D, Shultz K, *et al.* *Guidelines for reporting health research: a user's manual.* BMJ Books, 2014.
7. Simera I, Moher D, Hirst A, *et al.* Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. *BMC Med* 2010;8:24.
8. Moher D, Schulz KF, Simera I, *et al.* Guidance for developers of health research reporting guidelines. *PLoS Med* 2010;7:e1000217.

9. Staniszewska S, Brett J, Simera I, *et al*. Guidance for reporting patient and public involvement in health research: A Delphi consensus study to develop GRIPP2-SF and GRIPP2-LF. *BMJ* 2016.
10. Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126:376–80.
11. Centre for Reviews and Dissemination (CRD). *Systematic reviews: guidance for undertaking reviews in health care*. University of York, 2008. <https://www.york.ac.uk/>
12. INVOLVE: InvoNET website. www.invo.org.uk (accessed Jun 2012).
13. NHS Evidence. www.evidence.nhs.uk (accessed Jun 2012).
14. Kings Fund Library. www.kingsfund.org.uk/library (accessed Jun 2012).
15. National Library for Health. www.library.nhs.uk (accessed 2012).
16. Joseph Rowntree Fund Library. www.jrf.org.uk/Library (accessed Jun 2012).
17. Critical Appraisal Skills Programme (CASP). *CASP Checklists (URL used)*. Oxford: CASP, 2014.
18. Dixon-Woods M, Kirk D, Agarwal S, *et al*. *Vulnerable groups and access, to health care: a critical interpretative review*. London: National Co-ordinating centre for NHS service, 2005.
19. Hubbard G, Kidd L, Donaghy E, *et al*. A review of literature about involving people affected by cancer in research, policy and planning and practice. *Patient Educ Couns* 2007;65:21–33.
20. Williams PL, Webb C. The Delphi technique: a methodological discussion. *J Adv Nurs* 1994;19:180–6.
21. Tetzlaff JM, Moher D, Chan AW. Developing a guideline for clinical trial protocol content: Delphi consensus survey. *Trials* 2012;13:176.
22. Snap Chat. www.snapchat.com (accessed Sep 2011).
23. Harvey N, Holmes CA. Nominal group technique: an effective method for obtaining group consensus. *Int J Nurs Pract* 2012;18:188–94.
24. Gallagher M, Hares T, Spencer J, *et al*. The nominal group technique: a research tool for general practice? *Fam Pract* 1993;10:76–81.
25. Brown JID. *The World Cafe. Shaping our futures through conversations that matter*. Berrett koehler publishers. Inc, 2005:765.
26. Wilkinson JE, Rycroft-Malone J, Davies HT, *et al*. A creative approach to the development of an agenda for knowledge utilization: outputs from the 11th international knowledge utilization colloquium (KU 11). *Worldviews Evid Based Nurs* 2012;9:195–9.
27. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191–4.
28. Moher D. Guidelines for reporting health care research: advancing the clarity and transparency of scientific reporting. *Can J Anaesth* 2009;56:96–101.
29. Attree P. Growing up in disadvantage: a systematic review of the qualitative evidence. *Child Care Health Dev* 2004;30:679–89.
30. Dixon-Woods M, Sutton A, Shaw R, *et al*. Appraising qualitative research for inclusion in systematic reviews: a quantitative and qualitative comparison of three methods. *J Health Serv Res Policy* 2007;12:42–7.
31. Carney O, McIntosh J, Worth A. The use of the nominal group technique in research with community nurses. *J Adv Nurs* 1996;23:1024–9.
32. Lai VS, Wong BK, Cheung W. Group decision making in a multiple criteria environment: a case using the AHP in software selection. *Eur J Oper Res* 2002;137:134–44.
33. Morgan DL. Focus groups as qualitative research. *Planning and design of focus groups*. 2nd ed London: SAGE Research Methods, 1997.
34. Rogers MR, Lopez EC. Identifying critical cross-cultural school psychology competencies. *J Sch Psychol* 2002;40:115–41.
35. Linstone HA, Turoff M. The delphi method techniques and method? *Nursing Research* 1975;46:116–8.
36. Goodman CM. The Delphi technique: a critique. *J Adv Nurs* 1987;12:729–34.
37. Snyder-Halpern R. Indicators of organizational readiness for clinical information technology/systems innovation: a Delphi study. *Int J Med Inform* 2001;63:179–204.
38. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;32:1008–15.
39. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. *Int J Nurs Stud* 2001;38:195–200.
40. Cuhls K. *The Delphi Method. Fraunhofer institute for systems and innovation research*. Germany, 1998.
41. van Zolingen SJ, Klaassen CA. Selection processes in a Delphi study about key qualifications in Senior Secondary Vocational Education. *Technol Forecast Soc Change* 2003;70:317–40.
42. Munier F, Rondé P. The role of knowledge codification in the emergence of consensus under uncertainty: empirical analysis and policy implications. *Res Policy* 2001;30:1537–51.
43. Survey Monkey. 1999–2016. <https://www.surveymonkey.com/>
44. Zoomerang (online web survey). 1999–2016. <http://www.zoomerang.com/>
45. Google forms (Online web survey). 2007. www.google.com/forms
46. Survey Gizmo. 2005–2016. <https://www.surveygizmo.com/>
47. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*, 2008.
48. Wittnk DR, Bayer LR. “The measurement imperative” marketing research. 2003;6:14–22.
49. Altman D. *Practical statistics for medical research (chapman & hall)*, 1990.
50. Murphy MK, Black NA, Lamping DL, *et al*. Consensus development methods, and their use in clinical guideline development: a review United Kingdom, Health Technology Assessment. 1998;2:49.
51. Dunham RB. *Organisational behaviour: The delphi technique*. US: University of wisconsin school of business, 1998.
52. Anson R, Bostrom R, Wynne B. An experiment assessing group support system and facilitator effects on meeting outcomes. *Manage Sci* 1995;41:189–208.
53. Dickson G, Partridge J, Robinson L. *Exploring modes of facilitative support for GDSS technology MIS quarterly*. 1993;17:173–94.
54. Ives J, Damery S, Redwod S. PPI, paradoxes and Plato: who's sailing the ship? *J Med Ethics* 2013;39:181–5.
55. Brett J, Staniszewska S, Mockford C, *et al*. *The PIRICOM Study: A systematic review of the conceptualisation, measurement, impact and outcomes of patients and public involvement in health and social care research*. UK-CRC, 2010.
56. Forbat L, Hubbard G, Kearney N. Patient and public involvement: models and muddles. *J Clin Nurs* 2009;18:2547–54.