Using consensus group methods such as Delphi and Nominal Group in medical education research

Susan Humphrey-Murto, Lara Varpio, Carol Gonsalves & Timothy J. Wood

View supplementary material

Published online: 12 Nov 2016.

Submit your article to this journal

Article views: 5211

View related articles

View Crossmark data

Citing articles: 158 View citing articles

AMEE GUIDE

Using consensus group methods such as Delphi and Nominal Group in medical education research*

Susan Humphrey-Murto[a], Lara Varpio[b], Carol Gonsalves[a] and Timothy J. Wood[a]

[a]Division of Rheumatology, University of Ottawa, Ottawa, Canada; [b]School of Medicine, Uniformed Services University of the Health Sciences, Bethesda, MD, USA

ABSTRACT

Consensus group methods are widely used in research to identify and measure areas where incomplete evidence exists for decision-making. Despite their widespread use, these methods are often inconsistently used and reported. Using examples from the three most commonly used methods, the Delphi, Nominal Group and RAND/UCLA; this paper and associated Guide aim to describe these methods and to highlight common weaknesses in methodology and reporting. The paper outlines a series of recommendations to assist researchers using consensus group methods in providing a comprehensive description and justification of the steps taken in their study.

Introduction

Medical educators strive to use the latest findings from scientifically grounded research when making decisions regarding education and assessment in the health professions. Empirical evidence from educational research, however, is often limited or contradictory (van der Vleuten 2014). When faced with conditions of uncertainty or incomplete evidence, consensus group methods such as Delphi, nominal group technique (NGT), and the RAND corporation/University of California Los Angeles appropriateness method (RAND/UCLA) are widely used to synthesize expert opinions and enhance decision making (Jones & Hunter 1995; Murphy et al. 1998, p. 4; Campbell et al. 2001; Fitch et al. 2001).

Standard textbooks rarely provide clear guidelines to help researchers utilize these methods, and studies using them often lack methodological rigor (Boulkedid et al. 2011; Sinha et al. 2011; Diamond et al. 2014; Waggoner et al. 2016). The AMEE Guide describes these methods to provide a "how to" approach, highlight common weaknesses in methodology and reporting, and outline recommendations for reporting future consensus based studies.

Why use consensus methods in education?

Just as clinicians regularly make difficult choices about treatment options, health professional educators must make decisions in the face of uncertainty. In medical education, consensus group methods are used to determine components of a new or revised curriculum, develop items for assessment tools, define competencies, and develop educational resources (Humphrey-Murto et al. 2014).

Practice points

- Consensus group methods such as Delphi and Nominal Group Technique are commonly used, but there is a lack of standardization in definitions, methodology and reporting.
- Very few medical education textbooks describe these methods in sufficient detail to serve as a "how to use guide".
- We provide suggestions that researchers should consider when planning and publishing studies using consensus group methods, in order to maximize methodological rigor.

Consensus group methods – definition and rationale

Consensus methods are defined as a systematic means for measuring and developing consensus. The goal of these methods is to establish how well experts agree on a particular issue, and is based on the idea that accurate and reliable assessment can be best achieved by consulting a panel of experts and accepting the group consensus (Campbell et al. 2001; Tammela 2013). The case for using formal consensus group methods is based on assumptions about decision-making in groups; potential advantages include a wider range of knowledge and experience available, debate may challenge ideas and stimulate new ones, and group consensus may be seen as more credible (Murphy et al. 1998, p. 1).

The consensus group methods should have the following common features: anonymity, iteration, controlled

---

feedback, statistical group response and structured inter-action (Jones & Hunter 1995; Murphy et al. 1998; Vernon 2009). This differentiates informal group meetings from formal consensus methods.

Consensus group methods

Several consensus methods have been described. Two consensus methods that predominate are the Delphi and NGT. The RAND/UCLA is a hybrid of these two methods.

Delphi

The Delphi technique was initially described by the RAND Air Force Corporation in America in the 1950s (Murphy et al. 1998). The Delphi has been used in many fields including science and technology, health, business, communication, education, policy analysis (de Loe et al. 2016; Vazquez-Ramos et al. 2007). In medical education, the Delphi and modified Delphi account for ∼75% of papers using consensus group methods (Humphrey-Murto et al. 2014).

The Delphi method and process

The Delphi method includes the following stages: identifying a research problem, selecting participants, developing a questionnaire of statements, conducting anonymous iterative postal or email questionnaire rounds, collecting individual and group feedback between rounds and summarizing the findings (Jones & Hunter 1995; Murphy et al. 1998; Campbell et al. 2001). This process is repeated until the best possible level of consensus is reached, or until a pre-determined number of rounds have been completed. Participants never meet or interact directly in the classically-described Delphi method (Murphy et al. 1998) (see Table 1 available online as Supplemental Material).

The initial identification of the research problem usually involves a group of experts who share an interest in seeking solutions or recommendations to a particular problem. For example, a group of researchers wanted to develop a national neurology curriculum for internal medicine residents in Canada (Lazarou et al. 2011). The research team developed the questionnaire based on reviewing neurology textbooks and adding additional topics generated by the research team. The initial survey was composed of a comprehensive list of all topics in neurology.

Round 1 of a Delphi involves mailing out the survey to expert participants who rank their agreement with each statement, and may be allowed to add new items to the list. The number of participants in the Delphi has ranged from 4 to 3000 (Campbell et al. 2001). In the national neurology example, the participants were asked to rate each item on a 5-point scale (definitely include, possibly include, neutral, possible exclude, and definitely exclude). The participants in this study consisted of directors of internal medicine programs and neurology programs for all Universities across Canada as well as internal medicine residents from those institutions. In some Delphi studies, there is no initial questionnaire development; instead, the initial round is for idea generation from the participants. This is very relevant when little is known about a particular topic.

Between rounds, the research team collates the rankings and the anonymous collated results are sent back to participants for review. The quantitative data might include the mean, median and/or frequency distribution for each item. Participants would usually be able to see their ranking relative to other participants, and have an opportunity to re-rank the items. In some studies participants would also be asked to provide written feedback when their views differed substantially from the others.

The list of items and participants may vary for each round. In some studies the entire list of items would be sent for each round, while in others only those items in which there was a lack of agreement would be re-sent. The participant group may be the same for each round, or, as in this example, only the program directors are included in the second round. The number of rounds may vary from 2 to over 10. The final ranking or list of items may be sent to the participants for final verification.

The benefits of the Delphi method include the potential inclusion of a large number of participants who are geographically dispersed, and clearly are favored for international research. It is relatively inexpensive and avoids undue dominance by specific individuals by providing the greatest degree of anonymity but may limit discussion and debate.

Nominal group technique (NGT)

The NGT was developed by Delbecq and Van de Ven in the 1960s (Murphy et al. 1998, p. 3). The NGT shares several features of the Delphi, but in contrast is a structured face-to-face interaction usually involving 5–12 participants. It is often used for item generation and provides an opportunity for discussion (Jones & Hunter 1995). The steps are outlined in Table 1 (available online as Supplemental Material). The research team formulates a nominal question and gathers a panel of expert participants. Unlike the Delphi, a literature review is not always completed at this time, but background information is provided to the participants. The participants meet face to face and respond to the nominal question by recording his/her ideas independently and privately. These ideas are then shared with the group in a round-robin format, with each participant sharing one item from their list. These ideas are recorded by a facilitator who documents the responses until all participants have no more original ideas. Often, a flip chart or list of the responses is posted for all to see. The facilitator then leads a group discussion where each idea is discussed in turn, with similar ideas grouped together, and clarification provided. In some settings, the process may end here. Usually, however, individuals then vote privately on the items and results are then fed back to the group in aggregate (anonymously). Further discussion and voting may take place (Murphy et al. 1998). The entire process may last from 1.5 to up to 6 hours (Campbell et al. 2001). A facilitator is required to effectively run the session.

Advantages of the NGT include the generation of a larger number of ideas and the potential for discussion and debate. Limitations include a smaller number of participants than the Delphi, and the potential for dominant participants to unduly influence the group.

RAND/UCLA appropriateness method (RAND/UCLA)

The RAND/UCLA method was developed to enable the measurement of the overuse and underuse of medical and surgical procedures (Fitch et al. 2001). The concept of appropriateness refers to the fact that the expected health benefits should exceed the expected negative consequences. RAND/UCLA has also been referred to as a hybrid of the Delphi and NGT. Like Delphi, it begins with identification of a research problem, completion of a literature search, and development of a questionnaire of statements, which is sent out to participants. The questionnaire might involve a list of specific clinical scenarios or indications of patients who may present for a certain treatment. Participants also receive background information (literature review, definitions). In the first round, participants do not meet face-to-face; instead, they rank each item individually via email or mailed survey. In the original description, for each indication, the participants rate the benefit-to-harm ratio of the procedure on a scale of 1–9, where 1 means that the expected harms greatly outweigh the expected benefits, and 9 means that the expected benefits greatly outweigh the expected harms. A middle rating means that the harms and benefits are about equal for the patient described (Fitch et al. 2001).

Like the NGT, the next step involves a face-to-face meeting where collated results are fed back to each individual member of the group. The participants have an opportunity to discuss their ratings, in light of how other participants have voted after which private voting occurs again. Consensus should not be forced and the outcome may lead to a classification of "appropriate", "uncertain", or "inappropriate". A common scale used is median scores in the 1–3 range are classified as inappropriate, 4–6 range as uncertain and 7–9 range as appropriate. A rating of uncertain would also be allocated if there was disagreement; in other words, all participant ratings did not fall within any 3-point range (Fitch et al. 2001). The number of participants has traditionally been 9, large enough for diversity of representation yet small enough to allow discussion.

The RAND/UCLA method is not intended for idea generation and assumes there is supporting data available to create the initial questionnaire of a highly structured list (Fitch et al. 2001). As with the NGT an effective facilitator is required. Table 2 (available online as Supplemental Material) highlights some of the key differences between the methods.

The problem with consensus group methods

Despite the extensive use of these seemingly simple research methods, concerns regarding consensus group methods are longstanding (Sackman 1975; Goodman 1987; Hasson et al. 2000). The large number of modifications to the classical Delphi has led to considerable confusion (Crisp et al. 1997; Keeney 2001)

Four recent reviews using the Delphi in health care and policy-related research have systematically explored deficiencies in the use and reporting of consensus group methods. Collectively, these studies have noted deficiencies regarding: information provided to the participants at the start of Delphi, reporting response rates, feedback to participants, level of anonymity, outcomes after each round and the definition of consensus (Boulkedid et al. 2011; Sinha et al. 2011; Diamond et al. 2014; de Loe et al. 2016). In the field of medical education, a recent scoping review of consensus methods demonstrated similar findings (Humphrey-Murto et al. 2014). This research clearly suggests that it is important to move away from the use of labels and move toward a comprehensive description of the steps taken in a specific study.

Qualitative or quantitative: underpinning methodology of consensus group methods

Some authors consider these methods to be straddling both qualitative and quantitative methodologies (Stewart 2001; De Villiers 2005, p. 642). If one believes that the Delphi method supports a constructivist enquiry, then rigor should be defined in terms of trustworthiness criteria used in qualitative research (confirmability, credibility, transferability, and dependability) and not by positivist criteria of objectivity, namely validity and reliability (Campbell et al. 2001). The absence of an appropriate philosophical foundation results in inconsistent conceptualizations of the methods and may contribute to the inconsistency in methodology (Guzys 2015).

Recommendations for demonstrating methodological rigor for consensus group methods

When undertaking research using consensus methods, the following recommendations should be considered in order to add credibility to the research process and ensuing results (Table 3 available online as Supplemental Material).

Define the purpose or objective of the study

As with all research, the authors must provide a clear purpose for their study or line of inquiry. This will guide future decision-making regarding, for example, selection of the appropriate consensus group method and the definition of expert participants. Selection of one consensus method over another should be evident if the purpose is clearly stated. For example, if the purpose of the study is to develop international guidelines regarding ultrasound training for Rheumatologists (Pineda et al. 2010), it would be highly favorable to have input from a large number of geographically remote regions which would require a Delphi method.

Outline each step of the process: if modifications were made, provide a rationale for the choices made

As previously stated, the number of modifications and lack of standardization to the consensus group methods makes it essential that researchers clearly describe in sufficient detail the steps and processes used for each stage of the study. Simply stating a "modified Delphi" was performed with insufficient explanation of the actual process is, in our opinion, poor research reporting as it makes it difficult for a reader to judge the quality of the research. Providing justification for the choices made will also add credibility. One exemplary instance of clearly outlining the process and choices is a study that sought to clarify the minimum

necessary nursing ethics educational content and mastery level in Japan (Ito et al. 2011).

It should also be noted that consensus group methods are increasingly being considered one research method to be used within multi-method research projects (Harvey & Holmes 2012; Fletcher & Marchildon 2014; Bloor et al. 2015; Brady 2015; de Loe et al. 2016). As a result, a flow diagram or table to summarize the steps and participants at each stage adds clarity. For example researchers often begin with a local NGT to generate items that are later used in an international Delphi (CoBaTrICE Collaboration 2011).

Describe the selection and preparation of the scientific evidence for the participants

Research-based information should be provided to all participants early on in the process (Murphy et al. 1998). This information is best provided in a summarized form to facilitate review. Grading the quality of studies and information provided may reduce bias. Despite the original description by some authors that NGT does not require a search of the existing literature (Campbell et al. 2001), it is our opinion that this step is essential for a rigorous study. If there is no existing literature, then the authors must explicitly state the extent of the search. If different participants groups are used, the authors should clearly state what information was provided to each group.

Describe how items were selected for inclusion in the initial questionnaire: describe the process in sufficient detail

The original Delphi and NGT were described as beginning with an open-ended question to avoid biasing the participants, and is still suggested by some authors (Murphy et al. 1998; Sinha et al. 2011). Often, the items in the survey are generated based on a literature review and discussion with experts in the field. Researchers should make it clear which protocol – idea generation, providing a pre-determined list or a combination of both – was used in their study. The authors should describe, in sufficient detail, the methods used to develop these items and the process used for item selection. A useful example is the study by Ito where the authors clearly describe how the initial items were determined for their Delphi (Ito et al. 2011).

The initial question and items must be carefully worded to ensure clarity. Pilot testing with a small group of individuals is suggested before implementation. Another important aspect to consider is the length and complexity of the questionnaire because this may affect response rates (De Villiers 2005, p. 641)

Practically speaking, participants may be asked to respond to a statement by selecting "should definitely include, should definitely not include or undecided" or they may be required to rate their level of agreement with each statement using a Likert scale.

Describe how the participants were selected and their qualifications: if the NGT or RAND/UCLA is used, describe facilitator's credentials

Selecting participants for consensus group methods requires careful consideration. Most use "experts", but the definition of an expert can vary depending on the purpose of the study. Participants are usually individuals who are knowledgeable, representative of the area of inquiry and have practical experience. Studies have demonstrated that the composition of the panel can impact results and raises methodological concerns (Black et al., 1999; Campbell et al. 1999; Fitch et al. 2001; Keeney 2001; Hutchings & Raine 2006; Hutchings et al. 2006). Other studies have looked at diversity within a panel and suggest that diversity in the panel leads to better performance while in some instances heterogeneity may have an adverse effect if there is irreconcilable conflict between participants (Murphy et al. 1998).

Group size can vary tremendously depending on the purpose of the inquiry and the type of consensus group technique used. NGT typically uses 5–12 members, whereas Delphi can use 6 to thousands. Fewer than six participants in a Delphi is likely too few, whereas 12 or more is reasonable (Murphy et al. 1998). Arguably balance, or representation of multiple viewpoints and expertise, is more important than size (Bloor et al. 2015).

It is important to be aware that the Delphi requires a sustained commitment from participants who are questioned round after round on the same topic with a slightly modified questionnaire each time. Selection should consider choosing participants who will be committed to the task. One criticism is that participants willing to engage and continue in the process are more likely to be affected directly by the outcome and therefore are biased (Keeney 2001).

Whatever the makeup of the expert panel, the authors must provide a rationale and justify their choices. One group of authors has identified potential participants in health care research as the 9 "p"s: providers, professors, patients, payers, policy makers, private sector, public, purchasers and press (Boers et al. 2015). This can be easily transferred to education.

If using NGT or RAND/UCLA, the facilitator should be either a recognized expert in the field or a credible non-expert (Campbell et al. 2001). The facilitator should ensure that all viewpoints are equally discussed.

Describe the number of rounds planned and/or criteria for terminating the process

The currently recommended number of rounds is two to three; however, this is based on very little scientific evidence (Boulkedid et al. 2011). Attrition is a particular concern, with studies demonstrating significant decreasing response rates in successive rounds (Tammela 2013). Some authors feel strongly that the number of rounds should be determined *a priori* to avoid "false consensus", where participants become fatigued and agree just to end the process.

It has been suggested that authors should specify alternate stopping criteria such as terminating the Delphi process when a predetermined definition of agreement is reached, a maximum number of rounds had been completed or there is stability of responses (Diamond et al. 2014). Stability suggests that participant responses to each question across rounds is not changing and requires a

pre-determined definition such as a change in mean that is below a certain threshold (Vazquez-Ramos et al. 2007).

Clearly describe how consensus was defined

Some authors consider how consensus was defined as one of the most sensitive methodological issues and should occur before the process begins (Boulkedid et al. 2011; Diamond et al. 2014). A typical definition of agreement would be that 70% of respondents agreed or strongly agreed that an item should be included. Items where 70% disagreed or strongly disagreed should be discarded, and items not meeting these criteria would be sent to the next round for re-ranking. Others may decide a median score above a defined threshold is adequate (Tammela 2013). There is great variability in the definition, with a range of 51–80% (Hasson et al. 2000).

Report response rates and results after each round

There are two important aspects to consider with respect to the reporting of responses and results. The first is participant response rates. Response rates after each round should be reported while maintaining anonymity. The type of respondents should also be included, as it may have significant implications. For example, in an international Delphi, the percentage of respondents from each country would be relevant to report. The concern is that participants with minority opinions may be more likely to drop out, which may mean the degree of consensus reached in the final round is an overestimation (Sinha et al. 2011).

The second is the actual results after each round. Some have suggested that reporting each round separately gives a clear view of the themes generated and the strength of support for each round. For example, a study developing an instrument to assess a resident in subclavian central venous catheter insertion outlines the results for Delphi rounds 1 and 2 including which items were dropped or modified at each round (Huang et al. 2009)

Describe the type of feedback provided after each round

Feedback is considered an essential component of Delphi. Feedback to participants can include quantitative and/or qualitative data. It also involves two types of agreement: the extent to which individual participants agree with an issue, and the extent to which participants agree with one another. Quantitative feedback may include summary statistics such as the participants' score, participants' medians, range of scores and the proportion of participants selecting each point on a scale. Participants are provided an opportunity to change their ranking, but it should be made clear that they do not need to conform. Researchers may ask the participants who are outliers to provide written justification for their choices (qualitative data).

Some studies have demonstrated that the type of feedback provided, for example collective feedback from all participants or from a single group, may have a substantial influence on the outcome of the process (Campbell et al. 2001).

Describe how anonymity was maintained

One of the most important aspects of consensus group methods is the maintenance of anonymity. It is essential to avoid dominance by members who, for example, may be in a position of authority. Authors must clearly state how this was accomplished. It is achieved through the use of mail outs in Delphi and RAND/UCLA and private ranking in NGT.

Address potential methodological issues in the discussion

As previously noted, the lack of consensus of whether consensus group methods are "quantitative or qualitative" has led to confusion regarding how to assess validity, or credibility of these methods. Arguably, the first step is attention to the methodological issues such as those covered in this Guide. In the discussion the authors should address issues that may have impacted the results such as poor response rates between rounds, lack of participation from a select group or geographic region, or lack of consensus. When there is a lack of consensus, this often informs the future research agenda for a particular topic.

Conclusions

Consensus group methods are commonly used, but the considerable lack of standardization in definitions, use and reporting of these methods may hamper their effectiveness. Adequate attention to the issues discussed above will help researchers move away from the use of labels and toward a comprehensive description and justification of the steps taken in their study, ensuring that the results of the research are as credible and as useful as possible.

Disclosure statement

Notes on contributors

Susan Humphrey-Murto, MD, MEd, is an Associate Professor, Faculty of Medicine, University of Ottawa and Acting Director of the Medical Education Research Support Unit in the Department of Innovation in Medical Education (DIME).

Lara Varpio, PhD, is an Associate Professor in Medicine, and Associate Director of the Master's and PhD in Health Professions Education degree program at the Uniformed Services University for the Health Sciences, in Maryland, USA. Dr. Varpio specializes in medical communications and qualitative research.

Carol Gonsalves, MD, FRCPC, MMed, is an Assistant Professor, Faculty of Medicine, University of Ottawa and Clinician Investigator with the Ottawa Hospital Research Institute. Her medical education interests include needs assessment, curriculum development, and physician well-being.

Timothy J. Wood, PhD, is an Associate Professor in the Department of Innovation in Medical Education, University of Ottawa. He has also held positions as a PhD Researcher with the Academy for Innovation in Medical Education, University of Ottawa and Manager Research and Development, Medical Council of Canada.

References

Black N, Murphy M, Lamping D. 1999. Consensus development methods: a review of best practice in creating clinical guidelines. J Health Serv Res Policy. 4:236–248.

Bloor M, Sampson H, Baker S, Dahlgren K. 2015. Useful but no Oracle. Qual Res. 15:57–70.

Boers M, Kirwan J, Tugwell P, Beaton D, Bingham CO, Conaghan PG, D'Agostino MA, de Wit M, Gossec L, March L, et al. 2015. The OMERACT handbook. Omeract [cited 2016 Jan 15]. Available from: http://www.omeract.org/pdf/OMERACT_Handbook.pdf.

Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. 2011. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One. 6:e20476.

Brady SR. 2015. Utilizing and adapting the Delphi method for use in qualitative research. Int J Qual Methods. 14:1–6 doi:10.1177/1609406915621381.

Campbell SM, Cantrill JA, Hann M, Roland MO, Quayle JA, Shekelle PG. 2001. Consensus methods in prescribing research. J Clin Pharm Ther. 26:5–14.

Campbell SM, Hann M, Roland MO, Quayle JA, Shekelle PG. 1999. The effect of panel membership and feedback on ratings in a two-round Delphi survey: results of a randomized controlled trial. Med Care. 37:964–968.

CoBaTrICE Collaboration. 2011. International standards for programs of training in intensive care medicine in Europe. Intensive Care Med. 31:385–393.

Crisp J, Pelletier D, Duffield C, Adams A, Nagy S. 1997. The Delphi method? Nurs Res. 46:1–5.

de Loe RC, Melnychuk N, Murray D, Plummer R. 2016. Advancing the state of policy delphi practice: A systematic review evaluating methodological evolution, innovation, and opportunities. Technol Forecast Soc. 104:78–89.

De Villiers MJR. 2005. The Delphi Technique in health sciences education research. Med Teach. 27:639–643.

Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, Wales PW. 2014. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol. 67:401–409.

Fitch K, Bernstein SJ, Aguilar MD, Burnand B, LaCalle JR, Lazaro P, van het Loo M, McDonnell J, Vader J, Kahan JP. 2001. The RAND/UCLA appropriateness method user's manual. Santa Monica, CA: RAND Corporation.

Fletcher AJ, Marchildon GP. 2014. Using the Delphi method for qualitative, participatory action research in health leadership. Int J Qual Methods. 13:1–18.

Goodman CM. 1987. The Delphi technique: a critique. J Adv Nurs. 12:729–734.

Guzys D. 2015. Gadamerian philosophical hermeneutics as a useful methodologic framework for the Delphi technique. Int J Qual Stud Health Well-Being. 10:26291.

Hasson F, Keeney S, McKenna H. 2000. Research guidelines for the Delphi survey technique. J Adv Nurs. 32:1008–1015.

Harvey N, Holmes CA. 2012. Nominal group technique: an effective method for obtaining group consensus. Int J Nurs Practice. 18:188–194.

Huang G, Newman LR, Schwartzstein RM, Clardy PF, Feller-Kopman D, Irish JT, Smith CC. 2009. Procedural competence in internal medicine residents: validity of a Central venous catheter insertion assessment Instrument. Acad Med. 84:1127–1134.

Humphrey-Murto S, Ufholz L, Gonsalves C, Varpio L, Lynch S, Wood TJ. 2014. The use of consensus group methods in medical education research. Paper presented at: Oral presentation, The Ottawa Conference, Ottawa, 2014 April 26–30.

Hutchings A, Raine R. 2006. A systematic review of factors affecting the judgments produced by formal consensus development methods in health care. J Health Serv Res Policy. 11:172–179.

Hutchings A, Raine R, Sanderson C, Black N. 2006. A comparison of formal consensus methods used for developing clinical guidelines. J Health Serv Res Policy. 11:218–224.

Ito C, Ota K, Marsuda M. 2011. Educational content in nurse education in Japan: a Delphi study. Nurs Ethics. 18:441–454.

Jones J, Hunter D. 1995. Consensus methods for medical and health services research. BMJ. 311:376–380.

Keeney S. 2001. A critical review of the Delphi technique as a research methodology for nursing. Int J Nurs Stud. 38:195–200.

Lazarou J, Hopyanm J, Panisko D, Tai P. 2011. Neurology for internal medicine residents: working towards a national Canadian curriculum consensus. Med Teach. 33:e65–e68.

Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, et al. 1998. Consensus development methods and their use in clinical guideline development. Health Technol Assess. 2:1–88.

Pineda C, Reginato AM, Flores V, Aliste M, Alva M, the Pan-American League of Associations for Rheumatology (PANLAR) Ultrasound Study Group, et al. 2010. Pan-American League of Associations for Rheumatology (PANLAR) recommendations and guidelines for musculoskeletal ultrasound training in the Americas for rheumatologists. J Clin Rheumatol. 16:113–118.

Sackman H. 1975. A delphi critique. Lexington, MA: Lexington Books.

Sinha IP, Smyth RL, Williamson PR. 2011. 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 Med. 8:e1000393.

Stewart J. 2001. Is the Delphi technique a qualitative method? Med Educ. 35:922–923.

Tammela O. 2013. Applications of consensus methods in the improvement of care of paediatric patients: a step forward from a 'good guess'. Acta Paediatr. 102:111–115.

Van der Vleuten C. 2014. Medical education research: a vibrant community of research and education practice. Med Educ. 48:761–767.

Vazquez-Ramos R, Leahy M, Hernandez NE. 2007. The Delphi method in rehabilitation counseling research. RCB. 50:111–118.

Vernon W. 2009. The Delphi technique: a review. Int J Ther Rehabil. 16:69–76.

Waggoner J, Carline JD, Durning SJ. 2016. Is there a consensus on consensus methodology? Descriptions and recommendations for future consensus research. Acad Med (Perspective). 91:663–668.