Participant expectations and experiences of using the nominal group technique (NGT) in a NICE healthcare guideline: a pilot study

Victoria Katrina Roberts

D. Clin. Psy. Thesis (Volume 1), 2019

University College London

UCL Doctorate in Clinical Psychology

Thesis declaration form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Name: Victoria Katrina Roberts

Date: 18th July 2019

Overview

Part one offers a conceptual introduction of formal consensus methods. This is in the form of a review of the evidence concerning participant experiences of formal consensus methods. Three databases were searched and 10 papers were included in the review. Themes that were identified across the papers included increased participation, discussion as a helpful process, improved quantity and quality of ideas, and finding the consensus process interesting and enjoyable. The conceptual introduction concludes with recommendations for the practice of formal consensus methods when applied to complex problems such as when developing guidelines, including using a hybrid of methods to enhance their respectively identified strengths.

Part two presents an empirical paper of the present pilot mixed methods study that was based at the National Guideline Alliance (NGA), which creates healthcare guidelines for the National Institute for Health and Care Excellence (NICE). The project aimed to report on participant expectations and experiences of using a formal consensus method called the Nominal Group Technique (NGT). This was done initially through 12 qualitative interviews with healthcare guideline committee members and their associated technical team prior to the NGT session. Following this, participants were asked to complete a Likert quantitative questionnaire immediately after the NGT session to capture their experiences of it. Themes that were identified in the interviews largely agreed with the questionnaire data. The main themes were separated into four groups. Firstly, there were themes about the formal consensus method in terms of credibility of the method and effort and resource intensiveness of using the method. Secondly, a group of themes labelled methodology included guideline interpretation and implementation, the interpretation

of evidence, and the restrictiveness of the NICE process. Thirdly, a theme of group processes included the management of expertise, anonymity, leadership, and the benefits of discussion. Fourthly, a theme regarding the continuity of group members was considered important in the data. The data are corroborated to form recommendations for the use of NGT in healthcare guidelines as well as areas for future research.

Part three comprises of a critical appraisal of the research process. It illustrates some of the methodological considerations, contextual factors of working in a governmental environment, and personal reflections on the research including potential biases that could have impacted the process.

Impact Statement

This was the first study that collected qualitative views from healthcare guideline committee participants using formal consensus methods at the National institute for Health and Care excellence (NICE). It was designed to contribute to the development of "evidence-based methodologies" for consensus in healthcare guidelines. This is with consideration to the literature that suggests formal consensus method effectiveness is dependent on the context in which it is applied. Thus, research into the experiences of users of formal consensus methods in healthcare requires specific research that is specific to the particular governmental setting, participants and task.

NICE guideline creation is influenced by external pressures that require the efficient creation of documents with national implications. Under these conditions,

there is a drive to streamline procedures. However, with this tendency there is also the potential for the effectiveness of NGT to be compromised. This is particularly because not much is known about the essential components of the NGT method. The present paper offers a structure and method for the assessment of formal consensus methods including NGT to support its evaluation so that it remains a helpful method for guideline development. The use of a mixed methods design proved to be a helpful way of capturing views on formal consensus methods. Study participation was high, which suggests that it was successful as a method of evaluation that was acceptable to candidates. The present study has led to the development of feedback questionnaires that can be routinely used in guideline groups for improvement in NGT implementation.

Guideline committees are tasked with applying multidisciplinary knowledge to complex questions across a range of healthcare settings. This research is the first to investigate how the professional background of participants might impact their perceptions of benefits of NGT and on group decision-making in general. This not only contributes to a better understanding of the experience of group-decision making in committees, but also how NGT might support the needs of individuals in committees of different professional backgrounds. These differences are likely to be accentuated in the context of healthcare committees that depend on the effective exchange of multidisciplinary information.

Many important questions in healthcare guidelines have only low quality or very limited evidence. Informal consensus methods do not offer a systematic way to make use of data taken from sources other than the evidence-base, which includes evidence from non-randomised controlled trials or from experiential clinical

expertise. The present paper offers preliminary data on NGT as a potentially viable option for integrating evidence from multiple sources in a transparent way nu investigating the perceptions of committee members and the associated technical team. It is of particular importance for areas where recommendations are needed but only low quality evidence is available.

Table of Contents

List of Appendices

List of Tables

Part 1: Conceptual Introduction

Part 2: Empirical Paper

List of Figures

Part 1: Conceptual Introduction

Part 2: Empirical Paper

Acknowledgements

I would firstly like to thank my internal supervisor Professor Stephen Pilling. Thank you for your insight and support, I really appreciate all the time that you gave me. Also thanks to my external supervisor Dr Patrice Carter, for being an essential link between University College London and the National Guideline Alliance (NGA). I am particularly thankful also for the support of Phoebe Barnett, who helped transcribe and analyse the data, and often offered thoughtful contributions to the project.

I would like to thank all of the staff at the NGA who welcomed me as part of their team. I could not have carried out the research without their support and openness to collaborate. In particular, Jen Francis and Rami Cosulich for their time and patience when training me on how to implement a systematic review from the beginning.

I also appreciated the generosity and interest taken by the participants in the project. Their enthusiasm really inspired me to delve ever deeper into formal consensus methods as situated within healthcare guidelines. These participants gave up their time without compensation and often offered interesting, insightful, and honest opinions.

Last but not least, I would like to extend a thank you to my friends and family. Your love and support has humbled me throughout my clinical psychology journey, and I could not have made it this far without you. In particular I would like to thank Julia, Michael, Mum, and Dad.

Part 1: Conceptual introduction

A review of the literature concerning participant experiences of using formal consensus methods with a focus on healthcare guidelines

Abstract

Aims: The nominal group technique (NGT) is a formal consensus method to support decision-making that is most appropriate when evidence is limited or of low quality, which can be the case when developing healthcare guidelines. This manuscript aims to review participants' experience of developing healthcare guidelines using formal consensus methods, including NGT.

Method: Three databases (SCOPUS, PsycINFO, and PubMed), along with forwards and backwards reference search, were searched to derive peer-reviewed articles that focused on participant experiences of formal consensus methods. Inclusion criteria were that studies measured participant attitudes in terms of experience, and that they applied decision-making to "complex" questions, defined as requiring a pooling of unshared experience by the group to a problem with no single "correct" answer.

Results: The database and reference search returned a total of 5845 hits, of which a final 10 were deemed relevant for inclusion. These studies implemented formal consensus methods in a variety of different ways in terms of the method used, the time taken, and the amount of training given and the problems used.

Conclusions: There were considerable similarities in the themes identified across studies. The themes most valued by participants were increased participation, discussion as a helpful process, better quantity and quality of ideas, and finding the process interesting and enjoyable. These themes appeared to have complex interactions and impacts on participants, and these are discussed.

Introduction

This is a mixed methods research project to evaluate the experiences of committee members and technical staff on healthcare guidelines when using a formal consensus method. More specifically, the project evaluates the Nominal Group Technique (NGT; Van De Ven & Delbecq, 1971) as used in policy guidelines by the National Guideline Alliance (NGA) in developing guidelines for the National Institute for Health and Care Excellence (NICE). This thesis takes a macro-level view of the clinical psychologist, as is defined in the domain of community psychology (e.g. Moritsugu, Vera, Wong, & Duffy, 2015). More specifically, it follows a broader view of the clinical psychologist as engaging in policy and social change as well as working with individual clients and the teams around them. The NGA has increased the use of NGT in recent months, but the process has not been evaluated in terms of participant experience .This is particularly important because the scope of NGT has been broadened by the NGA to include circumstances when the review question is "too broad" and encompasses too much evidence. This is in contrast to how NGT is typically used in the research literature, as a method for when there is insufficient or contradictory evidence. Therefore, it is a particularly relevant time to review the feasibility and utility of NGT from the perspective of participants.

NICE develops guidance for the National Health Service, and is a non-departmental public body established by statute. NICE faces the major challenge of creating healthcare guidelines when the area is often characterised by a lack of good quality evidence (Ketola, Kaila, & Honkanen, 2007). The NGT is a method by which expertise can be deployed to develop recommendations despite a lack of good

evidence and is currently used by NICE. It has been recommended for areas where there is limited or contradictory evidence because it allows for the inclusion of expert opinion and a broader range of types of evidence than traditional informal consensus methods (Campbell & Cantrill, 2001; Jones & Hunter, 1995a). It has been widely used in healthcare context such as for identifying barriers to screening for colorectal cancer (Bajracharya, 2006), prioritising interventions for diabetes (Chasens & Olshansky, 2008), or identifying the training needs of community nurses (Carney, McIntosh, & Worth, 1996).

Despite the use of NGT and other formal consensus methods in guidelines, little is known about the experiences of those who use them. This is important because the perception of formal consensus methods is likely to affect their use and uptake by committees (Bini & Mahajan, 2016). Moreover, understanding and evaluating participant views can lead to the improvement of methods and consequently improved decision-making in committees.

The aims of this conceptual review are twofold. Firstly, it seeks to provide a narrative synthesis of the literature investigating the experiences of participants using formal consensus methods. Secondly, it hopes to evaluate whether formal consensus methods are feasible and relevant for decision-making in healthcare guidelines.

Group decision-making and informal consensus methods

Informal consensus methods can be defined as unstructured groups using free discussion with the aim of coming to a general agreement on a topic, for example as in a jury (Murphy et al., 1998). In groups, decision-making via informal consensus has been found to be subject to a variety of process influences, which have been labelled by some authors as "process losses" (Steiner, 1972). For example, group members might be less likely to present a variety of viewpoints due to the observed tendency of participants to agree with the majority view or one's perceived peers, labelled as an affect of "conformity" or "compliance" (e.g. Woudenberg, 1991). Indeed, group members are more likely to present information that is shared with the rest of the group rather than non-shared information that might support a different view (Winquist & Larson, 1998; Wittenbaum & Stasser, 1996). Individuals have been shown to have a preference for both receiving and giving shared versus unshared information, and that those giving shared information are perceived as more credible, proficient, and knowledgeable by other members of the group (Wittenbaum, Hubbel,& Zuckerman, 1999).

When decisions are cognitively demanding, such as is the case for when considering scientific evidence, the personal characteristics of group members may also contribute to the weighting of opinions. Emphasis tends to be placed on individuals who have greater perceived knowledge, credibility, trustworthiness, or confidence (e.g. Cramer, Brodsky & DeCoster, 2009; Neal, Guadagno, Eno, & Brodsky, 2012). Group methods that increase the participation of all individuals therefore could serve to reduce the biased sharing of opinions.

The above characteristics of group processes could be heightened by the

context of healthcare guidelines, as has been suggested by a number of authors(e.g. Pagliari, Grimshaw, & Eccles, 2001; Rycroft-Malone, 2001; Shekelle, Woolf, Eccles, & Grimshaw, 1999). One source of influence is the application of healthcare evidence as requiring interpretation by experts. This process involves the appraisal of multiple variables, including how the evidence may apply to different patient populations (Woolf, Grol, Hutchinson, Eccles, & Grimshaw, 1999), the potential costs and benefits of a procedure (Fitzgerald, Ferlie, & Hawkins, 2003), and local service funding and policy priorities that might impact implementation (Schünemann, Fretheim, & Oxman, 2006). Furthermore, high quality research is scarce in some areas of healthcare, and where this is the case the systematic reviews that are presented to expert panels may be comprised of low quality evidence (Chassin, 1989; Gill et al., 2011). Despite these potential limitations, the presentation of literature reviews has been shown to facilitate group agreement compared to the absence of a literature review (Raine et al., 2004).

A second source of influence is interpersonal guideline group pressures and social dynamics. The impact of powerful or dominant individuals in driving group decision-making is detectable even in groups that have no significant status differences or prior history of working together (e.g. Kramer, Kuo, & Dailey, 1997). Thus, encouraging participation through the pooling of privately generated views could be particularly important for effective group representation. These effects are likely to be increased in healthcare settings that function in a hierarchical way between medical staff, lay members, and non-medical committee members.

Related to the inclination to share similar information is the effect of group homogeneity, since group conformity is increased in groups that are perceived to be

more similar (Asch, 1952). Group agreement increases if groups are comprised of a single profession, and thus alternative explanations are not as fully explored (Murphy et al., 1998; Raine et al., 2004). Difficulties with homogenous group decision-making have been discussed elsewhere (Bang & Frith, 2017). These challenges cannot simply be overcome through increasing participation or improving group processes, since homogenous groups by definition have access to similar experiences, information, and potentially consequent biases. Good practice as defined by NICE requires guideline groups to be heterogeneous in terms of representing a representative range of professionals, service user and carers (National Institute for Health and Care Excellence, 2014). Therefore, there are likely to be a range of viewpoints drawing from different sources of evidence.

A third source of bias could be individual decision-making biases. Guideline committee members tend to interpret evidence in light of already held beliefs and assumptions (Raine et al., 2004), be biased towards theories and procedures associated with their own knowledge or experience (Chassin, 1989; Fitzgerald et al., 2003), and generalise evidence so as to overstretch the utility of interventions to other conditions (Kahan et al., 1996; Murphy et al., 1998). Indeed, groups can and do disagree on their fundamental interpretation of the same evidence (Hemingway et al., 2008).

In summary, the application of formal consensus methods to healthcare guidelines requires careful consideration. This is because the process of applying evidence to the development of healthcare guidelines requires the mobilisation of a particular set of skills as applied to imperfect evidence by group committee members from heterogeneous and hierarchically organised professional backgrounds.

Increasing the transparency of these processes through the use of formal consensus could support groups to use expertise and evidence in a more conscious, clear, and informed way. Formal consensus methods aim to support participation equality through the pooling of individual views, which is helpful particularly because of the tendency for groups to avoid discussing unshared information.

Unanimous agreement on the implications of contested and limited evidence is unlikely. What is more achievable is a consensus process that people can agree offers "accountability for reasonableness" (Daniels & Sabin, 2008). This means that decision–making procedures should follow a consistent and transparent process that is based on reasoned agreement. This way its contents can be debated by the public and become subjected to regulation and review. NGT has been suggested as a useful method for areas of limited or low-quality evidence since it allows for the incorporation of different types of data, including experiential and lower quality evidence (Black et al., 1999).

NICE asserts the importance of incorporating clinical expertise and service user and carer experiences when making guidelines to encourage a meaningful translation of evidence to practice guidelines (National Institute for Health and Care Excellence, 2014). The development of guidelines currently involves face-to-face meetings of experts and lay-members, which is a setting supportive of the use of NGT. One of the advantages of NGT in this setting is that it allows for efficient use of participant time. Increasing the transparency of decision-making processes may also enable inadvertent participant biases to be delineated from deliberately contributed expertise. This can be achieved through a process of recording the statement revisions and where they were sourced in a progressive way.

Healthcare guidelines are often based on low quality evidence from multiple sources. Tensions could arise from differences between the experiences of service users, the professional experience of clinicians, and the generalised evidence base. Guidance on how to manage these multiple perspectives is unclear, and often weighted towards prioritising the research evidence(Rycroft-Malone et al., 2004). In order for those who implement guidelines to be informed consumers, it is important that the translation of evidence to guidelines is transparently reported. This separates guideline interpretation from guideline production, and encourages the critical appraisal and evaluation of implementation and development of guidelines.

Formal consensus methods. Formal consensus methods offer a structured, systematic approach. They therefore attempt to mitigate some of the group pressures that contribute to ineffective decision-making (Murphy et al., 1998; Pagliari et al., 2001). There are several methods of formal consensus. The formal consensus methods that are most commonly used in healthcare are Delphi, RAND-UCLA Appropriateness Method (RAND), and the Nominal Group Technique (NGT).

Delphi. The Delphi method was initially developed to predict the impact of technology of warfare during the Cold War by the RAND corporation (Rescher, 1998; Yousuf, 2007). To use the Delphi method, a facilitator recruits members who have some expertise on a topic to a group. Next, the facilitator generates a set of statements that the experts rate for agreement. The responses from the members are gathered by the facilitator, and the facilitator gives anonymous individual feedback on how the group member's responses compare to the rest of the group. Following this, the group members are given the option to revise their responses. Statistical criteria are used to define consensus, and the responses converge across rounds of

questionnaires.

The Delphi technique has been evaluated by many authors (e.g. Bolger & Wright, 2011). It allows for the efficient collection of views from a large group of participants. The ability for these views to be collected via postal mail or the internet increases the accessibility of Delphi since it is without geographical constraint. However, a disadvantage of participation without face-to-face interaction is that participants are unable to directly resolve discrepancies through the discussion of ambiguities or uncertainties. The Delphi method has been used in a number of studies to develop clinical practice including guidelines for CBT for pain (Mobily, Herr, & Kelley, 1993), and preventative therapy to support the use of an antibiotic (Passannante, Restifo, & Reichman, 1993).

NGT. The NGT is a structured process involving a group of relevant expert members to gather and assess information on a given topic (Murphy et al., 1998). The procedure of NGT as described in the original paper (Van de & Delbecq, 1974) and is as follows. Firstly, individuals privately record their ideas in response to the topic. Secondly, responses are presented in turn either by the participants or the facilitator. Thirdly, ideas are each discussed for clarification and evaluation. Fourthly, individuals rate the ideas individually. Fifthly, the pooled group vote is presented for discussion by the group. Sixthly, this is followed by another round of individual rating. The final group outcome is an aggregate of the ratings.

*UCLA/RAND Appropriateness method (RAND).*This approach was developed in order to assess the uptake of medical and surgical procedures (Fitch et al., 2001). It uses two separate groups- a multi-disciplinary "expert" panel and a separate "core" panel. The core panel consist of technical team members. In the initial phase, the

core panel are responsible for providing information to the expert panel in the form of a synthesis of relevant data (i.e. a literature review) and also scenarios or statements describing a patient with particular characteristics. The expert panel are then tasked with rating the benefit-to-harm ratio of each procedure in the scenarios. In the next phase of RAND, the experts meet in person where their collated results are disseminated and discussed. This is followed by private rating of each statement or scenario. The RAND method is not intended for the creative generation of ideas, as it assumes data is available to compile the scenarios and statements (Fitch et al., 2001).

Formal consensus methods are well suited to instances when there is limited evidence as can be the case in healthcare research. This is because formal consensus methods are designed to better manage group processes and optimise group decision-making when integrating information from multiple sources. Previous research has assessed the effectiveness of formal consensus methods by measuring accuracy and superficial elements such as the quantity and quality of the contributions or ideas (e.g. Graefe & Armstrong, 2011; Hutchings, Raine, Sanderson, & Black, 2006). This might not be appropriate when applied to decision-making in healthcare guidelines. Quality healthcare guidelines require the interpretation and application of imperfect information to complex problems, and there may be more than one "correct" answer. Thus, one important measure of effectiveness might be assessed by the experiences of users of formal consensus methods. Understanding participant experiences allows us to identify their perceptions of acceptability, which may in turn increase their application and uptake (Bini & Mahajan, 2016). A critical realist perspective emphasises the importance of triangulation, which could be particularly important in the early developments of an area such as the individual experiences of participants

using formal consensus methods (Barker, Pistrang, & Elliott, 2002). Thus, a mixed methods approach was adopted in order to capture the experiences of participants.

Narrative synthesis of literature- Method and Results

The research question for this narrative synthesis was "What are participant experiences of the use of formal consensus methods in healthcare guidelines?" The aim of this search was, therefore, to identify studies reporting on a range of consensus methods. Three databases (SCOPUS, PsycINFO, and PubMed) were used for the search. The search strategy for bibliographic databases is illustrated below with a contextual search narrative to explain the decision-making at each stage of the literature search strategy (Cooper et al., 2018).

Data extraction

The search was initially not limited to a particular population in order to preserve search sensitivity and to be representative of the broad range of topic areas that engage formal consensus methods. However, studies were excluded at the full text stage for several reasons.

Firstly, included studies focused on the perspectives of participants rather than those of the researchers. The perspective of researchers could be considered distinct to that of participants. This is because their involvement in formal consensus methods is likely to bear focus on the creation, analysis, and synthesis of statements rather than the group discussion, group process, and decision-making. Therefore, research that captured the experiences or reflections of researchers was removed since it was beyond the scope of this review.

Secondly, studies were removed if they were not in the domains of health or

policy development. Examples of excluded studies were those that required groups to problem-solve a NASA moon expedition survival scenario or to estimate the percentage of the population over 65 years old.

Thirdly, studies were excluded if the decision-making process did not require a pooling of unshared expertise from participant. This is due to the importance of balancing shared versus non-shared information during committee decision-making. For example, studies that exclusively involved student populations and gave participants identical information to solve an unfamiliar problem were excluded.

Fourthly, findings were excluded if the consensus method used was not the focus of the study, for example if instead leadership style was manipulated. The intervention was defined as the method of consensus (e.g. Delphi, NGT, or RAND) and the outcome was considered the experiences of participants as measured by qualitative or quantitative means.

The measurement of "experiences" was defined as participant views towards consensus methods as described by their cognitions, emotions, and behaviour. Studies were included if they measured an individual's subjective and personal perceptions relating to the use of consensus methods. This could be achieved through direct means (e.g. self-report interviews or questionnaires), or indirect means (e.g. observation of consensus groups). Previous work has included Likert-scales and open questions that captured measures of enjoyment, appropriateness to the task, felt time efficiency, contribution, and difficulty (e.g. Boje & Murnighan, 1982; Graefe & Armstrong, 2011). If alternative methods of capturing experiences were noted when conducting the search, these were captured and the definition refined. The searches were limited to English papers in the subject areas of social sciences, psychology,

medicine, business, and decision-making.

Table 1: Table to illustrate the search strategy

Search syntax	Contextual narrative
1. formal consensus[Title] OR NGT[Title] OR nominal group technique[Title] OR Delphi [Title] OR RAND[Title] OR consensus methods[Title] OR consensus[Title] OR group decision*[Title] 2. experience*[Title] OR understand*[Title] OR observation*[Title] OR perception*[Title] OR view*[Title] OR impression*[Title] OR thought*[Title] OR attitude*[Title] OR judgement*[Title] OR evaluat*[Title] 3. guideline*[Title] OR healthcare[Title] 4. (LIMIT-TO (SUBJAREA , "MEDI") OR LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "PSYC") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "DECI")) AND (LIMIT-TO (LANGUAGE , "English")) 5. 1, 2 and 3 6. 1, 2 and 4	5. This search only returned 96 results in SCOPUS, and 37 results in PsycINFO. 6. This search did not limit to titles that included “guidelines” or “healthcare” but instead used subject areas of Medicine, social sciences, psychology, business, and decision-making as limiters. Search was also limited to papers in the English language.

A non-database search method was used in addition to the main search to identify relevant studies (Cooper, Booth, Britt& Garside, 2017). This involved forwards and backwards citation chasing of references (Levay, Ainsworth, Kettle, & Morgan, 2016). The following Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) diagram illustrates the sifting process.

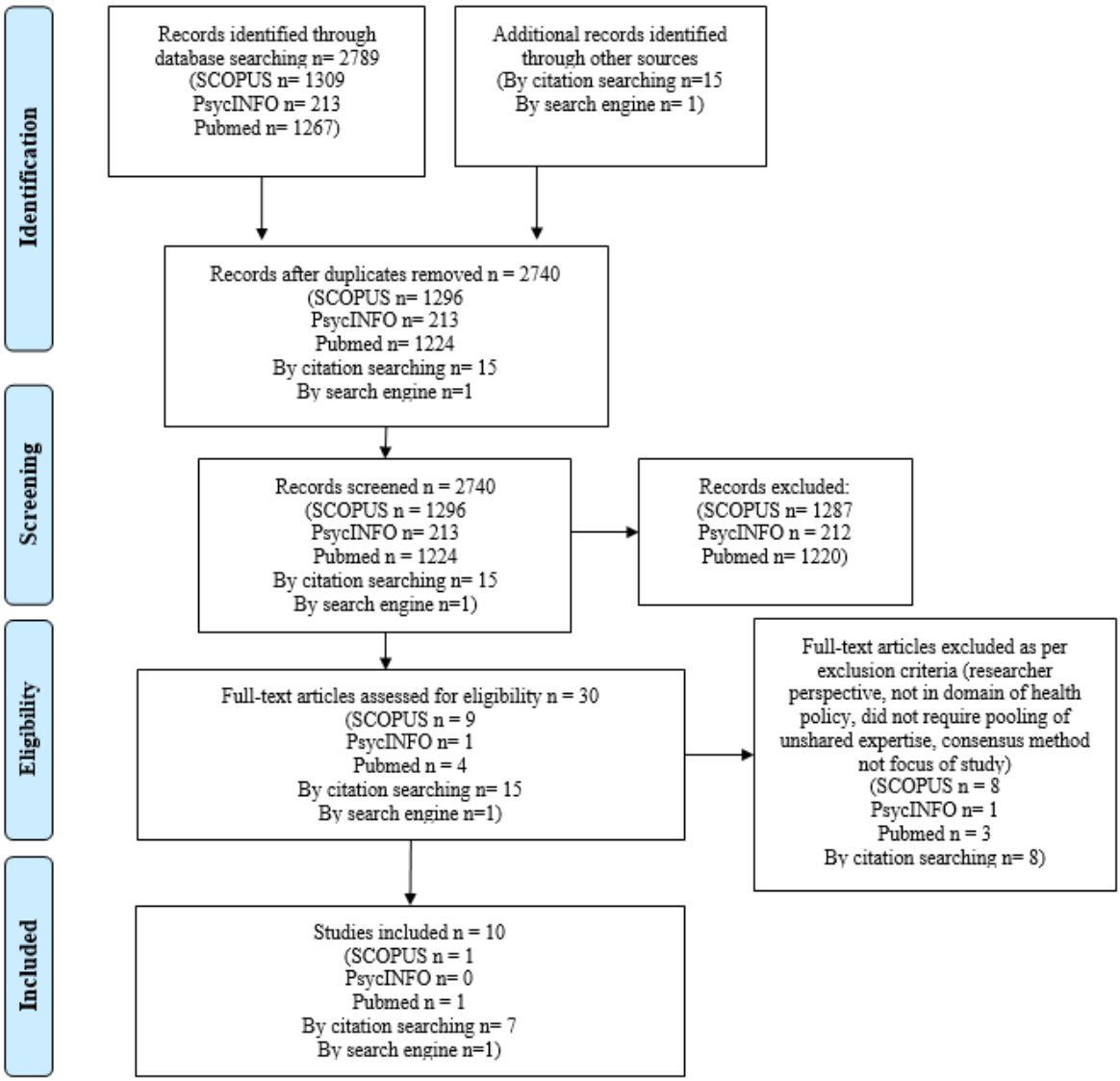


Figure 1: Flowchart of study selection process, PRISMA diagram (Moher, Liberati, Tetzlaff, & Altman, 2009)

Characteristics of studies

A total of 10 studies were included in this review (see Table 2 on pg. 34). Studies used a mixture of qualitative and quantitative data. The studies were highly heterogeneous from the perspective of analysis, design, and procedures. Comparing heterogeneous findings favours qualitative rather than quantitative synthesis because it leaves the data as situated in its context (Thomas & Harden, 2008). Therefore, priority was given to qualitative findings as a framework with which to synthesise the mixed methods approach, and these were analysed and are presented first.

The studies occurred in a range of settings involving meaningful problem-solving, including policy conferences (Gresham, 1986; Stephenson, Michaelsen, & Franklin, 1982) and university committees (Kramer et al., 1997; Van de Ven & Delbecq, 1974). Only two of the studies shared the percentage prevalence of participant endorsement of the reported qualitative themes within their data (Stephenson et al., 1982; Van de Ven & Delbecq, 1974).

The quantitative studies that investigated the formal consensus methods are challenging to compare due to the variation in methods and output (see Table 2). Studies within each type of formal consensus method ranged in terms of sample size (between three to 10 per NGT group and seven to 15 per Delphi group, although some articles did not specify group size), length of delivery (from one hour to across two days for NGT, and from a few weeks to several months for Delphi), and the way that outcomes were reported. Studies were similar by generally employing Likert-scales. However, comparison between the studies is limited since there was variation in the range of possible responses on the Likert scales. To manage this, studies that reported statistical significance were included in the analysis in the context of

supporting the qualitative data.

Three of the quantitative studies averaged participant responses for each question on a scale ranging from negative to positive with the central figure indicating neutral. Studies reported these findings as a percentage rather than statistical comparison (Bini & Mahajan, 2016; Landeta, Barrutia, & Lertxundi, 2011; Mcdougal, Brooks, & Albanese, 2005). Since it is challenging to compare across Likert scales, these findings shall be used descriptively alongside the other results.

Data analysis

Themes from each of the qualitative studies were analysed using thematic analysis (Braun & Clarke, 2006) and a "critical realist" epistemology (Fletcher, 2017). This states that the construction of individual experiences is achieved both through their own experiences of formal consensus methods and their perceptions of how they are viewed on a wider level by society and the general public. In this way, their experiences are understood as "real" and situated in their socio-cultural context (Thomas & Harden,2008).

Initially, the researcher (VR) repeatedly read the data to familiarise herself with it. Data was inputted into NVIVO qualitative analysis software(QSR International Pty Ltd., 2018). The researcher then began coding units of meaning remaining close to the data, rather than ascribing any higher-order categorisation (see Appendix A and B for an example of the early stages of coding). Following this stage, codes were linked to develop over-arching themes. The threshold for when a collection of codes were considered sufficient to be assigned a theme was discussed in supervision and with the other researcher (PB). It was agreed that themes would be expected to be described by at least two participants, or would be explicitly

emphasised verbally by a single participant. The researcher adopted a data-driven, inductive approach. However, the data was viewed through the lens of the research question, which also informed the generation of themes and inclusion of data. Themes were compared to the data and revised in an iterative fashion, which was checked for consistency and transparency by another researcher (PB).

Variation in the reporting of data of qualitative studies is relatively common (Sandelowski & Barroso, 2002). Indeed, some of the studies did not include actual participant quotations but only presented the themes they had drawn out of their data as the results. To resolve this problem, all text under the headings of "results" or "findings" was included as part of the results produced by this review, as has been previously suggested by (Thomas & Harden, 2008).

Thematic analysis developed an initial total of 53 NGT and 24 Delphi codes or units of meaning. Following iteration and discussion with other researchers, these were organised into higher-order themes. For example, the code of "anonymity" in Delphi was grouped under a general theme of participation since participants referred to it in the context of encouraging participation. Participation was grouped under an over-arching theme of "group processes", which was defined as factors relating to interpersonal interactions within the group (Brown, 2001). This was separate from a theme of methodological difficulties or intrapersonal positive feelings as a reaction to formal consensus methods.

The qualitative themes were grouped into "positive" and "negative" participant feedback. Qualitative data themes were reduced into seven negative and five positive Delphi themes, and ten negative and nine positive NGT themes. The quantitative research data produced two themes for Delphi, three themes for NGT,

and three themes for RAND. The RAND data produced no additional themes to the NGT and Delphi data. Instead, its themes were encompassed by those drawn from the NGT and Delphi research. Please see the appendix A and B for a more detailed illustration of this early stage of themes.

Table 2: Overview of included studies

First Author	Code	Year	Formal Consensus Method	Aim	N	Decision-Making Task	Results Captured	Data Analysis
Bini	**B**	2016	NGT	To report on the adoption rate of clinical practice guidelines created and implemented by a large orthopaedic group using the Delphi consensus method	200 Orthopaedic surgeons 85 participants divided into 5 groups (size unspecified)	15 clinical guidelines on the management of orthopaedic trauma	Self-report using online survey on 5 point Likert scale	Descriptive Statistics from questionnaire
Van de Ven	**V**	1974	NGT, Delphi, and Informal consensus	To compare informal consensus, nominal, and Delphi groups in idea quantity and participant satisfaction	420 student residents, housing, administrators, faculty academic administrators. Assigned to 20 NGT, 20 Delphi and 20 informal consensus groups	Defining the job description of part-time dormitory counsellors.	Self-report on 5 Point paper Likert Scale and also qualitative data from open ended question	Thematic analysis of qualitative data ANOVA to compare Likert scales between groups with post-hoc tests comparing groups
Stephenson	**St**	1982	NGT	To evaluate NGT in a naturalistic setting particularly using participants from heterogeneous backgrounds and a complex problem	45 Energy industries, energy researchers, news media, consumers, federal government, lawyers, architects, manufacturers, management experts, utility companies, and builders. Does not specify individual group size although describes groups as "small"	Developing a solar energy plan	Self-report on 6 point Likert scale and 2 open ended questions and	Thematic analysis of qualitative data Comparison of hypothetical mean of neutral response from informal consensus conference and experimental response to NGT method to yield statistical results
De Ruyter	**Ru**	1996	NGT, informal consensus	To present NGT as an alternative to focus groups for idea generation in market research	44 railway passengers, 3 NGT groups and 3 focus groups. 5, 7 , or 10 participants n each group	"Which services should the Dutch Railways offer in order to enhance its service quality?"	Self-report on 7 point Likert scale	Number of ideas analysed between groups using Chi-squared for parametric and Mann-Whitney for non-parametric tests

Gresham	**G**	1986	NGT	To determine whether participants would be satisfied with NGT across a range of problem solving and decision-making conferences	206 total participants. 178 at an International conference on food and water , 60 at a Texas agricultural extension service district conference, 9 at an English Language Institute Curriculum conference, and 67 at a Principal's Centre Academy	Food and water conference- Priority of issues in water and food policy Agricultural extension conference- Identification of problem areas in working with adults English language curriculum conference- Determine the content of a new course for the English language institute Principles centre summer academy- exploring ideas and assumptions arising in lectures and problem solve issues raised	Self-report on 5 Point Likert scale and open ended question to elicit general feedback.	Descriptive statistics and use of quotes to support attitudes
Kramer	**K**	1997	Informal, brainstorming, & NGT	Examine the impact of brainstorming and NGT on group processes and successive decision-making	200 students and staff divided up into groups of 5.	Developing a 2 hour Saturday afternoon program for 200 high school juniors and seniors.	Self-report on 5 point Likert scale and open ended question, observer rating of decision quality on 4 point Likert scale	Qualitative data analysed using content analysis Groups compared on Likert scales using ANOVA
West	**W**	2011	Delphi	To evaluate the costs and benefits of the Delphi methodology from the researcher and participant's perspective	7 psychotherapist specialists in trauma	"What are the supervision issues and needs that require consideration when supervising a counsellor/psychotherapist who is working with adult trauma?"	Written open ended feedback on process	
Shekelle	**Sh**	1996	RAND appropriateness	To assess the feasibility of using the RAND/appropriateness method for developing clinical guidelines in health care policy and research	27 members of lower back problems panel comprised of physicians and healthcare personnel	Healthcare scenarios requiring participants to make a judgement on how appropriate a particular treatment is	Self-report 5 point Likert scale	Calculated mean with 95% CI to derive statistical significance
McDougal	**M**	2005	Hybrid NGT and Delphi	To describe the process of using the formal consensus	38 faculty members attending the conference.	Develop core leadership training competencies for paediatric	Self-report 5 point Likert Scale	Descriptive presentation of results only

				method and outcomes	Group sizes were 5-6 people and 7 groups participated	pulmonary centres		
Landeta	**L**	2011	Hybrid NGT, Delphi, and focus groups	To propose a hybrid NGT-Delphi-focus group approach to formal consensus methods	9 Human resources managers from 40 Basque companies, small business workers, human resources and nursing managers from different hospitals. Group sizes were 10, 10, and 6 in the focus group, and 13, 18 15 in the Delphi rounds)	3 groups addressed following questions: 1) improving continuous management training 2) understanding what constitutes competitive advantage for family business, and 3) developing nursing competencies	Self-report 10 point Likert Scale	Descriptive presentation of results only

Results

The experiences of participants using formal consensus methods

Descriptive themes were grouped under “positive” (8) and “negative” (11) appraisals from individuals across the formal consensus data. The overarching themes of “positive” and “negative” are likely to have arisen from the framing of questions in the method of the studies, since the open-ended questions asking about participant experiences of the process were often grouped into questions about the relative strengths and weaknesses of the approach. Although some of the themes spanned both positive and negative feedback (e.g. group participation), they remained separated for analysis due to systematic differences in their method of capture.

Positive themes tended to be represented by qualitative and quantitative data but negative themes presented in the qualitative data only. Figure 2 illustrates the percentage spread of the themes (see Figure 2). The distribution of the negative themes can be found in Figure 3. Negative themes were only present in the qualitative data (see Figure 3).

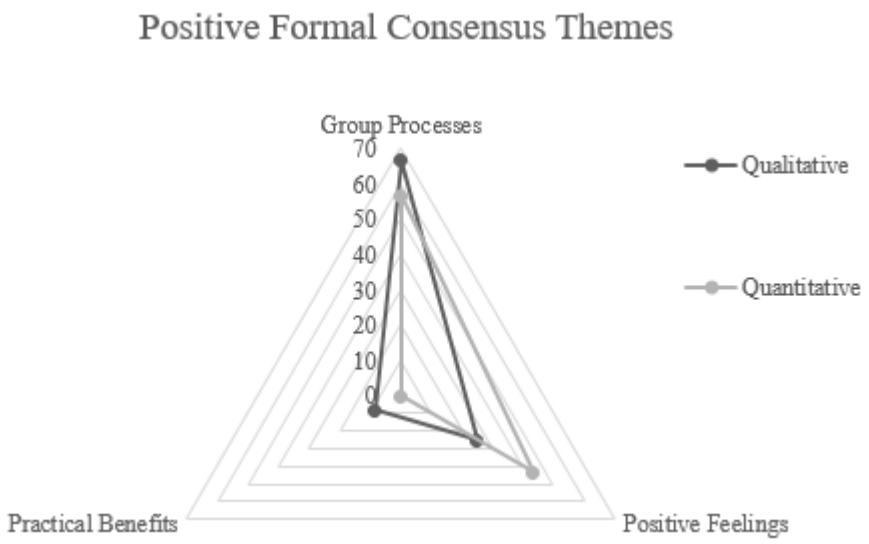


Figure 2: Chart to show percentage distribution of positive themes in data

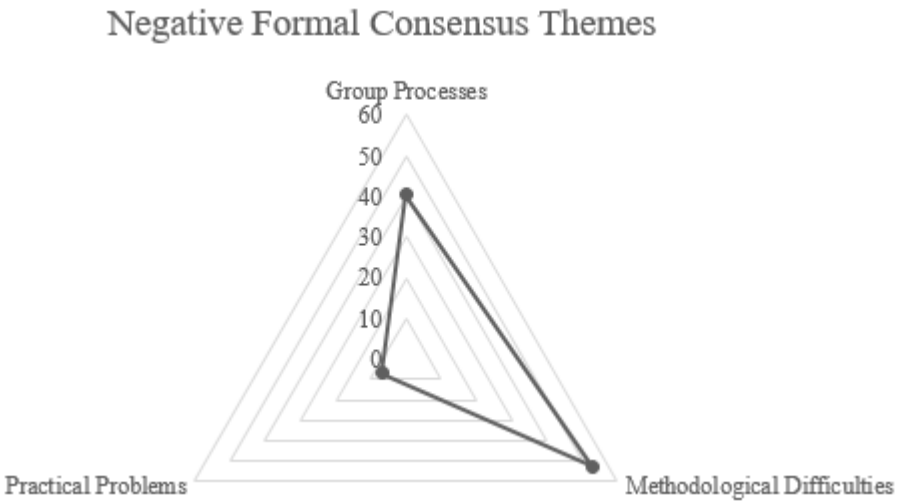


Figure 3: Chart to show percentage distribution of negative themes in data

Findings shall now be described in more detail below, with reference to particular studies using the coding assigned to each study. This is presented in Table 2.The limited data across the types of formal consensus methods (Delphi, RAND/Appropriateness, and Hybrid methods) resulted in the data being analysed as a whole rather than stratified by formal consensus type. A more detailed presentation

of the positive and negative themes is displayed in Figure 4 and Figure 5.

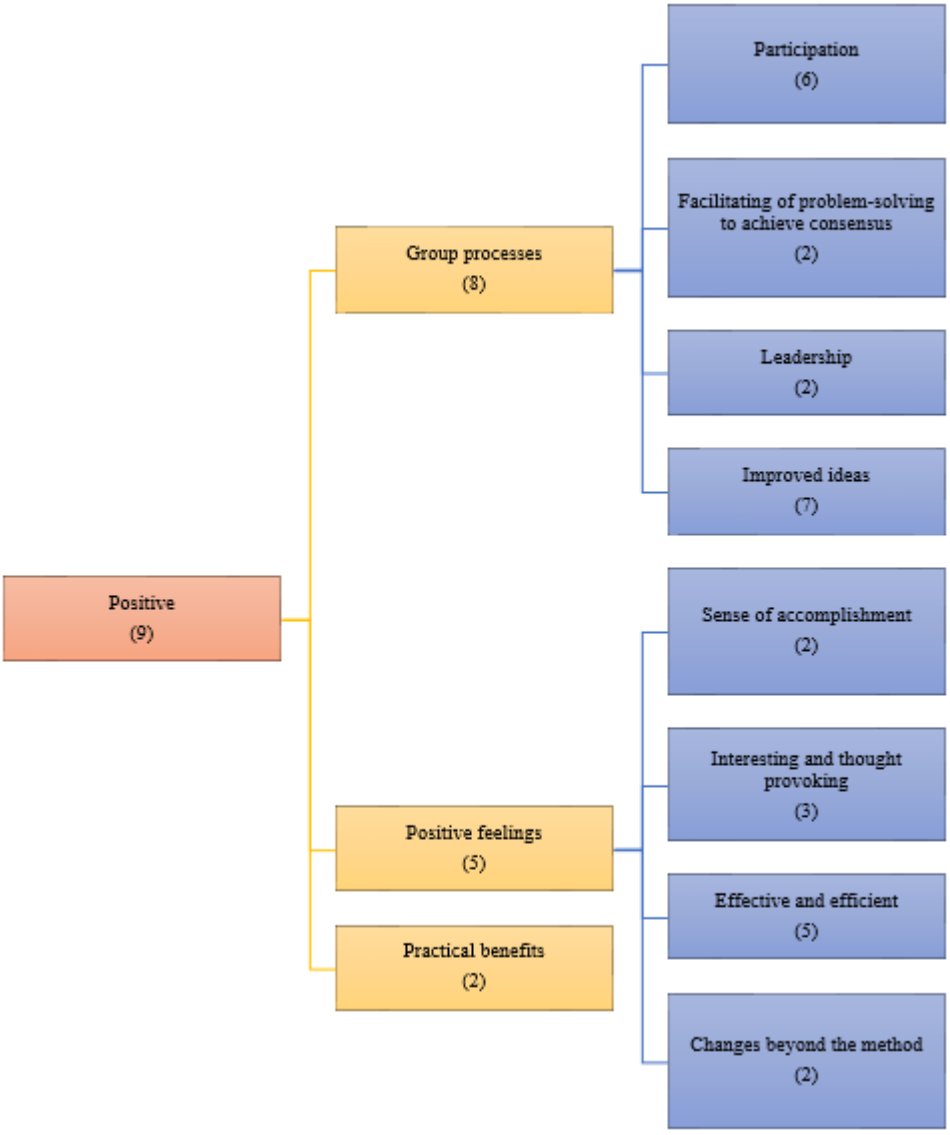


Figure 4: Concept map of positive themes, numbers in brackets indicate number of papers endorsing theme

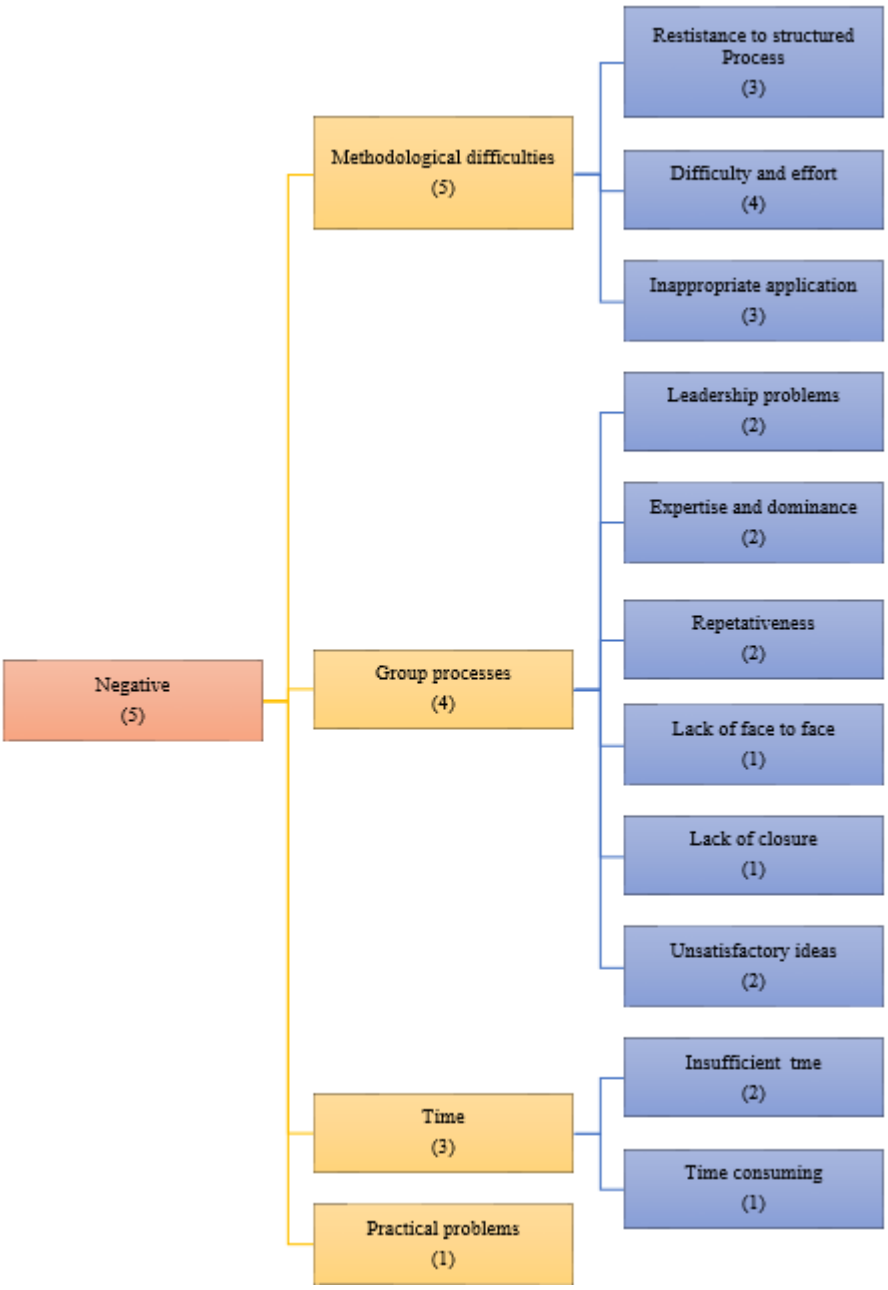


Figure 5: Concept map of negative themes, figures in brackets indicate number of papers endorsing the theme

Group processes. The largest positive theme was regarding "group processes". This theme was endorsed by a large majority of the included studies from both qualitative and quantitative data across all formal consensus methods (V, St, K, B, W, Sh, M, & L). The subthemes shall be presented below.

Participation. Participation presented as a prevalent theme in the data, and participants spoke in their qualitative feedback about feeling unrestricted. Both Delphi and NGT participants made reference to facilitating shared expertise. Delphi participants linked increased feelings of participation to freedom due to anonymity, whereas NGT participants focused more on reducing the influence of dominant individuals to enhance their felt participation in the process.

> *"Facilitating shared expertise" (St; 14.9% prevalence in feedback for NGT)*

> *"Absence of felt pressure from dominant individuals in group. Process allowed independence of thought and expression." (V; 40% prevalence in feedback)*

> *"The freedom associated with being anonymous." (V; 20% prevalence in Delphi feedback)*

Quantitative data supported increased participation for NGT, RAND, and hybrid methods. These studies reported significance across domains of communication when designing a 2-hour Saturday student program (K), ability to participate during a solar energy conference (St), and opinion heard in an orthopaedic surgeon conference (B) for NGT. The RAND and hybrid methods reported significant findings for a sharing of expertise at a healthcare conference and a business management meeting (Sh; L).

Improved ideas. "Improved ideas" was the second largest theme identified in

the NGT qualitative data. This was endorsed by all of the NGT studies and was referenced nine times in the data. The analysis differentiated between quality and quantity of ideas.

"Getting many ideas quickly and smoothly" (G)

"High quality and depth of ideas" (V; 45% prevalence in feedback)

This theme was also corroborated by significant quantitative findings in the majority of NGT studies, further lending support to the agreed benefit by NGT of improving ideas. Some of these studies found that NGT generated significantly more ideas when compared to focus groups for Dutch railway market research (Ru), brainstorming groups for devising a student programme (K), and informal consensus groups that were tasked with defining the job description of dormitory counsellors (V). NGT also produced ideas rated as significantly better quality by participants (St) and observers listening to tapes (Ru).

One of these studies compared NGT to focus groups with five, seven, and ten participants. The study found statistical significance for idea quantity in favour of NGT compared to focus groups of all group sizes (Ru). This study also compared idea quality as measured by three external observers who were experts in the area being discussed. These observers rated ideas in terms of their applicability, flexibility in application, and long term impact. The study showed that NGT groups had a tendency to present ideas rated as better quality. However, statistical significance for idea quality was only found for groups of seven members. The authors suggested that this was because of adverse effects of larger groups on individual perceptions about the value of their own particular contributions to the group. However, this view is not

completely supported by the self-report satisfaction data collected in the study. NGT groups were statistically more satisfied for groups of seven or ten participants and equally satisfied for groups of five participants when compared to focus groups. Satisfaction was inversely related to idea quality for focus groups, with larger groups becoming less satisfied. For NGT, satisfaction increased initially for smaller groups of five and seven participants, but then remained relatively stable for NGT groups of 10 despite differences in idea quality across group size. Thus, satisfaction does not appear to be linked to idea quality in this particular study as hypothesised by the researchers.

The single Delphi quantitative study reported high quantity of ideas for Delphi compared with informal consensus groups. However, the same study found the quantity of ideas to be equal to that of NGT (V).

Of the RAND and hybrid studies, only the RAND paper reported significantly more comprehensive ideas as rated by participants (Sh).

Facilitation of problem-solving or achieving consensus. Qualitative feedback from participants referenced formal consensus methods as facilitating problem-solving and achieving consensus. There were more references to this in the NGT rather than the Delphi feedback, with it being disclosed in two NGT and one Delphi study.

> *"NGT's ability to retain a strong task focus, balancing the problem-solving orientation against the social orientation of the group." (St, from NGT feedback)*

> *"The process of writing responses to the questions forced one to think through the problem" (V, from Delphi feedback)*

It is unclear whether the quantitative data offer support for NGT. One of the studies that used NGT to develop clinical guidelines during an orthopaedic conference reported that individuals found the evidence review helpful (B). However, this may not have been a factor specific to the NGT process and may have been a more general statement about evidence reviews. More specific quantitative data can be derived from one of the hybrid studies, which found participants rated strong agreement that the process had facilitated consensus when a group was tasked with developing leadership competencies for paediatric pulmonary care centres (M), and that they felt that it was a better method than informal consensus methods for judging how appropriate an appropriate treatment would be for lower back pain (Sh).

Leadership. The theme of leadership was elicited from participants in two studies about NGT only, and the theme prevalence was smaller than that of other themes.

"Unbiased role of the leader" (V; 10% prevalence from NGT feedback)

Quantitative data encompassing this theme was only captured by one of the hybrid studies. They found that participants rated leaders highly at a conference developing healthcare leadership competencies and outcome measures (M).

Positive feelings from participants. There was a relatively consistent disclosure of positive reactions to the formal consensus process across all of the methods.

> *"Meeting was constructive, and developed a positive approach and attitude among group members" (V, 35% prevalence of feedback for NGT)*

"Enjoyable" (W, feedback for Delphi)

Sense of Accomplishment. Participants spoke about feelings of accomplishment across the qualitative data for NGT and Delphi:

> *"Objectivity of the approach to assessing the problem was geared to getting something done, shows someone is interested and it is a start in the right direction" (V, 25% prevalence in feedback for Delphi)*

> *"Sense of accomplishment, task closure, or pooling of judgement" (St, 14.9% prevalence in feedback for NGT).*

Likert scores supported this theme in the quantitative data for the NGT, Delphi, and RAND studies.

Interesting and Thought Provoking. Participants disclosed finding the process interesting.

> *"Gained new knowledge by listening to the different ideas of others" (V, 65% prevalence in feedback for NGT)*

> *"Thought provoking to read other's comments'" (W in feedback for Delphi)*

The quantitative data reported on positive feelings from participants in a general way. Participants at an orthopaedic conference endorsed questionnaire items that the NGT process enhanced the understanding of the subject matter (B). NGT was also rated as significantly more satisfying than Delphi and informal consensus groups by university residential participants (V). In the same study, there was no difference in satisfaction between Delphi and informal consensus groups. Both hybrid studies also noted positive feelings from participants. Faculty members highly rated the hybrid Delphi-NGT consensus process when they used it to develop

paediatric leadership training competencies (M). A hybrid Delphi-NGT method was communicated as satisfactory by managers from business and nursing backgrounds (L). This was in terms of interacting with other colleagues, learning something new, and contributing knowledge for research (L). Individuals also rated the RAND method as a significantly better method (Sh).

Effective and efficient. A reference to efficiency and effectiveness was made once in the qualitative data when participants spoke about Delphi.

> *"An expedient, practical way for participation in decision-making by a wide variety of people. The repetitive feedback and multi-questionnaire approach is a convenient, sensible, approach to investigating a complex problem" (V; 35% prevalence from Delphi feedback)*

This theme was most prominently featured in the NGT quantitative data in four of the six quantitative studies, driven by researcher questions. A high proportion of individuals rated NGT as an effective process for reaching consensus (B). At the same conference, participants gave a medium-high rating for NGT producing better guidelines. Individuals also rated their experience of NGT as medium-high in terms of being a good use of conference time (G). NGT and brainstorming were rated by participants as significantly more effective for group processes than informal consensus methods (K). Another study found significance for participant ratings of NGT for effectiveness when compared with a neutral rating (St).

Changes beyond the method. A final theme placed within the theme of positive feelings that featured only in the quantitative data referred to effects beyond the formal consensus itself. A high proportion of conference attendees stated the method would result in them changing their clinical practice (B). Other studies reported medium ratings that they would use NGT (G) and a hybrid method (L) in

the future (G).

Practical benefits. A theme of relatively smaller prevalence concerned practical benefits. This was presented by individual feedback regarding NGT. Comments focused on group size, clarity and accuracy of recording and instructions, and the informality of the meeting.

> *"Clarity of instructions and clearly organised meeting" (V; 15% prevalence from NGT feedback)*

> *"The opportunity to be involved while I didn't have to attend any meeting" (V; 10% prevalence in Delphi feedback)*

There was no measurement of practical benefits in the quantitative data.

Negative themes. Of particular interest is that only one of the quantitative studies asked participants about negative reactions to formal consensus methods. Therefore, the majority of data regarding negative themes is discussed in the context of qualitative findings. Participants endorsed negative themes at a lower rate than positive themes across all formal consensus styles; percentage endorsement did not surpass 35%.

Methodological difficulties. Of the negative themes, the largest was the theme that described methodological difficulties. This theme consisted of quotes that discussed the procedures of formal consensus methods.

Resistance to structured process. A theme mentioned by three NGT and one Delphi qualitative study was regarding general resistance to the structured process.

> *"General resistance to structured process- desire for more flexibility" (V, 20% prevalence of NGT feedback)*

"Don't feel completely represented by a piece of paper. It lacks personal feeling of helping a cause" (V, 25% of Delphi feedback)

Two specific subthemes appeared to permeate both NGT and Delphi findings. One was a specific dislike of writing of ideas.

"Dislike of questionnaires, I just dislike writing anything" (V, 15% prevalence of Delphi feedback)

"Dislike writing my ideas: felt like taking a test again, although it is the only way to get it on paper" (V, 10% prevalence of NGT feedback)

Another was a dislike of the voting procedure whereby participants found the voting too restrictive.

"The voting system totally distracted from the need to discuss the specifics of the various issues' proposed for recommendation." (G, NGT feedback)

"Felt restricted in ranking only five priorities" (V, 25% prevalence from Delphi feedback)

Specific to the NGT data was a sense that the structure placed too many limits on the discussion.

"Pressures to categories ideas on flip-charts in discussion" (V, 10% prevalence in NGT feedback)

Difficulty and effort. Issues concerning the difficulty of formal consensus methods were presented in the large majority of qualitative NGT and Delphi studies. Participants talked about the effort of using formal consensus methods, and likened the learning of the process to the application of a new skill that required effort, adequate instruction, and training.

"There was no guidance or at least poor instructions as to whether we were striving for issues, problems, or solution strategies" (G, feedback about NGT)

"Sheer number of words on the questionnaire and found it daunting at times and potentially off-putting." (W, feedback about Delphi)

The RAND study presented a single example from the quantitative data endorsing the theme of "effort". Participant scoring of effort for the RAND method as compared to informal consensus methods reached statistical significance.

Inappropriate Application. Some individuals felt that the context in which Delphi and NGT methods were being applied was inappropriate in relation to the specific task or resources allocated.

Inappropriate applications of NGT, deficient group output, and preplanning deficiencies (St, 5.5% prevalence of NGT feedback)

Participants felt NGT and Delphi were being incorrectly applied to a topic question, which was a subtheme that was identified for both of them.

"The question posed to the group involved in the NGT was too broad, too vague for the NGT to have been most successful" (G, NGT feedback)

"Questionnaire was too general, too open-ended" (V, 30% prevalence of Delphi feedback)

Time. Both Delphi and NGT discussed problems with time.

"Too little time, too much time, or inefficient use of time" (St, 21.8% prevalence from NGT feedback).

More specifically, NGT and Delphi participants made reference to the formal consensus method being time consuming, although this theme was more prevalent

for the Delphi rather than the NGT data.

> *"Too much waiting while the ideas are written down" (V, 5% prevalence from NGT feedback)*

> *"Extra guidance on the time needed for the second and third questionnaire would have been welcomed" (W, Delphi feedback)*

More particular to the NGT data was feedback that time was insufficient.

> *"There was far too little time to do a good job" (G)*

Group processes. The theme of group processes was identified in the negative as well as the positive clusters of themes, with very similar attributes being discussed by participants.

Leadership problems. A subtheme that presented in two NGT studies was problems with leadership in terms of experience, competence, and personality.

> *"I think the success of any group is directly related to the ability and personality of the NGT leader; in my case, his inexperience showed and was a detriment to how the discussion was handled." (G, feedback about NGT).*

> *"Inaccurate recording of ideas by leader" (V; 5% of NGT feedback)*

Expertise and dominance. Both NGT and Delphi participants discussed problems with group expertise.

> *"Belief that they (or other participants) did not possess an adequate knowledge of the subject area or the objectives of the workshop" (St; 10.9% prevalence from Delphi feedback)*

> *Difficulty of the problem statement. Felt I was asked to answer questions*

I did not feel qualified to answer (V; 20% prevalence from Delphi feedback)

Some subthemes emerged as specific to the particular consensus method. Within the NGT feedback data, an additional subtheme further to an insufficiency of expertise was a feeling that knowledge was too diverse in the group.

"The groups were too diverse so only general issues came to the top. The groups should be much more homogeneous and technically specific." (G from NGT feedback)

Repetitiveness. A subtheme that was specific to all of the included qualitative Delphi studies was a sense of repetitiveness of presentation.

"Repetitive" (W, feedback from Delphi)

"Too great a quantity of ideas. Feedback list was too long with many relatively closely related ideas." (V, feedback from Delphi)

Lack of closure. A theme of lack of closure was derived from the NGT and Delphi qualitative data. This theme was discussed in the studies within the context of participant expression of a sense of low felt accomplishment and of lack of changes beyond participating in the consensus method process.

Lack of closure in knowing what happens next; don't know if this meeting will have an effect (V, 20% prevalence of NGT Feedback)

Lack of closure in knowing what happens next. Don't know if this survey will have an effect (V, 15% prevalence of Delphi Feedback)

Lack of Face to Face. Lack of face to face contact was a theme specific to the

Delphi studies. This theme was characterised by the absence of real time feedback to aid decision-making and a lack of stimulation. It is of note that the theme of stimulation and interest shared some quotations with those coded as part of a lack of face to face. This could echo some of the positive feedback described earlier regarding discussions as interesting and thought provoking.

> *"Anonymous recipient of information: didn't know who I was expressing myself to, or who my "group" was. Trying to figure out the kind of response that was wanted or understandable to the "group"." (V, 35% prevalence of Delphi feedback)*

> *"Discussion with others would have been more interesting and stimulating" (V, 10% prevalence of Delphi feedback)*

Unsatisfactory ideas. Dissatisfaction with ideas was a theme synthesised from the NGT studies only. This related to feelings that there were too many closely related and redundant ideas, that ideas were not diverse enough, and that there were too many ideas.

> *"Damaging, embittering, and levelling process which eliminates important ideas based on detailed knowledge and imagination" (G, NGT feedback)*

Discussion

This conceptual review provides an overview of the experiences of participants using formal consensus methods specifically in the context of healthcare guidelines.

Critical appraisal of studies

Only four of the ten included studies were directly relevant to healthcare guidelines. Six other studies were included as they were considered relevant because they applied formal consensus methods to question that did not have a clear "correct" answer. Instead, they required the sharing of heterogeneous expertise which was considered analogous to decision-making in health care settings. A strength of the studies was that they were all naturalistic and included participants that were both interested in and personally impacted by the outcome of the guidelines generated, which is also a similar scenario to healthcare guidelines.

A challenge when comparing these studies is the variation in the amount of time dedicated to the formal consensus method, particularly with regards to NGT. Some of the studies used NGT over one (M) or two (B) days, whereas others used it over 1.5 (St) or 2 hours (K). There seemed to be no systematic differences between the actual time taken and the feedback regarding time in the data however, which supports time as a general theme rather than as an element that is impacted by the specific implementation of formal consensus methods. For example, efficiency was mentioned by participants in the study that spanned the longest time (B), and time was discussed both as being inefficient and insufficient in a study that offered one of the shortest times for NGT that applied it over two hours (V).

Leadership effectiveness is particularly important for NGT outcomes (e.g. Souder, 1977). Therefore, it is important to note that there was variation of training offered to leaders in the formal consensus method, from a group being given instructions on the NGT principles followed by a 5-minute practice (K) to selecting group leaders who are already experienced in formal consensus methods (M). Leadership style and competence were identified as part of the negative themes from participants with specific reference to lack of training and expertise. It is of interest that the highest frequency of comments (which were negative) regarding this theme was derived from a study that stated the NGT leaders had been specifically trained to implement the NGT at the conferences (G).

The included studies used self-report measures through Likert scales and open-ended questions to capture experiences, with only one study triangulating results with the ratings of observers listening to taped sessions (K). More generally, self-report findings are limited by the participants' abilities to access and portray their experiences (Barker, Pistrang, & Elliott, 2002). The quality of the qualitative data could have been impacted by social desirability biases as well as inaccuracies in participants' memory recall (Mortel, 2008). However, a strength could have been that feedback was anonymised within most of the studies, which could have encouraged honest expression. It is unclear whether recall distortion would be an important aspect of experience or whether it would introduce unhelpful bias. There was no specifically validated Likert scale for the experiences of formal consensus methods by participants, so the studies largely used ad-hoc measures of experiences. This means that the Likert scales between studies should be compared with caution, because questionnaires could be subjected to systematic biases within the data such as framing effects.

A major limitation of the quantitative data is the lack of a control group in 70% of the studies. Researchers managed this limitation by using the central point of the Likert scale as a hypothetical comparison (St), presenting only descriptive statistics (G, M, L), or using the 95% confidence interval to postulate statistical significance and compare means (Sh). Some of the studies compared different formal and informal consensus methods. For example, NGT versus Delphi versus informal groups (V), NGT versus informal groups (Ru), NGT versus brainstorming versus informal groups (K), and Hybrid NGT and Delphi versus focus groups (L). The variety of comparison methods and applications supports a tentative aggregation of the findings. Thus, the derived themes are intended as general commonalities across formal consensus methods rather than specific attributes of any particular formal consensus approach.

Disagreement of qualitative and quantitative data

Some these were represented exclusively in the qualitative or quantitative feedback. Researchers tended to ask participants about positive themes through the use of Likert scales, and thus the negative themes were derived almost solely from qualitative participant comments. This bias in the data underscores the importance of using qualitative as well as quantitative data for gathering feedback. The practical benefits of formal consensus methods were commented upon exclusively in qualitative interviews rather than Likert-feedback. Participants commented on benefits such as no requirement to attend a meeting for Delphi, or having clearly communicated instructions. Researchers asked more frequently about formal consensus method efficiency and effectiveness, and whether the outcomes of formal consensus methods changed practice. There could be several reasons for disparities

between quantitative data capture and qualitative participant expression.

Firstly, there might be a difference in language between researchers and participants. For example, the effectiveness and efficiency as defined by researchers might relate to participant themes of practical benefits, time, difficulty and effort, or quality and quantity of ideas. Secondly, this disparity may be reflective of differences in priorities between participants and researchers of the method. More specifically, participants might prioritise usability of formal consensus methods whereas researchers might emphasise efficiency and effectiveness in terms of the outcomes of the method. It is important to consider and have awareness of the potential differences in priorities between participants and researchers when selecting a formal consensus method so that it is acceptable to its users rather than just reflecting researcher priorities.

Agreement of qualitative and quantitative data

Group Processes. Group processes were a major component of feedback across all formal consensus methods from a qualitative and quantitative perspective. One element of group processes that covered both positive and negative feedback was participation, which included the management of expertise. Improved participation is a key goal for the use of formal consensus methods to mitigate social pressures and improve the sharing of expertise (Black et al., 1999). Interestingly, the studies that focused the most on participation were those using RAND or Delphi-NGT hybrid consensus methods. This was specifically in terms of facilitating shared expertise. These studies are all from a healthcare context. It could be that the sharing of expertise is believed to be a priority for healthcare professionals, rather than managing power differentials or increasing participation more generally. This could

be seen as mirroring the multi-disciplinary approach used by health professionals in the NHS since it emphasises a pooling of expertise(Mental Health Taskforce, 2014). Because the RAND and hybrid data were derived purely by researcher-driven questionnaire capture, it is unclear whether participants themselves would highlight other components within the theme of participation, such as minimising the impact of dominant individuals.

Difficulties with managing expertise formed part of the negative feedback and can be explained through commonalities as well as differences between Delphi and NGT. Participants tended to report lacking either personal or group expertise when using Delphi or NGT respectively. This is understandable given the emphasis for effective decision-making to include group members of heterogeneous expertise (e.g. Black et al., 1999), which is likely to result in members feeling unable to contribute to all aspects of discussion. It may be important to allow participants the option of not responding to some of the questions. Studies have supported evidence of an "equality bias", where groups have a tendency to treat their members as equal even though genuine differences in ability exist (Mahmoodi et al., 2015). More research is needed to better understand how equal participation can be best managed in heterogeneous guidelines groups in order to balance participation with quality of output.

A particular contribution of Delphi to feelings of lacking expertise could be related to isolated contribution, which renders participants unable to seek clarification on posed questions. This could heighten a sense that individuals are unable to answer given their knowledge base (e.g. Yang, 2013). Feedback regarding managing expertise that was specific to NGT was that group knowledge was felt to

be too diverse. Comments from NGT group members seemed to be related to a lack of agreement, which participants felt reduced the effectiveness of the NGT process. This concurs with research into the merit of discussion to support sharing of expertise, promote effectiveness, and distinguish reliable from unreliable expertise (Bahrami et al., 2012; Klein & Epley, 2015; Minson, Liberman, & Ross, 2011), despite research evidence that participants generally believe informal group discussion to be of limited value to performance (Mercier, Trouche, Yama, Heintz, & Girotto, 2014). It could be that group members felt that this increased disagreement was not managed within the restrictions of the NGT process since it might require extra time for clarification and deliberation through discussion. Quotations within the negative feedback from NGT labelled "limitations to discussion" lend support to this. Within this subtheme, participants spoke about feeling unable to follow up on all ideas and wanting a longer discussion period.

There were positive and negative group process components that were specific to the Delphi method. Participants identified the anonymity offered by Delphi as a positive aspect of the process. Anonymity is restricted for NGT because of the collection and clarification of individual ideas as a group. The benefit of anonymity has been demonstrated experimentally by manipulating the NGT process to enable it. In these instances, anonymity has been associated with improved idea quantity and quality as compared to standard NGT (Sullivan, 1978). However, the negative feedback from the Delphi data identified a lack of face-to-face interaction as difficult for participants, since they said it resulted in a lack of real-time feedback and also made the process less stimulating. The identification of the strengths and weaknesses of anonymity tool for improved participation could support a hybrid version of Delphi and NGT where anonymity is supported to enhance participant

experiences.

Idea quality and quantity was another strong theme that included both qualitative and quantitative data. For idea quantity, the quantitative studies offered a mixture of participant report and actual count of ideas. These data agreed with each other, with participants agreeing with the actual count by researchers on the quantity of ideas generated. Judging the quality of ideas posed a challenge to the quantitative studies, since idea quality may well be subjective in the area of policy and healthcare. For example, a measure of quality for a participant of formal consensus methods might be the comprehensiveness of the agreed-upon policy, whereas an observer might view the utility of produced guidelines as a measure of quality. Studies approached this in a range of ways. Some research used external raters who were knowledgeable in the field to rate data quality on effectiveness, creativity, feasibility, and interest to users (K), whereas other studies captured participant self-report (St; Ru; B; & V).

The strongest support overall for both improved quality and quantity of ideas was for NGT. Those participating in the Delphi method offered some support for idea quantity and variety, which was found in the quantitative data only. The quantitative results from the RAND study reported significance relating to the quality of ideas only.

Despite difficulties in the measurement of idea quality and quantity, it is clear that it is an important feature of formal consensus methods from the perspective of participants. Previous qualitative research has interviewed market researchers on their perceptions of NGT. The findings agree that NGT is a helpful method for the production of ideas (Boddy, 2012); however the authors highlight the importance of

equal participation in the evaluation of those ideas following their generation in order to enhance their application and implementation to the assigned task. A linking of improved participation and better ideas is also in agreement with previous research findings that encouraging members to reveal more information that is unshared with other members improves decision-making (Wittenbaum, Hubbell, & Zukerman, 1999) and might have the largest impact on attitude change (Vinokur, Burnstein, Sechrest, & Wortman, 1985).

The negative feedback regarding idea quantity and quality suggested that participants felt the structured approach limited the diversity of ideas, and that some found the increase in quantity of ideas overwhelming given the limited time. There is often an assumption in the literature that increased idea quantity and quality is preferable (e.g. McMahon, Ruggeri, Kämmer, & Katsikopoulos, 2016). However, there is reason to exercise caution. Within the Delphi negative feedback, participants discussed the increase in ideas as resulting in a repetitive output. For NGT, larger group size has been correlated with a reduced felt sense of involvement in the final decision, which has been postulated to be a disadvantage of allowing equality of participation (Green & Taber, 1980). It could also link to the negative "lack of closure" experienced by the participants who felt less involved in the process and outcome. Furthermore, when compared to informal consensus groups, NGT members report reduced certainty about their own ideas. This was not linked to reduced overall satisfaction with the process, and has been linked to decreased certainty in an individual member's personal view in response to increased exposure to more ideas in the group as with the NGT process (Hegedus & Rasmussen, 1986).

Leadership was a theme specific to NGT that occurred in both the positive

and negative feedback. Positive feedback focused on the importance of the leader being unbiased. Positive and negative comments were gathered from groups about different leaders even when addressing the same question (e.g. Van de Ven & Delbecq, 1974), which lends support to individual leader style as experienced by participants independent of question type. Negative feedback related to the impact of lack of experience or guidance from leaders. This links to previous findings that leaders who are highly structured but considerate about group member's feelings are most favourable according to group members (e.g. Souder, 1977).

Positive feelings. Positive feelings were a strong theme across all formal consensus methods in both the qualitative and the quantitative data. Qualitative feedback described feelings of finding the process enjoyable, having a sense of accomplishment and finding the process interesting, whereas quantitative data focused on the effectiveness and efficiency of the method and changes beyond the group session. This feedback supports formal consensus methods as favourable from the perspective of participants. Task accomplishment has been linked to satisfaction in group decision-making (Wegge & Haslam, 2005). This sense of task closure might overlap in the data with the theme of "changes beyond the method", which linked to the implementation of the generated ideas beyond the meeting. The importance of decision-making impacting beyond the consensus task might be a theme that lends itself as particularly important for guideline development, where stakeholders are invited as part of a committee also to foster acceptance and implementation of the produced guideline (e.g. Boddy, 2012; Bini & Mahajan, 2016).

Methodological difficulties. A major theme from the negative feedback was relating to methodological difficulties. Some individuals found the formal consensus

process restrictive, although closer inspection of quotes relates this in part to a lack of familiarity with the process. Similarly, participants spoke about the additional cognitive effort and adaptation required to use formal consensus methods, including the need to be concise in their ideas and focus on solutions only for NGT, and needing to read through a lot of text for Delphi. Participants felt that additional resources were not allocated for this, and linked it to the competence of the leaders as well as the clarity of the instructions they were given. There is limited literature regarding this area, but there are a few studies offering evaluation of NGT from a researcher perspective (Allen, Dyas, & Jones, 2004) and evaluation of Delphi from a researcher (West, 2011; Yang, 2013)and a participant (Williams & Webb, 1994) view. These studies discuss augmented workload as a common theme of formal consensus methods compared to informal consensus methods. It may be that more emphasis needs to be made on the increased resources required when implementing formal consensus methods.

There were also quotes that reflected the inappropriate application of formal consensus methods. This subtheme was expressed more often in NGT as the topic question being inappropriate and in Delphi as the statements being too general. Indeed, there has been some research into when different types of formal consensus methods are best applied (e.g. Hutchings, Raine, Sanderson, & Black, 2006). For example, that NGT is best suited for generation but Delphi instead for aggregating expert opinion on an established topic (Nair, Aggarwal, & Khanna, 2011). Individuals also referred to problems with time management. For NGT, there was mention of both too much and too little time, whereas Delphi triggered more feedback about the task being time consuming.

Limitations of the review

The present mixed methods review was not intended to be systematic review of the literature. Rather, the purpose was to offer a narrative synthesis of the research to date. A narrative synthesis rather than a systematic approach was chosen due to the wide breadth of study design and data reporting types, probably due to the inclusion of interdisciplinary journals including those from management, cognitive psychology, medicine, and health policy. Due to a limited evidence base, research was collected from as early as the 1970s. Statistical reporting and methodology has evolved considerably over time, with increasingly clearer data reporting guideline requirements for journal publication. Despite this, the mixed methods systematic review appraisal tool (Heyvaert, Maes, & Onghena, 2013) can be loosely used as a framework when appraising the limitations of this mixed methods review.

The tool initially focuses on the review question, and whether it merited a mixed methods design. The question of participant experiences seems best answered using a qualitative design, and triangulation is important for the credibility of qualitative data (Barker, Pistrang, & Elliott, 2002). Since including quantitative studies can offer a form of triangulation (Braun & Clarke, 2006), a mixed methods approach seems useful when answering the question of participant experiences. In addition, including only qualitative studies would have resulted in only five studies being included. Thus, the review could have failed to include a considerable number of themes or narratives.

Next, the tool focuses on how successfully the components of the study are integrated into the research question. The investigation of participant experiences is more relevant to qualitative rather than quantitative research (Barker, Pistrang &

Elliott, 2002). Therefore, the review primarily led with a qualitative thematic analysis of the studies and used the quantitative results in the second phase to descriptively augment the generated themes. However, a limitation of this relates to how well these components were interpreted in the review analysis. The qualitative papers used in the review often lacked the specific qualitative data derived from participants. The themes as developed by the researchers in the studies often lacked the specific utterances of the participants. Without the complete dataset, an aggregated synthesis across studies is limited in completeness (Kuper, Lingard, & Levinson, 2008). Important aspects of the data may be missed when several studies are aggregated since the reported themes without quotations will lacks some of the "richness" of the original data set (Cassell, Symon, & King, 2014). Therefore, there is an increased risk of a biased interpretation of the data.

Inconsistencies in the reporting of results by qualitative studies is relatively common, partly because of the variety of approaches that it encompasses (Dixon-Woods, Shaw, Agarwal, & Smith, 2004). Indeed, some of the research reviewed included the percentage coverage of particular themes in the dataset, and some did not. As part of the overall review, it was decided to include the theme percentage prevalence information when possible. The aim of this was to provide further richness and context to those specific themes in terms of how dominant they had been in the particular study. However, it is unclear how these percentage endorsements can be relevant when comparing across studies of different contexts. Therefore, the relative dominance of themes should be interpreted with caution and placed in the context of the overall narrative of the review.

The appraisal tool includes an assessment of how divergences and

inconsistencies are discussed between the qualitative and quantitative aspects of the data. This was an important feature of this review, and similarities and differences across qualitative and quantitative elements of the review were discussed separately. The tool also notes the importance of methodological rigour in both qualitative and quantitative aspects of the review. The qualitative component of the analysis paid careful attention to good practice of qualitative analysis, including the use of triangulation, using quotations to support themes to increase credibility, and articulating the theoretical approach and application of the thematic analysis to allow the reader to assess the "trustworthiness" of the data (Nowell, Norris, White, & Moules, 2017). However, the quantitative analysis was less rigorous, and therefore is a weaker component of the review. The quantitative studies provided mostly a descriptive account of participant views, and often did not use statistical analysis to assess the significance of results. Therefore, the quantitative component of the mixed methods review is intended to offer a descriptive account of participant experiences rather than a rigorous and potentially generalisable account.

The present review attempted to manage some of these limitations through the use of several means, including triangulating the data by using mixed methods, situating the data using quotations, and credibility checks of themes with another researcher and by being explicit in terms of the approach to analysis. Nevertheless, these considerations remain relevant and this review would benefit from being replicated following the publication of further relevant studies.

Conclusions

This review highlighted the experiences of participants using formal consensus methods as applied to "complex" problems without any single obvious answer. In summary, group processes are a major theme within formal consensus methods. Within this theme, increased participation has been identified as valuable and important to users. However, the impact of this on participant experiences is multifaceted, and requires careful consideration in terms of its implementation. Participant experiences of formal consensus methods could be improved, for example, by allowing for participants to opt out of participation if they feel unable to contribute. Encouraging participation is likely to be augmented by preserving anonymity, although some face-to-face feedback is important to preserve feedback, clarification, and to encourage positive feelings of accomplishment. A pressure to reach consensus at the expense of free discussion is likely to negatively impact participant experiences and potentially the quality of ideas, since discussion may have specific merits that need to be unhindered by over-restrictions in formal consensus methods.

Studies generally commented that they enjoy the formal consensus method process, and in particular that they find it interesting and appreciate feeling that they are accomplishing a task. It seems to be particularly important for participants that the outcome from the consensus process can be linked to implementation and change beyond the meeting. For formal consensus methods to be most effective, additional time and resources should be offered for training and using the method. Additionally, individuals should be informed of the time commitment involved in partaking in formal consensus methods so they can make an informed choice about whether to

participate.

Some authors have suggested a hybrid approach to draw together the strengths of different approaches, including the anonymity and reliability of Delphi with the time-saving face-to-face feedback of NGT (e.g. Hutchings, Raine, Sanderson, & Black, 2006). The data support also preserving the more free discussion element of informal consensus methods as an important tool for problem solving.

Finally, the present review revealed gaps in the literature, for example collecting quantitative data on potentially negative aspects of participant experiences. An extension of the present review could be to analyse the literature regarding researcher perspectives on the development of formal consensus methods in order to contrast these with the views of participants.

References

Allen, J., Dyas, J., & Jones, M. (2004). Building consensus in health care: a guide to using the nominal group technique. *British Journal of Community Nursing*, *9*, 110–114.

Anson, R., Bostrom, R., & Wynne, B. (1995). An Experiment Assessing Group Support System and Facilitator Effects on Meeting Outcomes. *Management Science*, *41*, 189–208.

Asch, S. E. (1952). Group forces in the modification and distortion of judgments. In *Social psychology* (pp. 450–501). Englewood Cliffs: Prentice-Hall, Inc.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012). What failure in collective decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 1350–1365.

Bajracharya, S. M. (2006). An Assessment of the Perceived Barriers and Strategies to Promoting Early Detection of Colorectal Cancer: A Practitioners' Perspective. *International Quarterly of Community Health Education*, *26*, 23–44.

Bang, D., & Frith, C. D. (2017). Making better decisions in groups. *Royal Society Open Science*, *4*, 1–22.

Barker, C., Pistrang, N., Elliott, R. (2002). *Research Methods in Clinical Psychology: An introduction for students and practitioners*. Chichester, UK: Wiley.

Bini, S. A., & Mahajan, J. (2016). Achieving 90% Adoption of Clinical Practice

Guidelines Using the Delphi Consensus Method in a Large Orthopedic Group. *The Journal of Arthroplasty*, *31*, 2380–2384.

Black, N., Murphy, M., Lamping, D., McKee, M., Sanderson, C., Askham, J., & Marteau, T. (1999). Consensus development methods: A review of best practice in creating clinical guidelines. *Journal of Health Services Research and Policy*, *4*, 236–248.

Boje, D. M., & Murnighan, J. K. (1982). Group Confidence Pressures in Iterative Decisions. *Management Science*, *28*, 1187–1196.

Bolger, F., & Wright, G. (2011). Improving the Delphi process: Lessons from social psychological research. *Technological Forecasting and Social Change*, *78*, 1500–1513.

Bowman, J. M., & Wittenbaum, G. M. (2012). Time Pressure Affects Process and Performance in Hidden-Profile Groups. *Small Group Research*, *43*, 295–314.

Braun, V., & Clarke, V. (2006). Using Thematic Analysis In Psychology. *Qualitative Research in Psychology*, *3*, 77–101.

Brown, R. (2001). *Group processes : dynamics within and between groups*. Cornwall: Blackwell Publishers.

Campbell, S. M., & Cantrill, J. A. (2001). Consensus methods in prescribing research. *Journal of Clinical Pharmacy and Therapeutics*, *26*, 5–14.

Carney, O., McIntosh, J., & Worth, A. (1996). The use of the Nominal Group Technique in research with community nurses. *Journal of Advanced Nursing*, *23*, 1024–1029.

Cassell, C., Symon, G., & King, N. (2014). Using Templates in the Thematic Analysis of Text. In *Essential Guide to Qualitative Methods in Organizational Research* (pp. 256–270). SAGE Publications Ltd.

Chasens, E. R., & Olshansky, E. (2008). Daytime Sleepiness, Diabetes, and Psychological Well-Being. *Issues in Mental Health Nursing*, *29*, 1134–1150.

Chassin, M. (1989). Developing good medical standards we all can live with. *The Internist*, *30*, 6–8.

Cooper, C., Booth, A., Britten, N., & Garside, R. (2017). A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. *Systematic Reviews*, *6*, 234.

Cooper, C., Dawson, S., Peters, J., Varley-Campbell, J., Cockcroft, E., Hendon, J., & Churchill, R. (2018). Revisiting the need for a literature search narrative: A brief methodological note. *Research Synthesis Methods*, *9*, 361–365.

Cramer, R. J., Brodsky, S. L., & DeCoster, J. (2009). Expert witness confidence and juror personality: their impact on credibility and persuasion in the courtroom. *The Journal of the American Academy of Psychiatry and the Law*, *37*, 63–74.

Daniels, N., & Sabin, J. E. (2008). Accountability for reasonableness: an update. *BMJ*, *337*, a1850–a1850.

De Ruyter, K. (1996). Focus versus nominal group interviews: a comparative analysis. *Marketing Intelligence & Planning*, *14*, 44–50.

Dixon-Woods, M., Shaw, R. L., Agarwal, S., & Smith, J. A. (2004). The problem of

appraising qualitative research. *Quality and Safety in Health Care*, *13*, 223–225.

National Intitute for Care and Clinical Excellence (2016). *Mental health problems in people with learning disabilities: prevention, assessment and management*. NICE: London.

Faulmüller, N., Mojzisch, A., Kerschreiter, R., & Schulz-Hardt, S. (2012). Do You Want to Convince Me or to Be Understood?: Preference-Consistent Information Sharing and Its Motivational Determinants. *Personality and Social Psychology Bulletin*, *38*, 1684–1696.

Fitch, K., Bernstein María, S. J., Aguilar, D., Burnand, B., Lacalle, J. R., Lázaro, P. van het Loo, M., McDonnell, J., Varder, J., & Kahan, J. P. (2001). The RAND/UCLA Appropriateness Method User's Manual. *Santa Monica, California: RAND Coorporation.*

Fitzgerald, L., Ferlie, E., & Hawkins, C. (2003). Innovation in healthcare: how does credible evidence influence professionals? *Health and Social Care in the Community*, *11*, 219–228.

Fletcher, A. J. (2017). Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology*, *20*, 181–194.

Gill, P. J., Wang, K. Y., Mant, D., Hartling, L., Heneghan, C., Perera, R., … Harnden, A. (2011). The Evidence Base for Interventions Delivered to Children in Primary Care: An Overview of Cochrane Systematic Reviews. *Public Library of Science One*, *6*, e23051.

Graefe, A., & Armstrong, J. S. (2011). Comparing face-to-face meetings, nominal groups, Delphi and prediction markets on an estimation task. *International Journal of Forecasting*, *27*, 183–195.

Gresham, D. J. (1986). *Expressed Satisfaction with the Nominal Group Technique Among Change Agents*. Graduate College of Texas A&M University.

Hemingway, H., Chen, R., Junghans, C., Timmis, A., Eldridge, S., Black, N., … Feder, G. (2008). Appropriateness criteria for coronary angiography in angina: reliability and validity. *Annals of Internal Medicine*, *149*, 221–231.

Heyvaert, M., Maes, B., & Onghena, P. (2013). Mixed methods research synthesis: Definition, framework, and potential. *Quality and Quantity*, *47*, 659–676.

Hutchings, A., Raine, R., Sanderson, C., & Black, N. (2006a). A comparison of formal consensus methods used for developing clinical guidelines. *Journal of Health Services Research & Policy*, *11*, 218–224.

Hutchings, A., Raine, R., Sanderson, C., & Black, N. (2006b). A comparison of formal consensus methods used for developing clinical guidelines. *Journal of Health Services Research and Policy*, *11*, 218–224.

Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. *BMJ (Clinical Research Ed.)*, *311*, 376–380.

Kahan, J. P., Park, R. E., Leape, L. L., Bernstein, S. J., Hilborne, L. H., Parker, L., … Brook, R. H. (1996). Variations by specialty in physician ratings of the appropriateness and necessity of indications for procedures. *Medical Care*, *34*, 512–523.

Ketola, E., Kaila, M., & Honkanen, M. (2007). Guidelines in context of evidence. *Quality & Safety in Health Care*, *16*, 308–312.

Klein, N., & Epley, N. (2015). Group discussion improves lie detection. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 7460–7465.

Kramer, M. W., Kuo, C. L., & Dailey, J. C. (1997). The Impact of Brainstorming Techniques on Subsequent Group Processes. *Small Group Research*, *28*, 218–242.

Kuper, A., Lingard, L., & Levinson, W. (2008). *Qualitative Research : Critically Appraising Qualitative Research*. *337*, 687–689.

Landeta, J., Barrutia, J., & Lertxundi, A. (2011). Hybrid Delphi: A methodology to facilitate contribution from experts in professional contexts. *Technological Forecasting and Social Change*, *78*, 1629–1641.

Larson, J., Christensen, C., Franz, T., & Abbott, A. (1998). Diagnosing groups: the pooling, management, and impact of shared and unshared case information in team-based medical decision making. *Journal of Personality and Social Psychology*, *75*, 93–108.

Levay, P., Ainsworth, N., Kettle, R., & Morgan, A. (2016). Identifying evidence for public health guidance: a comparison of citation searching with Web of Science and Google Scholar. *Research Synthesis Methods*, *7*, 34–45.

Mahmoodi, A., Bang, D., Olsen, K., Zhao, Y. A., Shi, Z., Broberg, K., … Bahrami, B. (2015). Equality bias impairs collective decision-making across cultures.

Proceedings of the National Academy of Sciences, *112*, 201421692.

Mcdougal, J. A., Brooks, C. M., & Albanese, M. (2005). Achieving Consensus on Leadership Competencies and Outcome Measures. *Evaluation & the Health Professions*, *28*, 428–446.

Mental Health Taskforce. (2014). *The Five Year Forward View For Mental Health*.

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to Tango. *Personality and Social Psychology Bulletin*, *37*, 1325–1338.

Mobily, P. R., Herr, K. A., & Kelley, L. S. (1993). Cognitive-behavioral techniques to reduce pain: a validation study. *International Journal of Nursing Studies*, *30*, 537–548.

Moritsugu, J., Vera, E., Wong, F. Y., & Duffy, K. G. (2016). *Community Psychology* (Fifth Ed). Oxon: Routledge.

Murphy, M K, Sanderson, C., Black, N. A., Askham, J., Lamping, D. L., Marteau, T., & Mckee, C. M. (1998). Consensus development methods, and their use in clinical guideline development. *HTA Health Technology Assessment NHS R&D HTA Programme Health Technology Assessment*, *2*.

Murphy, MK K, Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment (Winchester, England)*, *2*, i–iv, 1–88.

National Institute for Health and Care Excellence. (2014). Developing NICE guidelines: the manual. *Process and Methods Guides*, p. 245.

National Institute for Health and Clinical Excellence. (2007). *The guideline development process: an overview for stakeholders, the public and the NHS*.

Neal, T. M. S., Guadagno, R. E., Eno, C. A., & Brodsky, S. L. (2012). Warmth and competence on the witness stand: implications for the credibility of male and female expert witnesses. *The Journal of the American Academy of Psychiatry and the Law*, *40*, 488–497.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods*, *16*.

Pagliari, C., Grimshaw, J., & Eccles, M. (2001). The potential influence of small group processes on guideline development. *Journal of Evaluation in Clinical Practice*, *7*, 165–173.

Passannante, M. R., Restifo, R. A., & Reichman, L. B. (1993). Preventive therapy for the patient with both universal indication and contraindication for isoniazid. *Chest*, *103*, 825–831.

Raine, R., Sanderson, C., Hutchings, A., Carter, S., Larkin, K., & Black, N. (2004). An experimental study of determinants of group judgments in clinical guideline development. *The Lancet*, *364*, 429–437.

Rescher, N. (1998). *Predicting the future : an introduction to the theory of forecasting*. Albany: New York Press.

Rycroft-Malone, J. (2001). Formal consensus: the development of a national clinical guideline. *Quality in Health Care*, *10*, 238–244.

Rycroft-Malone, Jo, Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack, B. (2003). What counts as evidence in evidence-based practice? *Nursing and Health Care Management and Policy*, *47*, 81–90.

Rycroft-Malone, Jo, Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack, B. (2004). What counts as evidence in evidence-based practice? *Journal of Advanced Nursing*, *47*, 81–90.

Sandelowski, M., & Barroso, J. (2002). Finding the Findings in Qualitative Studies. *Journal of Nursing Scholarship*, *34*, 213–219.

Schünemann, H. J., Fretheim, A., & Oxman, A. D. (2006). Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. *Health Research Policy and Systems*, *4*, 13.

Shekelle, P G, Woolf, S. H., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: developing guidelines. *BMJ (Clinical Research Ed.)*, *318*, 593–596.

Souder, W. E. (1977). Effectiveness of Nominal and Interacting Group Decision Processes for Integrating R and D and Marketing. *Management Science*, *23*, 595–605.

Steiner, I. D. (1972). *Group process and productivity*. New York: Academic Press.

Stephenson, B. Y., Michaelsen, L. K., & Franklin, S. G. (1982). An Empirical Test of the Nominal Group Technique in State Solar Energy Planning. *Group & Organization Studies*, *7*, 320–334.

Sullivan, J. J. (1978). *An experimental study of a method for improving the effectiveness of the nominal group technique*. University of Florida.

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology*, *8*, 45.

Van de, A. H., & Delbecq, A. L. (1974). The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes. *Academy of Management Journal*, *17*, 605–621.

Van De Ven, A., & Delbecq, A. L. (1971). Nominal Versus Interacting Group Processes for Committee Decision-Making Effectiveness. *Academy of Management Journal*, *14*, 203–212.

Van de Ven, A. H., & Delbecq, A. L. (1974). The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes. *Academy of Management Journal*, *17*, 605–621.

Vinokur, A., Burnstein, E., Sechrest, L., & Wortman, P. M. (1985). Group decision making by experts: Field study of panels evaluating medical technologies. *Journal of Personality and Social Psychology*, *49*, 70–84.

Wegge, J., & Haslam, S. A. (2005). Improving work motivation and performance in brainstorming groups: The effects of three group goal-setting strategies. *European Journal of Work and Organizational Psychology*, *14*, 400–430.

West, A. (2011). Using the delphi technique: Experience from the world of counselling and psychotherapy. *Counselling and Psychotherapy Research*, *11*, 237–242.

Williams, P. L., & Webb, C. (1994). The Delphi technique: a methodological discussion. *Journal of Advanced Nursing*, *19*, 180–186.

Winquist, J. R., & Larson, J. R. . J. (1998). Information pooling: When it impacts group decision making. *Journal of Personality and Social Psychology*, *74*, 371–377.

Wittenbaum, G. M., Hubbell, A. P., & Zuckerman, C. (1999). Mutual enhancement: Toward an understanding of the collective preference for shared information. *Journal of Personality and Social Psychology*, *77*, 967–978.

Wittenbaum, G. M., & Stasser, G. (1996). Management of Information in Small Groups. In *What's Social about Social Cognition? Research on Socially Shared Cognition in Small Groups* (pp. 3–28). United States: SAGE Publications, Inc.

Woolf, S. H., Grol, R., Hutchinson, A., Eccles, M., & Grimshaw, J. (1999). Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines. *BMJ (Clinical Research Ed.)*, *318*, 527–530.

Woudenberg, F. (1991). An evaluation of Delphi. *Technological Forecasting and Social Change*, *40*, 131–150.

Yang, H. (2013). Determining the effectiveness of the delphi method for quantifying the drivers of demand for health and social care. *University of Southampton, School of management, Masters Thesis, 77pp.*

Yousuf, M. I. (2007). Using Experts` Opinions Through Delphi Technique. *Practical Assessment, Research & Evaluation*, *12, 4, 1-8.*

Part 2: Empirical Paper

A mixed methods study of experiences of informal group consensus methods and expectations of using the Nominal Group Technique in a real healthcare guideline committee and technical team

Abstract

Aims: This is a feasibility pilot study that aimed to capture committee member and technical team experiences of current informal consensus practices, previous experiences of formal consensus methods, and expectations for the planned use of the Nominal Group Technique (NGT) formal consensus method.

Method: Twelve participants, including committee and technical team members across two guideline groups engaged in semi-structured interviews before using NGT. All of the committee members of one guideline group then answered Likert-scale questions about their experience after using NGT.

Results: Themes were extracted from the interviews and corroborated by the quantitative data. Themes included: Formal consensus (credibility, effort and resource intensiveness), Methodology (guideline interpretation and implementation, interpretation of evidence, and the restrictiveness of NICE process), Group processes (management of expertise, anonymity, leadership, and discussion), and Continuity of group members. Data were further analysed in the context of participant professional background.

Conclusions: Participants identified beneficial elements across formal and informal consensus approaches, and it is likely a hybrid of methods is best suited to healthcare guidelines given their task of combining diverse and complex knowledge to achieve specific guidelines. The results are interpreted in the context of theory and recommendations are made on the future use and conduct of consensus methods.

Introduction

Guideline development in the field of healthcare requires the individual and group interpretation of research evidence by stakeholders, which leads to a range of viewpoints (e.g. Dopson, FitzGerald, Ferlie, Gabbay, & Locock, 2002). The National Institute for Health and Care Excellence (NICE, 2014) invites stakeholders from varied professional backgrounds to committee meetings. Inviting the stakeholders that are affected by guidelines could not only be considered a professional and ethical obligation by NICE, but it could also increase the credibility and uptake of clinical guidelines. In addition, research shows that participation in group decision-making increases feelings of fairness, acceptance, and implementation by those involved (Moscovici & Doise, 1994).

The emphasis on group collaboration for guideline development requires consideration of the challenges of group decision-making. It has been shown that group decision-making is not necessarily superior to that of individuals(e.g. Wu & Seidmann, 2015). This is due a number of group process characteristics that can present themselves. Groups can exhibit a failure to appreciate “sunk costs”. This means that groups tend to persist with clearly defective projects. Thus, groups have an aversion to abandoning a defective project early on in the process to incur a smaller loss and avoid a larger future loss (Smith, Tindale, & Steiner, 1998). Groups can also present with “conformity pressures”, which is illustrated by group members acting in agreement with a majority view despite having personal beliefs to the contrary (Asch, 1952). A process known as “social loafing” can also occur, which is described by a reduction in the effort made by individual group members to opt instead for other members of the group to carry the workload (Henningsen, Cruz, & Miller, 2000).

A significant contributing factor to poor group decision-making is information sampling and confirmation biases (Nickerson, 1998; Stasser & Titus, 1985, 1987). The authors assert that the primary task for group decision-making is to reach consensus, which is dependent on effective information exchange between group members. The total knowledge of the group should be greater than that of each individual group member. This will be most optimal when members each hold unique information that is relevant to the task. Therefore, the sharing of this information would provide new information to the other group members. Such knowledge is referred to as previously "unshared" information. Groups have been shown to favour the discussion of shared versus unshared information (e.g. Larson, Foster-Fishman, & Keys, 1994; Stasser, Stewart, & Wittenbaum, 1995; Winquist & Larson, 1998). Therefore, decision-making is generally more likely to be influenced by shared rather than unshared group decision-making (e.g. Stasser & Stewart, 1992).

Several reasons for this bias towards shared information have been proposed. It could be that shared information simply has a higher probability of being discussed due to the increased likelihood of recall and introduction by group members (Stasser & Stewart, 1992; Stewart & Stasser, 1998). Another reason could be the drive for social validation and support, which is more likely to be the case if others confirm and recognise the information being discussed (Parks & Cowlin, 1996), and also of mutual enhancement, where individuals are socially reinforced by others for exchanging shared information (Wittenbaum et al., 1999).

Research has shown that offering unshared rather than shared information can support higher decision quality in a murder mystery task (Galinsky & Kray, 2004),

medical doctor scenario-based consultation (Larson, Christensen, Franz, & Abbott, 1998) and person profiling tasks (Winquist & Larson, 1998). However, this is limited by biases that occur once unshared information is presented, since individuals will tend to filter information based on initially preferred solutions (Brownstein, 2003). Indeed, the collective majority view of individuals prior to a meeting is usually predictive of the group decision following discussion (Gigone & Hastie, 1997). Group members have been shown to be more likely to present unshared information if it agrees with their initial preference (Dennis, 1996), and bias information that is revealed on the basis of their preferences (Edwards & Smith, 1996; Greitemeyer & Schulz-Hardt, 2003).

To mitigate some of the impact of group decision-making process losses, several approaches have been suggested. These have been divided into four categories: the use of procedural structure, effectively managing the task, effectively managing interpersonal dynamics, and offering sufficient training (Beranek, Beise, & Niederman, 1993).

Structured procedures have been found to work best when clear instructions are provided to group members (Hall & Watson, 1970), time is taken to properly formulate the problem (Volkema, 1983) and generate ideas (Ball & Jones, 1977; Bowman & Wittenbaum, 2012), when idea sharing and evaluation phases are separated (Camacho & Paulus, 1995; Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter, & Frey, 2006; Smith, 1998; Van de & Delbecq, 1974), and when the selection of a final decision is delayed (Hoffman, 1979; Sutton & Hargadon, 1996).

The task behaviours that have been shown to be most helpful are explicit discussion of task procedures (Hackman & Kaplan, 1974), having specific criteria for

discussion and solution identification (Hirokawa & Pace, 1983; Yearwood & Stranieri, 2010), using factual evidence (Bang & Frith, 2017; Hirokawa & Pace, 1983), and focusing on the goals of the exercise (Dalkey & Helmer, 1963; Durham, Locke, Poon, & McLeod, 2000).

The third area of effective interpersonal management is comprised of encouraging participation (Hoffman & Maier, 1959; Stasser, 1992), deploying a constructive resolution of conflict (López, 2004), encouraging consensus rather than simply using majority vote (Hall & Watson, 1970), the use of active listening (Sypher, Bostrom, & Seibert, 1973), and discussion of the interpersonal processes themselves (Hackman & Kaplan, 1974; Maznevski, 1994).

Finally, there is evidence to show that training is important, and that training improves outcomes whether it is for leaders, group members, or external facilitators (Chen, Sicrar, Hwang, & Hwang, 1998; Hall & Williams, 1970; Maier & Maier, 1957; Miner, 1979).

There have been specific interventions developed to be used within group discussion to enhance information discussion, including the encouragement of dissent using techniques such as devil's advocacy or dialectical inquiry (Schulz-Hardt et al., 2006). However, a more comprehensive approach is needed to account for the broad range of procedural, social, and intellectual variables that influence group decision-making processes.

Formal consensus methods as a solution for group process losses

Formal consensus methods combine many of the factors that enhance group decision-making in order to manage some of the process losses found in groups (Murphy et al., 1998; Pagliari et al., 2001). Participation is actively encouraged through individuals simultaneously and privately recording their ideas. This avoids the potential for individual contributions to be limited by conversational turn-taking, and allows for reduced social pressure due to the potential for anonymity (Postmes & Lea, 2000). The structure of formal consensus methods also separates the voting and discussion phases (Black et al., 1999).

A further strength of formal consensus methods is the transparent recording of the decision-making process of the group. This is particularly useful for healthcare guidelines since they require the incorporation of knowledge from a variety of sources, which includes clinical expertise, economic considerations, and research. Using non-factual information is not conducive to good quality decision-making, as previously discussed (Hirokawa & Pace, 1983). Therefore, it is important that the sources contributing to decision-making processes and outcomes are explicitly recorded. Because of this, formal consensus methods could be considered helpful for use in healthcare guidelines where there is often limited or low quality research evidence (Murphy et al., 1998).

Group members must interpret a range of evidence to develop healthcare guidelines. Unlike informal consensus methods, formal consensus methods allow for the systematic recording and feedback of the decision-making process. Thus, the factors that contributed to a particular decision, for example how a specific view was developed and how many committee members endorsed it, can be transparently

described. Consequently, different sources of information, such as clinical expertise, can then undergo the same evaluation as any research findings used, which has been argued is important for evaluating the impact and utility of clinical expertise (e.g. Stetler et al., 1998).

Common types of formal consensus method

Formal consensus methods include the NGT, UCLA/RAND Appropriateness method (RAND), and Delphi methods. The procedure from the original paper of NGT (Van de Ven & Delbecq, 1974) is as follows: Firstly, statements are generated by the committee through a individual writing task. Secondly, these statements are presented in turn by each individual, without discussion. Ideas are summarised and written on a board. Thirdly, ideas are discussed and evaluated. Fourthly, there is another round of silent and independent individual voting. The final result is the pooled votes of all the group members.

RAND (Fitch et al., 2001) has been considered to be a form of NGT whereby a committee of experts are brought together to answer healthcare decision-making questions. The members develop scenarios that manipulate the critical factors or cues relevant to intervention decision-making. The experts rate the chosen interventions on a nine-point scale ranging from extremely appropriate to extremely inappropriate. Ratings are initially done privately by individuals before aggregating the votes together for discussion as a group. Next, participants re-rate the scenarios privately and individually without discussion. The appropriateness score is defined as the median rating.

The Delphi technique (Rescher, 1998; Yousuf, 2007) begins with a

questionnaire that is developed by a researcher or facilitator. The questionnaire is comprised of broad questions around a topic of interest. This is posted or emailed to participants who are asked to review and return their answers. These statements are collected and summarised by the facilitator, who then sends these out to be voted upon by participants, with an option for participants to give feedback regarding the rationale behind their responses. The facilitator gathers the votes and sends each individual their own score alongside the pooled vote of the group, which also includes their comments. Participants then vote again on the statements. This is usually repeated until a predefined level of consensus is reached, most commonly after two or three rounds (see Rowe & Wright, 2011 for review).

NGT is currently being employed and evaluated by NICE across a number of guidelines in development such as in the area of mental health problems and learning disabilities (NICE, 2016). NGT could be considered as the most appropriate method of formal consensus for guideline development for the following reasons. Firstly, it allows for the efficient use of time. It can be completed in a few hours in the presence of the entire committee. Secondly, it is flexible in terms of content of included data, which is important for healthcare guidelines that often draw upon and need to be applied to multiple sources. The NGT has been particularly recommended for areas with low quality or contradictory evidence because it enables the incorporation and application of different types of knowledge, unlike the scenario-based RAND (Campbell & Cantrill, 2001; Jones & Hunter, 1995b). Authors have also suggested that NGT is well suited to answering complex questions, unlike the Delphi method which focuses more on forecasting (Kopp, Selbmann, & Koller, 2007).

The credibility of NGT has been approached from multiple angles by research. Studies have compared NGT outcomes with the empirical outcomes from healthcare research. Cohen's kappa (Cohen, 1960) is generally used as a statistic to understand the extent of agreement between outcomes with the consideration of the possibility that they may agree by chance. These studies have found "moderate" overlap or 68-71% agreement when comparing a RAND formal consensus method with a systematic review of the literature (Nicollier-Fahrni, Vader, Froehlich, Gonvers, & Burnand, 2003; P. Wortman, Smyth, Langenbrunner, & Yeaton, 1998). NGT and Delphi have shown between moderate and substantial agreement, with kappa statistics ranging from 53% to 76% agreement between separate expert panels when they rate the same clinical scenarios (Tobacman, Scott, Cyphert, & Zimmerman, 1999; Washington et al., 2003). The reliability of NGT and Delphi has been measured by giving different expert panels the same topic questions and assessing for the agreement between them. For NGT, this agreement has been reported as moderate or good, ranging between 45-83% between different groups of experts on the same topic (Coulter, Marcus, & Freed, 1998; Eriksen et al., 1996; Paul G. Shekelle et al., 1998) Research has reported that agreement between Delphi groups is "good" with a range of 62-96% (Kastein, Jacobs, van der Hell, Luttik, & Touw-Otten, 1993).

Hutchings, Raine, Sanderson, and Black (2006) compared Delphi and NGT methods across multiple healthcare groups, which comprised of 213 general practitioners in total. Their task was to rate the appropriateness of four treatments (brief psychodynamic interpersonal therapy, cognitive behavioural therapy, behavioural therapy, and antidepressants). When compared to Delphi, NGT was found to produce more within-group agreement and more shifts in group members'

views. NGT participants were also more likely to rate treatments as favourable. Delphi groups were shown to be more consistent in their voting between rounds of meetings than NGT groups. The authors suggested that reliability could have been reduced for NGT partly because of its smaller size compared to Delphi groups, and also because of the increased likelihood of group shifts in views depending on exposure to persuasion and social dynamics. Another view, taken by some authors, is that the aim of formal consensus methods is not necessarily to implement a positivist scientific inquiry, but rather a hermeneutic reflective interpretation of complex inputs as seen through the lens of experts, and therefore quantitative measures of validity and reliability are inappropriate (Guzys, Dickson-Swift, Kenny, & Threlkeld, 2015). It could be that a better form of validity and reliability would be characterised by the results and utility of outputs from formal consensus methods. This could encompass, for example, how representative the produced statements are of the key features of the topic area, or how well the produced statements can be applied and implemented beyond the group meeting in line with the overarching group goals.

Participant experiences as an important measure of formal consensus methods

For healthcare guidelines to be implemented, it is important that they are perceived as credible, and this is one of the tasks of the multidisciplinary committee (Graham & Harrison, 2005). Therefore, it is important to investigate the perceptions and experiences of committee members. This allows for the evaluation of qualitative content validity in addition to the previous approaches that have researched the reliability and validity of formal consensus methods from a quantitative viewpoint. There has been a single study that captured the experiences of healthcare guideline

members using NGT (Bini & Mahajan, 2016). In this paper, participants rated NGT as effective overall for reaching consensus. Individuals also reported NGT as enhancing their understanding of the evidence base, and that they felt it enabled their own and the opinions of others to be heard. Individuals also said that using NGT increased the likelihood that they would adopt the guideline recommendations.

Despite the assumption that a major strength of NGT is increased participation of group members (e.g. Murphy et al., 1998), there has been no research investigating whether this is indeed a priority for healthcare committee members, and what their expectations are in terms of NGT being an appropriate or satisfactory enhancement to more traditional informal consensus methods. There has been much more focus in the evidence base on the quality of decisions as measured by external quantitative factors, and on the feasibility of the use of formal consensus methods as described above. There has also not been any investigation regarding whether professionals from different backgrounds experience guideline decision-making differently, despite there being evidence that this is the case in the multidisciplinary team healthcare literature for group decision-making (e.g. Lanceley, Savage, Menon, & Jacobs, 2008), and that there are individual differences in response to formal consensus methods based on social preferences (Hirokawa, Ice, & Cook, 1988; McCroskey, 1977).

The present study is a feasibility pilot study of the use and evaluation of NGT in a NICE healthcare guideline. Committee members and technical staff from a NICE healthcare guideline were interviewed before using NGT. The primary aim of the study was to capture participant expectations of using the NGT method. However, it was understood that participant expectations would be situated in their

past experiences of formal and informal consensus methods. Therefore, more general questions regarding formal and informal consensus method experiences were included as part of the data capture. Data was collected through the use of qualitative semi-structured interviews and a quantitative questionnaire. The data were further examined for differences in views between healthcare and technical team members.

Method

Participants

Participants included committee and technical team members. They were recruited from two NICE healthcare guideline groups. Committee members are typically recruited to be part of a guideline group following an interview selection process. They are recruited due to being experts in the guideline topic area and tend to include doctors, nurses, other health professionals, patients, and carers. Technical team members consist of experts in synthesising evidence such as systematic reviewers, health economists, and guideline leads.

Twenty-one people across the two guidelines consented to being interviewed. They were grouped by professional type (medical doctor, non-medical, lay-member, chair, systematic reviewer, and guideline lead) and allocated a number. The two guideline groups were identified as either group “A” or “B”. An online random number generator (https://www.random.org/) was used until fourteen participants were selected across the two guideline groups. These people were then approached to complete a qualitative interview.

Seven people were not able to be interviewed due to being unavailable

within the study timetable, and these were replaced with additional participants identified by the random number generator. A final twelve members were interviewed of the nineteen initially contacted. Doctors included GPs and consultants. The term “non-medical” was used for other healthcare staff for the purpose of the study. This group consisted of professionals that had not completed a medical degree, and included nurses, biochemists, pharmacists, and dietitians. An overview of the demographic details of the participating and non-participating members from the qualitative and quantitative parts of the study can be found in tables 1 to 3. It is of note that there was an over-representation of white ethnic background in the dataset as compared with the demographic of healthcare staff. However, this reflected the demographic of the overall guideline.

Demographics for the participants included in the quantitative phase of the study are displayed by the data in Table 4. These data were from only one of the two guideline groups because of limitations within the study timeline. All the participants on the guideline committee participated in the quantitative questionnaire. The participants that used the NGT were defined as the committee members, and so the technical team did not complete the questionnaire and were not included as part of the quantitative data.

Table 1: Demographics of included participants for qualitative data

Code	Age	Gender	Ethnicity	Profession	Prior Formal Consensus Experience
A3	60s	Male	White British	Medical Consultant	None
A7	60s	Male	White British	Pharmacist Non-medical	Previous experience once of Delphi method
A10	40s	Female	White British	Dietitian Non-medical	Previous experience once of Delphi method
A13	30s	Female	White British	Lay-member	None
A15	30s	Female	White Other	Guideline Lead	Previous multiple experiences of Delphi method
A18	30s	Female	White Other	Systematic Reviewer	None
B5	60s	Male	White British	Chair/GP Medical	Previous experience once of Delphi method
B11	40s	Female	White British	Midwife Non-medical	None
B16	40s	Female	White British	Lay-member	None
B20	30s	Female	White British	Systematic Reviewer	None
B21	30s	Female	White Other	Systematic Reviewer	Previously used NGT once
B22	40s	Female	White Other	Health Economist	Previous multiple experiences of using NGT

Table 2: Demographic data of non-responder participants for qualitative data

Code	Age	Gender	Ethnicity	Profession	Stated Reason
A4	50s	Female	Asian	Medical doctor Consultant	Did not respond to invitation and follow-up email
B13	30s	Female	White British	Lay-member	Did not respond to invitation and follow-up email
B19	40s	Female	White British	Senior Systematic Reviewer	Was on leave
B7	50s	Female	White Other	Medical doctor Consultant	Did not respond to invitation and follow-up email
B8	60s	Male	White Other	Medical doctor Consultant	Did not respond to invitation and follow-up email
A2	50s	Male	White British	Medical doctor Consultant- Chair of Guideline	Responded too late to participate
B19	60s	Male	White British	Medical doctor Consultant- Clinical Advisor to Guideline	Unable to find mutually convenient time with researchers

Table 3: Demographic data of participants included in quantitative questionnaire data

Code	Age	Gender	Ethnicity	Profession
A2	50s	Male	White British	Medical doctor Consultant and chair to guideline
A3	40s	Male	White British	Medical doctor Consultant
A4	50s	Female	Asian	Medical doctor Consultant
A5	60s	Male	White British	Medical doctor Consultant
A7	60s	Male	White British	Pharmacist Non-medical
A21	60s	Male	White British	Biochemist Non-medical
A8	40s	Female	White British	Pharmacist Non-medical
A20	40s	Female	White British	Medical doctor Consultant
A9	40s	Female	White British	Dietitian Non-medical
A6	40s	Male	Asian	Medical doctor Consultant
A12	30s	Female	White British	Lay-member
A11	40s	Female	White British	Nurse Non-medical
A13	30s	Female	White British	Lay-member
A10	40s	Female	White British	Dietitian Non-medical

Ethical Approval

Ethical Approval was granted by the University College London Research Ethics Committee (project number CEHP/2018/569, see Appendix J). Participants were briefed about the study and given consent forms and information sheets (see

Appendix E and F) for the qualitative and quantitative parts. Members were invited to comment on their concerns. A few members expressed thoughts about the preservation of anonymity. This was of particular concern because of several reasons. The guideline group was a relatively small group with some roles allocated to a single person (e.g. chair), and the membership of these roles were published along with the guideline. Therefore, consultation was sought to ensure the anonymity process was satisfactory. Actions that resulted from the consultation included the use of aliases when analysing and presenting the data, and refraining from disclosing the specific topic area of the guideline. There was also an agreement made that participants could view the final report including quotations, and the right to withdraw their data within a given timeframe.

Procedures

The researcher delivered a presentation that provided an overview of the study and formal consensus methods at a guideline committee meeting. Following this, written consent to take part in the qualitative component of the study was sought from individuals (see Appendix E for consent form). Participants who consented were followed up with an email inviting them to a telephone or in person interview lasting 30-40 minutes. These interviews consisted of open-ended semi-structured questions regarding their experiences of decision-making in guideline committees and expectations of using formal consensus methods (see Appendix G for copy of interview schedule). There was no compensation for time in the study.

The Nominal Group Technique (Bernstein et al., 1992) procedure is established as part of NICE methodology guidance (National Institute for Health and Care Excellence, 2014). The established procedure is that the technical team generate

consensus statements from sources of evidence that are agreed upon within the committee. Selected topic advisors are available to support the technical team with the statements. These statements are then presented to the committee for anonymous written voting and feedback. There is also the option for committee members to not offer a response to a statement should they wish to opt out of a particular section. Responses are then collected for analysis, those above 80% agreement are kept for presenting to the final guideline, and those below 60% agreement are discarded. Statements with 60-80% agreement are revised depending on the comments received and presented back to the committee for clarification and shared revision. There is then a second round of voting. The final statements and results are then presented and discussion ensues to develop the finalised recommendations.

Design

Measures.

Interviews. The qualitative interviews were conducted prior to the NGT session to better understand the expectations of participants about using NGT and their prior experiences of using formal consensus methods. The interview schedule consisted of semi-structured questions focused on the aims of the study. Namely, what participants' views were on decision-making in committees, what their experiences had been of formal consensus methods so far, what their expectations were for their implementation generally and in these specific guidelines, and how individuals thought formal consensus methods were viewed by themselves and their colleagues. The interview schedule can be found in Appendix G.

The interview questions were developed drawing on the literature review

within this thesis as a basis for some of the questions. More specifically, the interview questions were informed by a combination of the Likert data, open-ended questions, and the resultant themes derived from previous studies investigating participant experiences of formal consensus methods (De Ruyter, 1996; Gresham, 1986; Kramer, Kuo, & Dailey, 1997; Landeta, Barrutia, & Lertxundi, 2011; Mcdougal, Brooks, & Albanese, 2005; Shekelle & Schriger, 1996; Stephenson, Michaelsen, & Franklin, 1982; West, 2011). Consultation was also sought from technical team and committee members of other guidelines to support the development of the questions. Finally, supervision was used as a space to discuss how the questions related to the theory and research question. A semi-structured format was selected to support engagement and data capture. The interview schedule was developed over a series of pilot interviews and meetings with members of the research team who also worked on clinical guidelines. Refinements consisted of clarifying the questions and also identifying suitable interview prompts.

Likert-scales. The Likert-scales were used during the NGT session to capture the current experiences of participants using the NGT. Items were both positively and negatively worded to reduce response bias. The scales required a response on a 7-point Likert scale. High scores indicated high levels of the variables. Respondents were asked to compare the NGT meeting to other committee meetings that used informal consensus. Additional space for comments to supplement the rating was provided, and participants were asked to indicate their background (chair, doctor, non-medical, lay-member, or other). Currently there is no validated standard instrument to measure the experiences of committee members using consensus methods. Therefore, the questionnaire used in the present study was influenced by several previous studies (Graefe & Armstrong, 2011; Bini and Mahajan, 2016) that

captured participant experiences in the context of clinical guidelines. A diagrammatical explanation of the procedures of the study can be found in Figure 6.

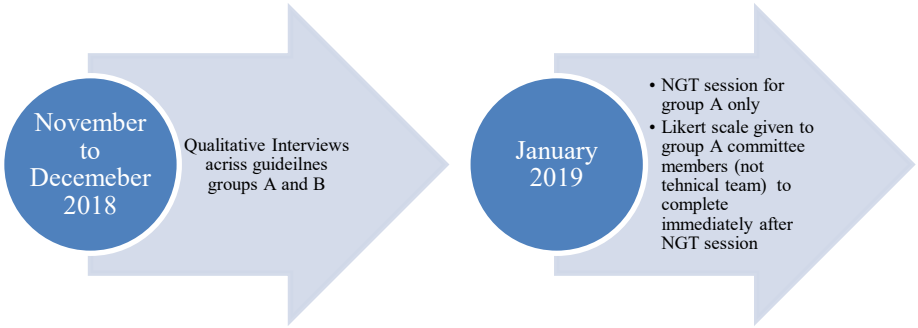


Figure 1: Figure to depict the chronological procedure of the study

Qualitative Analysis. Interview data were transcribed for analysis. Thematic Analysis (TA; Braun & Clarke, 2006) was used to develop individual and overarching themes from the data. The researchers (VR and PB) familiarised themselves with the data by reading through all of the transcripts, and these transcripts were uploaded to NVIVO, qualitative analysis software (QSR International Pty Ltd., 2018) and initial codes were generated. Codes were created that closely described the individual units of data rather than applying any higher-order categorisation. Themes that linked codes were developed to form higher-order meanings.

Meaning was partly informed by previous theory and research regarding experiences of formal consensus methods. For example, equality of participation was an important theme from previous research and was therefore considered likely to be included as a theme. In the present research this aspect of NGT seemed to be

expressed in relation to managing diverse expertise in the group. Thus, similar codes were grouped together under more general themes. It became apparent that some of the themes were not relevant to the research aims, and these were therefore discarded. For example, participants at times used the interview to ask the researcher about the practicalities of implementing the NGT, which was not relevant to their expectations. Themes were reviewed and redefined in an iterative process.

Credibility Checks. Elliott, Fischer, & Rennie (1999) provide guidance for qualitative research, which was used as guidance for the implementation of the current study. As recommended, the participants have been situated within their context by describing their demographic characteristics (Barker, Pistrang, & Elliott, 2002). Themes were also grounded in the data by offering multiple quotes per theme. A bracketing interview was carried out with another researcher to explore potential researcher biases, and a consequent subjectivity statement was written by the researcher (Preissle, 2008). Bracketing describes the process by which a researcher renders their expectations for their findings explicit and thereby attempts to approach the data as separate from their own assumptions with "fresh eyes" (Smith, Flowers & Larkin, 2009).

Two researchers (VR & PB) provided inter-rater reliability checks of themes. They coded the data separately and met to discuss the themes generated. This was an interactive process and occurred four times to fully develop an initial list of themes. Next, all of the data was re-coded with the new agreed theme structure. 50% of the transcripts were double coded and achieved an inter-rater agreement of 85% thus enhancing analysis credibility (Baker & Pistrang, 2005).

Subjectivity Statement. I am a female trainee clinical psychologist in my

early thirties. I believe that formal consensus methods can enhance group decision-making for healthcare guidelines, and that they can particularly give voice to lay-members who may feel less confident about their opinions. I also have the opinion that formal consensus methods pose significant resource challenges for the technical team, and so therefore that it may not be embraced favourably. I have attempted to "bracket" my assumptions (Barker, Pistrang, & Elliott, 2002) during the research, which is a process I have supported through the use of supervision and data checking.

Quantitative Analysis. Statistical analysis was not considered appropriate for the quantitative data. However, the quantitative results were used in the context of the qualitative data to support an exploration of the findings.

Results

Qualitative Data

The data analysis produced four groups of nine overarching themes, with nine subthemes. These are displayed in Figure 1. The responses were also stratified by professional background. The data are presented initially for all participants, then delineated by profession, and finally in more detail with example quotations.

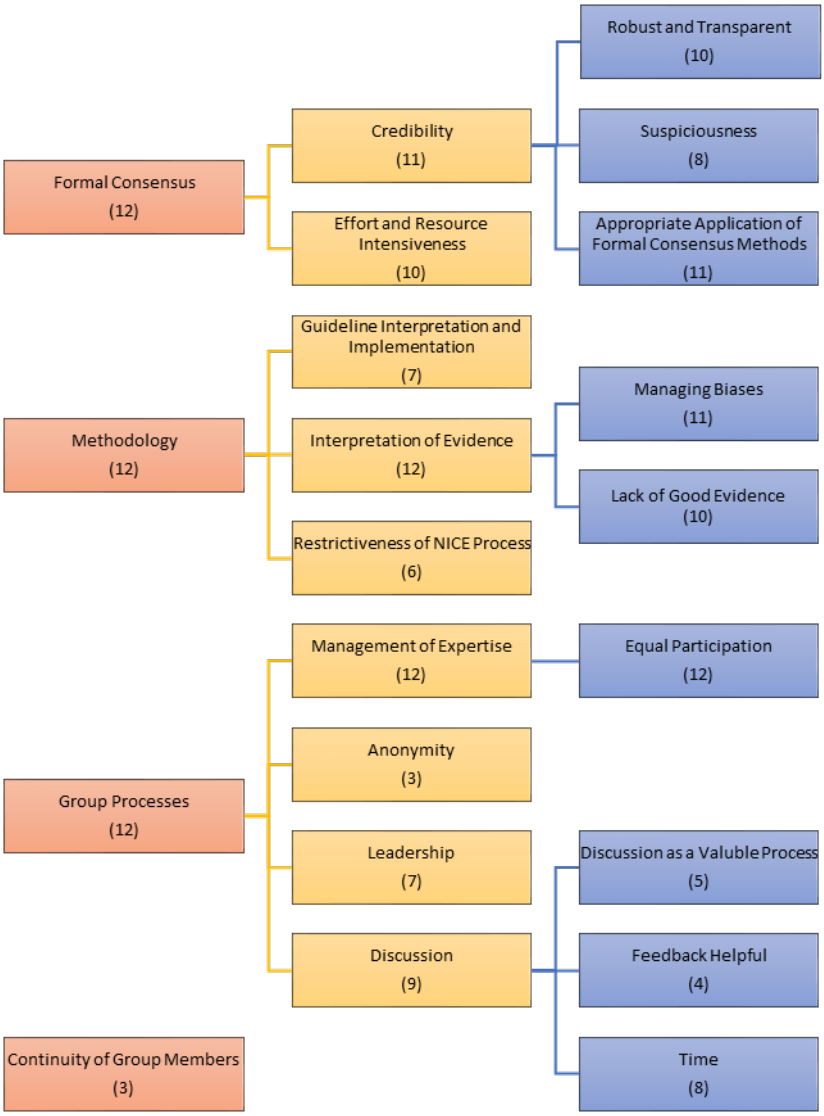


Figure 2: Themes developed from qualitative interviews. The numbers of participants endorsing themes are shown in brackets.

Group 1: Formal consensus

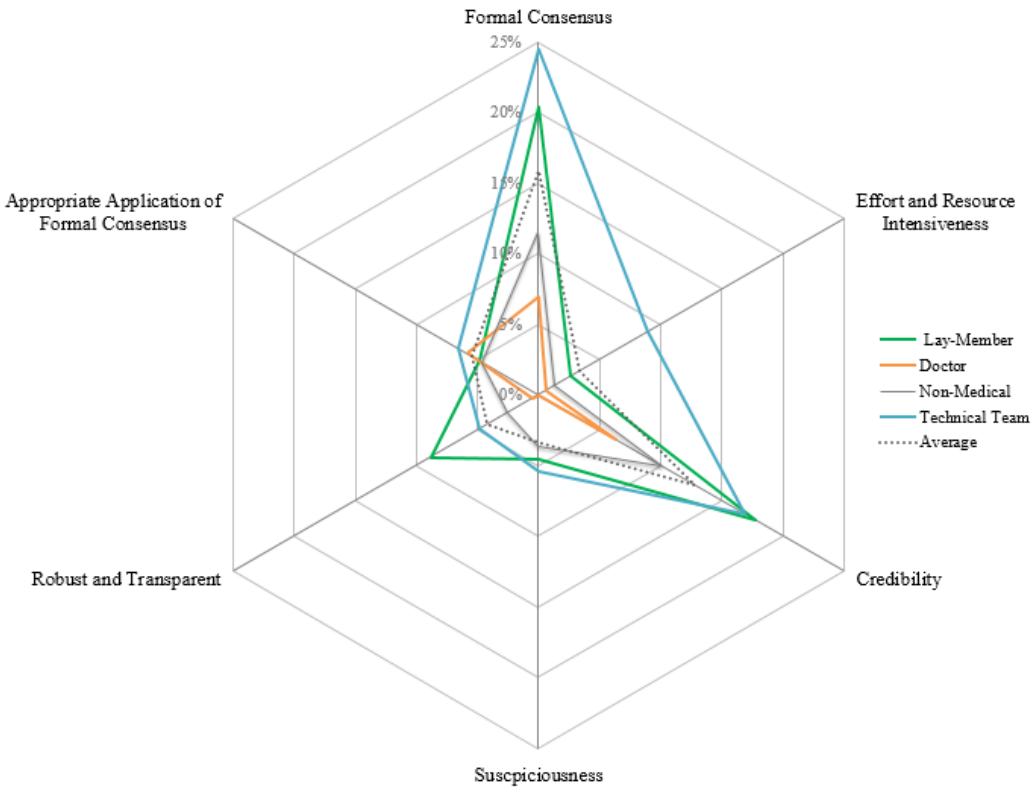


Figure 3: Chart showing percentage endorsement within cluster of themes of formal consensus of each theme across profession.

The first group of themes related to formal consensus methods. Participants seemed to be balancing the additional resources required for formal consensus methods with their supplementary value. Figure 2 shows the percentage spread and the average endorsement across the themes and professional backgrounds of the participants. Members of the technical team referred to formal consensus methods the most in their interviews, and seemed most concerned by the resource intensiveness.

The credibility of formal consensus methods was the largest theme mentioned by participants across all professional backgrounds. Both the technical team and lay-members spoke the most about feeling suspicious about formal consensus methods,

and that robustness and transparency were important. Lay-members were most concerned about robustness and transparency. The technical team raised most concerns about suspiciousness, particularly how useful the formal consensus method was going to be.

Credibility, robustness, and transparency. The credibility of formal consensus methods seemed to be important to participants. Interviewees often spoke about the importance of the formal consensus methods as being robust and transparent.

> *"I would hope that the formal method will work better in terms of us seeing how we're arriving at the decision". B5*

> *"...I think consensus would be extremely useful, so that there's a formal way of everybody contributing something, and then distilling it down so that you have something that is reasonable. Because you can't just say, "We're not doing to make a recommendation because there's no evidence" A10*

This was both in terms of reporting the formal consensus method as well as the process of decision-making.

> *"What I've read of consensus methods, although the process is transparent, they're perhaps not transparently reported in publications". A18*

Credibility and suspiciousness. In addition to this theme was a sense of suspiciousness not only about guideline committee decision-making, but also relating to formal consensus requiring a relinquishment of control of the outcome to a formalised structured process.

> *"...thought that's a very strong statement and everyone agrees with it,*

and there's one person that doesn't. But actually that one person's voice could be really important… even though he might have greater expertise than the rest of us who are voting…Do think that there's a lot of suspicion about the way in which closed groups come to decisions and things like that, and I actually quite like the openness of formal consensus in some ways, that you can say, in the absence of clinical data this is the robust way in which we have come to a decision as a group; we haven't just sat there and chit-chatted over tea and biscuits". B16

Interviewees seemed to be particularly concerned about the output of formal consensus methods being of a low quality, specifically in terms of statements being vague, as mentioned by A15 below.

"Also I think the Delphi statements are sometimes also quite vague. They kind of end up being the smallest common denominator to something" A15

Credibility and appropriate application of formal consensus methods. Limiting the use of formal consensus seemed to be important to participants. All participants felt that formal consensus methods could be best used when evidence was limited.

"I'm not against it, I think it's good. It just needs to be more clearly defined and it needs to be sort of like- I still think it should be the exception rather than the rule". A15

"I think that for me it would seem the most reasonable circumstance to use formal consensus would be where there's just a dearth of data there's nothing clinical that we can hang something on, and so actually what we need is some way of demonstrating that we've been through a robust process". B16

Effort and resource intensiveness. There was discussion around the additional effort and resources required to use formal consensus methods. Participants made reference to increased time for implementation and learning a new process. Some members expressed concern about the increased responsibility of the

technical team of systematic reviewers in developing the initial consensus statements to be reviewed, whereas others talked about it being interesting to learn a new skill.

> *"I'm nervous that the statements are going to be good enough and accepted by the committee. I've done it using the protocol and the processes, but again I'm new to consensus, the topic area. I'm not a clinician, so it's a new topic area. I don't have the same expertise clinicians would have…From a personal perspective, it's definitely very interesting. It gives me a break from doing systematic reviews… From a personal perspective as a team member, it's been a really good experience so far. Very interesting". A18*

Some participants felt that the added benefits of using NGT were not justified by the additional resources required.

> *"I wasn't entirely convinced it was necessarily…it was more for robustness and having a system of coming to agreement." A15*

Group 2: Methodology

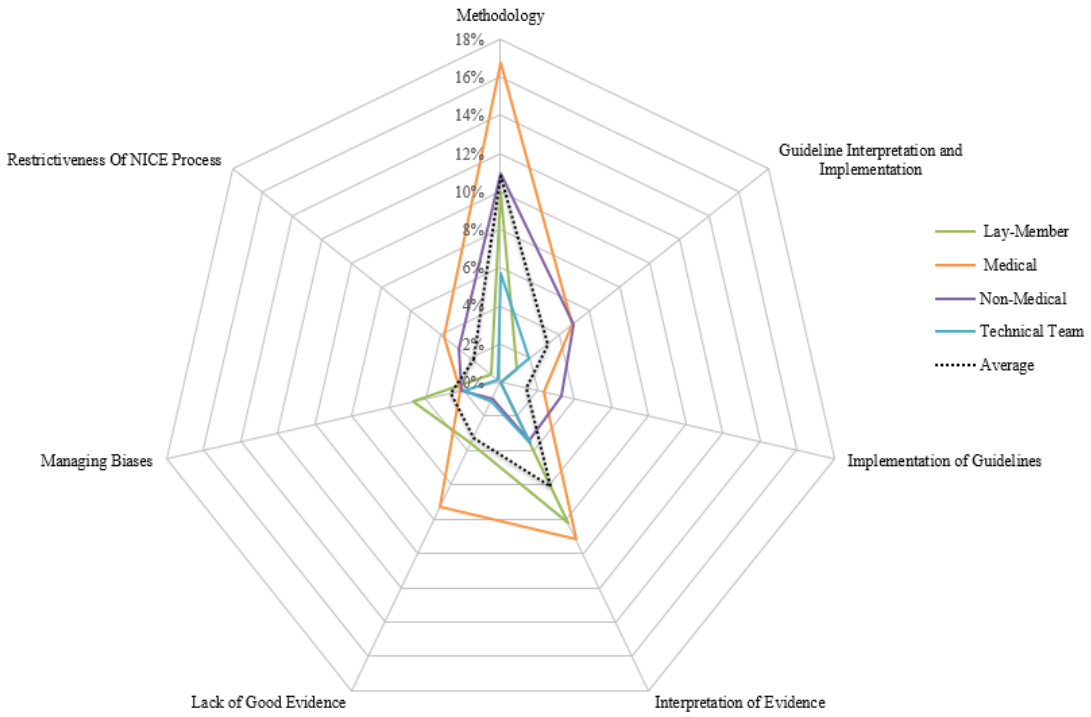


Figure 4: Chart showing percentage endorsement within theme of methodology of each subtheme across profession

Methodology of guideline formation seemed to be an important group of themes from the interviews. This group of themes included the implementation and interpretation of guidelines, the interpretation of evidence, and the restrictiveness of the NICE process. The professional group of doctors in the committee spoke most about methodology, and the technical team referred to it the least. The doctor and non-medical members referred to the restrictiveness of the NICE process most often, in contrast to the technical team and lay-members who mentioned it much less. Lay-members and doctor staff made the most reference to the interpretation of evidence, and non-medical members the least. They diverged from this when considering the context of evidence interpretation, with lay-members speaking more about managing

biases, and doctors talking more about there being a lack of good evidence.

Methodology and the interpretation and implementation of guidelines. Guideline use was an area that participants spoke about as being important. This was divided up into how a guideline that used a formal consensus method might be interpreted and how it might be implemented. Despite clinicians (doctors and non-medical members) most often endorsing methodology, the reasons they did so seemed different to that of the technical team members. There was a sense that clinicians prioritised methodology as a means to enhancing the utility of a guideline whereas the technical team would be more concerned with the methodology behind deriving recommendations as a way to improve the trustworthiness of a guideline in terms of perceived validity and reliability.

> *"From a methodological perspective, I think yes, because I would assume that a guideline is more valid if it has used a formal process as opposed to an informal process. But I am not sure that view is shared by clinicians". A18.*

Participants often linked the implementation of guidelines to the use of formal consensus methodology because it allowed for a more complete guideline that was easier to follow.

> *"If you put nothing in, then it makes it difficult to implement the whole guideline… if you can actually put that there was some sort of vigorous method behind how you looked at something that isn't evidence based then that's more likely to be accepted". N7*

Methodology, and the interpretation of evidence through biases. Participants discussed issues around the interpretation of evidence. There was discussion around how to manage the biases of individuals when interpreting data.

One of the ways people spoke about managing biases was through using evidence.

> *"If the evidence is there and clear and there's plenty of it, then it's quite easy to make a decision. If it's limited, then you start to just get people's opinions again." B11*

However, interviewees also described feeling powerless to biases even in the face of evidence if views were strong.

> *"…some people may have strong views even after they see the evidence, but when the evidence is there, you can show the evidence, you can emphasise on the evidence and the importance of the evidence, so they cannot say anything again, although some people can" B22*

Methodology, and the interpretation of evidence in the context of a lack of good evidence. A subtheme of a lack of good evidence was consistently mentioned by interviewees. They spoke of preferring to base recommendations on evidence, but that there was often a lack of good evidence for the healthcare questions that they wanted to answer.

> *"I think there are people around who think you shouldn't write a guideline unless there's definite evidence. Well that isn't the real world because the whole point is to guide people. In many ways it's the other way around. You need guidance when there isn't a controlled trial because that's why it's harder to pin down…I think we've done the best that we can do because most of the questions that we have are very limited in their evidence" A3*

Methodology and the restrictiveness of the NICE process. Participants also commented on the NICE process more generally in terms of fitting the evidence into a structured system.

> *"I understand that effectively if you've got a NICE guideline, the level of evidence expected is high quality randomised control trial and I think*

that's absolutely fine. It's then how you deal with that second tier of evidence." A3

"... there has been some discussions in some of the meetings about, "Well actually what we really wanted to ask was this" and then they said, "Well actually these were the questions that have come out so I can't go back and start doing the whole data review again."... we thought we knew what we wanted to ask, but as the evidence comes out there's maybe things that actually we might want to ask in a slightly different way, which might give us more of the answer we're looking for. But trying to do all that it's almost like starting again. So it's just not really feasible. N7

"...like a straightjacket to some extent" A15

Group 3: Group processes

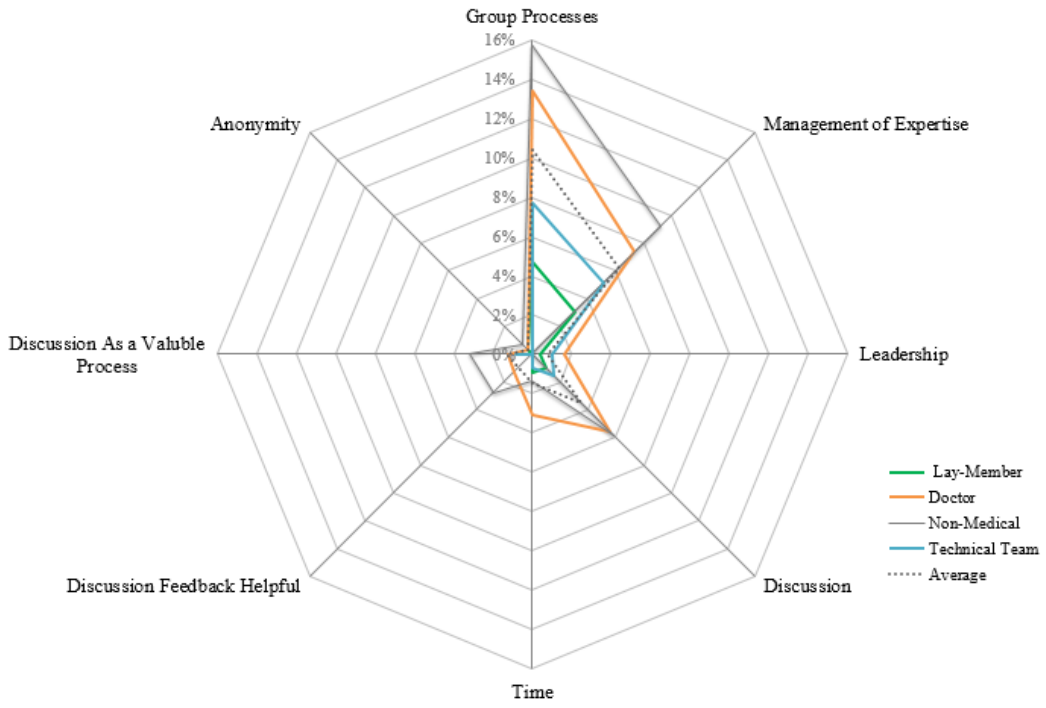


Figure 5: Chart showing percentage endorsement within theme of Group Processes of each subtheme across profession

Group processes was the third theme. Non-medical committee members commented on group processes the most, and lay-members the least. Medical doctors and non-medical individuals mentioned the management of expertise the most, and

lay-members the least. Non-medical members spoke about the process of discussion being helpful for enabling feedback, whereas medical professionals referenced discussion as an unhelpful process that was time consuming. Medical professionals also expressed more comments relating to leadership.

Group Processes and the Management of expertise through equal participation. Individuals discussed group processes in their interviews. A large number of participant comments concerned management group expertise, and how important this was for effective group decision-making. There were also comments on the focus on the impact of equal participation on managing dominant individuals.

> *"Because in a scenario where there's a less structured approach, some people will feel more confident about speaking out and putting forward their point of view than others. So it doesn't take into account different people's styles of contributing". A10*

> *"But both of them are such dominating personalities in the room that it's very hard for anyone else to oppose them. So difficult is the answer to making decisions in this committee, it's difficult". B20*

Some people referred to experiences of individual social power in shaping discussion. One participant specifically spoke about the impact of gender on power in committee participation.

> *"There's a lot of concern about doing a formal consensus, where it's coming from and I think also even if the formal consensus comes out with, "That is what the answer is" I still think those male dominants will try and twist it to how they think it should be". B20*

Some individuals spoke about managing equal participation at the level of committee selection.

"...essentially you're getting the decisions from those who are more vocal and those that are more confident in making decisions than those that aren't. Perhaps that's why it's important in selecting committee members to select those individuals that are very confident and are very happy to voice their opinions as opposed to people who are perhaps less dominant individuals. Perhaps their opinions could be lost in that setting because they're less likely to voice and contradict others.

They're more likely to just go with the flow of those dominant individuals". A18

Individuals also commented on worries concerning what the effect of giving people equal input might be.

"...there's lots of different expertise around the table then, in a sense, the weight needs to be shifted according to where the expertise is." A3

Lay-members discussed managing expertise the least, and made reference to feeling supported to voice their own opinions, which included having specific designated lay-member sessions and an appointed regular time slot with the chair.

"So, what I found would help in the first meeting the chair actually set aside a specific agenda item which was for laypeople to contribute our voices if we felt that we were being perhaps unable to speak during parts of the meeting where the clinical experts were sharing their views. But actually as the committee has progressed that's turned out to be unnecessary. From my point of view – I don't want to speak for the other layperson of course – but I've found that the committee has actually been quite welcoming and quite interested in hearing women's voices. So, I've felt there's been a quite open space for me to share my views and my ideas on what we should be doing. So, it's been quite a positive experience for me so far in the committee." B16

Non-medical professionals' comments were coded as concerning the management of expertise the most frequently. This was referred to in relation to practical constraints of time for discussion, the management of dominant individuals, and lacking confidence in one's own expertise.

"...we're all there because we've got an interest, a particular interest in this field, and it's a passion with most of the people there, so people do get passionate when they're speaking and sometimes it can be difficult then to have an equal contribution for everybody, because somebody speaks, who's very passionate about what they're saying, and you think, 'I'm not sure I can say anything, I can put another point of view because of that.'" A10

Group processes and anonymity. Anonymity was a theme that was identified as part of the group process feedback. Individuals spoke of balancing anonymity to encourage participation with the benefits of face-to-face interaction. One benefit of removing anonymity was suggested to be having the opportunity to meet with colleagues from different disciplines.

"...was it anonymous, these comments? ... Yes, so definitely that could help" A13

"Well, for a start, you know most of the people around the table anyway. Some members of the committee won't do. If you're representing gastroenterology, for example, that's the only time you'll tend to meet neonatologists or clinicians in neonatal care." A3

Group Processes, Discussion, and Discussion as a valuable process. The element of discussion in committee meetings was also a theme in the interviews. People described the process of discussion helpful in formulating thinking.

"...usually they discuss more informally between themselves, and they sometimes say 'Oh yeah, I've also observed this issue.' And then, through discussion, they manage to define it better. So these informal discussions can be quite helpful, to help the committee members maybe even understand themselves better what they're worried about or what they would like to improve" B21.

There was some conversation about how freeform discussion might be coupled with something more structured to increase its benefit:

"...I think discussion is always good because it's dynamic and because you have that instant response to somebody else's thoughts, and so obviously I think a bit of discussion is inevitable and really useful, but maybe based on something more structured beforehand". A10

Group Processes, Discussion, and Feedback as Helpful. Some participants noted the importance of discussion as a method of feedback that not only supported further thinking but also functioned as a corrective process during committee meetings. Some participants spoke of the benefit of feedback when comparing their own individual responses to that of the group at large, and also that this could enable individuals to see how they have contributed to a guideline.

"...I think discussion is always good because it's dynamic and because you have that instant response to somebody else's thoughts" A10

"...I guess people will be able to see what their own contribution is and how it fits in with everybody else". B5

"...let's say there's a score of six of zero to ten and I've put something as six and the vast majority put it as seven or eight and you might think, "Well actually have I read that question or fully understood it?" So there's a bit of thinking behind it about the feedback". A7

There was mention of face-to-face contact being preferable.

"..it's not the same as sitting in a committee where you've seen people and the way they approach their decision-making and where their areas of interest and expertise are". A3

A subset of individuals talked about finding committee discussions interesting.

"....I happen to be one of the people that knows the evidence on most things reasonably, well, but it's always quite interesting to hear people chip in from a different perspective and I think that's a healthy thing".A3

Group Processes, Discussion, and Time. There were many quotes regarding time-management in terms of discussion length.

> *"…sometimes we do just talk around items for some significant amount of time". A13.*

There was also reference to discussions tending to be lengthier when there was not a strong evidence-base, since it was more important to draw upon expertise.

> *"If you don't have anything to work with there's a lot of discussion, a lot of to-and-froing. It's harder to draw them back and say, okay, what is the recommendation here?" A18*

Participants also commented on the dangers of prioritising consensus rather than allowing for discussion.

> *"…the process of going through doesn't go under the surface and explore people's reasons for holding such a view…" B5*

> *"I did find it frustrating because it was, well, okay this is what the consensus has been, and then the door was shut on further discussion if you felt strongly that it didn't quite match what the data showed." A10*

Leadership. The role of leadership was spoke about and collected as a theme. The meeting chair and proper leadership was seen as a useful way of managing group processes.

> *"…the chair… does not allow people to dominate, but some other times chairs are more gentle, so they will not stop people". B22*

Continuity of Group Members

A final theme concerned disruptions in the continuity of personnel, both in terms of committee and technical team members between meetings.

> *"…they haven't been exposed to the same technical team and they've had that break in the middle. If that happened, it could upset their confidence…." A18*

> *"…it's difficult to get the entire panel together for every meeting, so I do find that sometimes if a particular member has not been at the previous meeting and we may revisit the same topic at the subsequent meeting and they are there we end up going back round almost the same discussion". A13*

Quantitative Data

All participants present completed the questionnaire. The questionnaire consisted of a Likert scale with scores from 1 to 7, where higher values indicated more positive experiences. Data were collected from a small sample of only 14 participants, therefore no statistical analysis was applied and results should be interpreted with caution. The data contained eleven outlier responses and failed tests of normality, hence median rather than mean values were used. The data are displayed alongside confidence intervals of the median in Table 4.

The most consistent findings were that participants felt more able to participate when using NGT in comparison to their experiences of other committee meetings that used informal consensus methods. Other questions did not report systematic differences between the NGT and informal consensus committee meetings.

Median ratings indicated that participants generally considered NGT and

informal consensus methods equivalent for nine of the twelve questions. Participants reported no difference in terms of their rating of their own contribution to the group. They also rated NGT as similar to informal consensus methods for group discussion (focus on the clinical question and how effectively disagreements were managed). However, there did appear to be a trend in favour of NGT for the group discussion being more productive. Participants rated the recommendations produced by NGT as similar to informal consensus methods in terms of clinical utility and personal satisfaction. When asked about the formal consensus method, people tended to feedback that the NGT method felt more satisfying. However, there were no differences disclosed between NGT and informal consensus methods for the statements effectively capturing the important aspects of the review question, improving their own understanding of the evidence, and how well presented the statements were.

Table 4: Median and Confidence Intervals for Questionnaire Data. Values represent self-report scores from a 1-7 Likert scale, where larger values indicate more positive experiences.

Question Group	Question	Median (95% Confidence Interval for the Median)
Your experience in the group discussion	How much did you feel able to participate?	5 (4-6)
	To what extent did you feel your time was well spent?	4.5 (3-5)
	How would you rate your overall contribution?	4 (4-5)
The recommendations decided by the group	How clinically useful do you feel they are?	4 (3-6)
	How satisfied were you with them?	4 (4-6)
Group discussion	How unproductive would you say the discussion was? (reverse scored)	5 (4-6)
	How focused was the discussion on the clinical question?	4 (3-6)
	How effectively were disagreements managed?	4 (4-5)
Formal consensus method	How satisfied were you with the consensus method?	5 (3-5)
	How effectively did the statements capture the important aspects of the review question?	4 (2-5)
	Did the use of statements worsen your understanding of the evidence? (reverse scored)	4 (3-6)
	How well presented was the evidence?	4 (3-6)

Discussion

The present study aimed to explore the views of committee and technical members regarding the expectations of NGT, including whether they thought it could enhance informal consensus methods. The data captured included experiences of informal consensus methods and past experiences of formal consensus methods other than NGT. These were all incorporated into the analysis because it was considered important to place the expectations of NGT in the context of the views of current informal consensus method practices and previous experiences of formal consensus methods more generally.

There was the sense overall that the use and application of NGT required careful thought, and that ensuring the credibility of the method needed to be balanced with the amount of resources required to implement it. Individuals were largely of the view that NGT should be used when there is limited evidence, although they reflected that this was often the case in healthcare guidelines. They shared concerns about an increased susceptibility to interpretation of evidence through biases in the face of limited or low-quality evidence. However, participants emphasised that providing guidance in the context of a lack of research evidence was important to ensure guidelines could be fully implemented and that a transparent mode of decision-making such as in NGT was a good fit with this.

The group processes that people identified as most helpful in committee meetings were effective leadership, allowing for anonymous input, managing expertise through encouraging equal participation, and time for face-to-face discussion. Discussion was identified as valuable to elicit feedback and develop ideas. However, there was a sense that discussion needed to be limited to enable

better time management. The cohesion of the group was also considered important, and interviewees spoke about the impact that disruptions through a lack of continuity of the committee members and technical team would have on the group's effectiveness.

It is of interest that there is considerable overlap between the themes found in the present empirical paper and the conceptual introduction that reviewed literature concerning a broader use of formal consensus methods. This could support a sense of themes as common and robust across formal consensus methods and approaches, and lend support to the development of hybrid methods to best enhance the respective qualities of various formal consensus methods.

Formal Consensus

Within the cluster of themes of formal consensus methods, individuals spoke about balancing the credibility of formal consensus methods with the effort required to implement them. Participants expressed doubt whether useful outcomes could be gained when relinquishing control to a systematic method and being restricted by the structure of the formal process. These feelings are in contrast to evidence that shows formal consensus methods increase the number of ideas and feedback due to the individual rating system (Jarboe, 1988; Van de & Delbecq, 1974; White, Dittrich, & Lang, 1980). This is thought to be achieved because of reduced "production blocking" that tends to occur and decrease the number of ideas when people are required to manage social turn taking in order to contribute ideas, rather than being able to offer them concurrently (Diehl & Stroebe, 1987; Powell, 2003).

Despite these challenges, interviewees referred to the advantages of the

robustness and transparency of the method, not only in the process but also in the reporting of guideline development. Even though this has been an important area identified by the participants in this study, the clear and consistent reporting of formal consensus methods has been identified as a problem by researchers. Authors have commented on the low quality of reporting of the formal consensus method process, and consequent high variation in their application (Hasson & Keeney, 2011; Humphrey-Murto, Varpio, Gonsalves, & Wood, 2016; Humphrey-Murto et al., 2017). Both the participants in this study and researchers appear to agree that this is an area that requires improvement as a priority.

Interviewees spoke about the resource intensiveness of NGT, not only in terms of the time to conduct it, but also staff resources in terms of the learning of a new process by committee members and the technical team. The view appeared to be that this process was interesting and novel, but also effortful. Previous research has discussed the additional resources required by formal consensus methods, and suggests that participants receive proper additional resources to support their implementation (e.g. see Yang, 2013 for a discussion of comparitive cost implications of Delphi). The actual resource difference between informal and formal consensus methods requires further research. However, the additional cost felt by the participants of a potentially novel method remains valid, and could mean additional support might be required.

In terms of differences by profession, the technical team paid most attention to the resource intensiveness of using NGT. This is understandable considering that, in this form of NGT, the technical team were responsible for generating the consensus statements, as has been suggested by past authors (Kea & Sun, 2015).

Interestingly, the technical team also made more references than other professional groups to NGT being appropriately applied, specifically when evidence is limited. This could be related to their appreciation of the resource intensiveness, and therefore increased awareness of the need to restrict the use of NGT.

The lay-members and the technical team spoke most about robustness and transparency of the method. They also referred to feeling suspicious towards the formal consensus method and about the development of guidelines in general. These themes might be linked conceptually as well as apparently co-occurring in the data. Participants often described appreciating NGT for being robust and transparent, which could be a way to counteract suspicions towards guideline development. This link between suspicion and methodological quality has been suggested by authors elsewhere (R. Graham, Mancher, Wolman, Greenfield, & Steinberg, 2011).

Lay-members discussed robustness and transparency most often, particularly in terms of how the guideline might be perceived by readers. This could be because lay-members are more aware of the variation in types of evidence that are being offered, including their own experiential evidence. Authors have suggested that lay-member involvement requires careful consideration to be meaningful (van de Bovenkamp & Trappenburg, 2009), since lay-members can at times feel their experiential knowledge is not considered as seriously as scientific evidence (Jarrett, 2004). Therefore, lay-members may have highlighted the importance of robustness and transparency to a greater extent than other professions. Further investigation is required to explore this possibility. The predominance of these themes supports NGT as a helpful way to support guideline development.

Methodology

The methodology group of themes focused not only on NGT, but also the perceptions of methodology more generally at NICE. Participants talked about prioritising evidence over experience to buffer against biases of interpretation. However, they also spoke about the tendency to prioritise certain types of evidence, such as randomised control trials (RCTs), and that this felt unsatisfactory particularly when there was evidence from other research designs or when there was secondary outcome data. This is in keeping with the culture of "evidence-based practice", which emphasises RCTs, meta-analyses, and systematic reviews as a "gold standard" to reduce the focus on financial and therapeutic aspects of an intervention (National Institute for Health and Clinical Excellence, 2007; Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996).

Guidelines are required to combine person-centred knowledge with evidence-based care and are often required most in areas where there is uncertainty and the evidence is less clear (Jo Rycroft-Malone et al., 2003). As a committee of professionals who accompany the technical team, the offering of experiential expertise is an essential contribution of committee members. Formal consensus methods offer a helpful way of collating information from a broad range of evidence types. The feelings expressed of methodology restriction during the NICE process could be a reflection of attempting to maintain a culture of hierarchical evidence in a guideline process that requires the use of alternative evidence. This could suggest an overemphasis on research evidence as equalling truth, when a more accurate perception is that evidence is open to a range of social constructions and interpretations by its users (Higgs & Titchen, 1995; Wood, Ferlie, & Fitzgerald,

1998). It has been argued that research evidence is best suited to generalizable factors, whereas expertise-driven evidence is appropriate for the person-centred practice which is perhaps more relevant to applied healthcare guidelines (Eraut, 2000; Higgs, Jensen, Loftus, & Christensen, 2008; Higgs & Titchen, 1995).

Doctors spoke most often about the methodology and the technical team the least. All healthcare professionals spent more time discussing feeling restricted by the NICE process, and about guideline interpretation and implementation. This could be because healthcare professionals are more acutely aware of the differences between clinical practice and the guideline and research questions (Grol et al., 1998).

Lay-members spent more time considering the management of personal biases in guideline development groups. In particular, they referred to prioritising evidence rather than experiential understanding, which included their own contributions. As previously discussed, this could be supported by research into the role of lay-member and experiential knowledge in guideline groups (van de Bovenkamp & Trappenburg, 2009), which could result in lay-members feeling that their views are not valued because they are not grounded in research evidence.

Drawing together data from the theme clusters of formal consensus and methodology, it could be that different professions might adopt formal consensus methods for varied reasons. Lay-members and technical team members might prefer them because they are robust and transparent, whereas medical doctor and other health professional members might use them because they allow for better guideline implementation.

Group Processes

The theme cluster of group processes was broken down into themes of management of expertise, leadership, anonymity, and discussion.

The management of expertise was one of the largest and was the most referenced theme in the interviews. This was often in the context of a sense of increasing equal participation, which was consistent with the quantitative results. Individuals responded that they felt more able to participate in the NGT meeting than in other meetings. Interviewees expressed interest and appreciation about the range of feedback from their colleagues. They paid particular attention to the variety of professional backgrounds present in meetings, which created challenges for understanding where the individual expertise lay since it represented a dynamic construct specific to the particular area being discussed. This expertise was not always systematically identified because of dominance differences between committee members. It was thought by participants that NGT would be a helpful response to this, which has been mirrored in the literature (see Murphy et al., 1998 for review). However, some participants were cautious, stating that the equality of formal consensus methods might be a disadvantage because it did not distinguish between a genuine lack of expertise but instead gave everyone equal weighting.

There have been versions of formal consensus methods that assign "weighting" to particular response to enhance their influence on the final decision. However, this is generally not advised since it is unclear how expertise should be defined for the assignment of weights (e.g Black et al., 1999), and there is lack of research investigating how weighing impacts outcomes**.**

When focusing on leadership, individuals tended to talk about leaders in terms of their ability to efficiently manage the discussion. This included keeping the group on topic, encouraging participation, and managing discussion length. Indeed, leadership style has been identified as important for effective decision-making either through improving decision quality (George, Dennis, & Nunamaker, 1992; Somech, 2006) or group cohesion and processes (Anson, Bostrom, & Wynne, 1995). This is believed to be because an effective facilitator increases the sharing of shared and unshared information by summarising and repeating information to the group that is relevant to the task (Larson et al., 1998; Larson, Christensen, Abbott, & Franz, 1996; Larson et al., 1994; Vinokur, Burnstein, Sechrest, & Wortman, 1985; Wortman, Vinokur, & Sechrest, 1988). Leaders are also likely to play a central role in encouraging a non-judgemental and supportive atmosphere to foster information exchange (Faulmüller, Mojzisch, Kerschreiter, & Schulz-Hardt, 2012).

Anonymity has been shown to improve the quantity of ideas generated (Postmes & Lea, 2000), probably because it reduces the need for censorship. Anonymity has been shown to reduce the influence of social power in small group decision making (Parks & Sanna, 1999; Postmes, Spears, Sakhel, & De Groot, 2001), which concurs with the expressed desires of the participants in the present study for anonymity to increase freedom of expression. However, they also spoke about frustrations relating to anonymity in the past, since it limited the ability to provide specific feedback and a sense of collaboration with particular individuals. Previous research has linked anonymity to increased conflict, social loafing, unproductive discussion, and inability to reach consensus (Jessup & George, 1997; Valacich, Jessup, Dennis, & Nunamaker, 1992). Furthermore, anonymity might encourage a reduced sense of accountability for individuals involved, which has been associated

with reduced decision quality (Scholten, van Knippenberg, Nijstad, & De Dreu, 2007).

To manage the costs and benefits of anonymity, authors have suggested the inclusion of both silent voting and collective discussion for optimising group consensus and the integration of ideas (Souder, 1977). It could be that initial face to face meetings to support group cohesion could be followed by anonymous rounds of voting to manage the tendency for social biases of information exchange (Driskell, Radtke, & Salas, 2003; Levi, 2001; Roch & Ayman, 2005)

The theme of discussion as an important process to develop thinking and increase the exchange of ideas agrees with evidence that increased discussion time increases opportunities for the exchange of unshared information and therefore improves decision-making (Bowman & Wittenbaum, 2012). Increasing discussion time could also mitigate concerns that the consensus process might reduce dissent, with individual opinions becoming lost (Powell, 2003). Extending the length of group discussion has been associated with an improved depth of information processing by groups (Kelly & Karau, 1999; Parks & Cowlin, 1995). Group discussions have been illustrated as progressing through periods of information exchange, opinion exchange, and proposal exchange (Bales & Strodtbeck, 1951; Hirokawa, 1990; Wiltshire, Butner, & Fiore, 2018). Solutions that score most favourably are more likely to be discussed later in increasing detail each time, a process termed “spiralling” rounds of “reach-testing”(Poole & Roth, 1989; Scheidel & Crowell, 1964). This sequence was preferred in one study by the majority of individuals over more linear methods of formalised decision-making (Pavitt, 2009).

Research around discussion length in group decision making has tended to

focus on ad-hoc teams created in laboratory settings. This experimental design is subjected to a variety of limitations since it limits testing of established groups with high cohesiveness and familiarity, which is likely to impact discussion and consensus. Groups that are familiar with each other are likely to focus more on interpersonal rather than task-related group discussion (Okhuysen, 2001), more likely to have established group norms that may interfere with unbiased information processing (Gigone & Hastie, 1997), and are more likely to exchange unshared information (Gruenfeld, Mannix, Williams, & Neale, 1996). More research in more naturalistic settings is needed to better understand how these factors might impact the utility of free discussion in groups.

Evidence has shown that despite discussion being traditionally viewed as unhelpful to efficient performance of groups (Mercier, Trouche, Yama, Heintz, & Girotto, 2014), it has been found to be helpful for encouraging the sharing of expertise, increasing effectiveness, and evaluating information credibility (Bahrami et al., 2012; Klein & Epley, 2015; Minson, Liberman, & Ross, 2011).

Non-medical professionals discussed group processes the most, and lay-members the least. Lay-members also discussed managing expertise the least. At first glance, this distribution might seem surprising as lay-members are typically considered to hold the least power in guideline groups. However lay-members spoke about feeling supported to voice their own opinions, including have specific designated lay-member sessions and an appointed regular time slot with the chair.

Another power differential that is perhaps less conspicuous is between the doctors and non-medical group members. In fact, it was these two professional groups that talked about group processes the most. Non-medical team members

spoke most about discussion being helpful to develop ideas and also to get feedback from colleagues. Conversely, medical professionals expressed frustrations at how lengthy discussions were in meetings, and identified the helpfulness of leaders to impose limits on discussion the most. This pattern of responding could indicate that non-medical professions appreciate the use of discussion as a way of managing dominant individuals more than medical doctor professions. This could be because increasing discussion affords for more time allocation to information processing and exchange and discourages premature consensus (Brodbeck, Kerschreiter, Mojzisch, & Schulz-Hardt, 2007).

Some authors have suggested that the evidence-base can be a challenge for established medical professionals who feel their expertise is sufficiently complete to answer clinical questions (Lipman, 2000). In agreement with this hypothesis, individuals from medical and non-medical backgrounds mentioned the management of expertise the most, and lay-members the least. Status has been noted to impact decision-making. Participation has been linked to member status and relevant expertise (Vinokur et al., 1985). Furthermore, the majority opinion is more like to be adopted when status is unequal in a group for tasks involving judgement with no "correct" answer, such as in the case of guideline development (Kirchler & Davis, 1986).

The management of expertise could be linked to a wider theme of social power. One participant made reference to "male dominants", indicating a narrative of gendered power in committee groups. Medicine as a largely male dominated profession has been discussed in the wider literature (e.g. Reichenbach & Brown, 2004). Further research in this area could help understand how wider social

narratives impact decision-making within committees and whether formal consensus methods could help manage these processes.

Continuity of Group Members

Interviewees spoke of the disruptions that changes in staffing or inconsistent attendance by committee members had on decision-making. Group cohesiveness can be defined as the tendency for a group to collectively work together to achieve mutual goals or meet shared emotional needs (Carron & Brawley, 2000). Group cohesion has been shown overall to be related to group performance (Hogg, 1992; Klein & Mulvey, 1995). It is unclear whether commenting on the inconsistency of group members could be a cause of reduced group cohesion, or whether it is as a result of difficulties with group cohesion.

A potential reason why cohesiveness might be particularly important for guideline committee can be gained from a model developed by Forbes & Milliken (1999) which focuses on group effectiveness and efficiency. In this model, cognitive conflict and group cohesiveness have an inverse relationship. Cognitive conflict has been associated with negative feelings (Nemeth & Staw, 1989), and groups harbouring higher levels of conflict have reduced member satisfaction and expressed desire to remain in the group (Jehn, 1995; Schweiger, Sandberg, & Ragan, 1986). Group heterogeneity of expertise, as emphasised in healthcare guidelines, could increase dissent and sharing of new knowledge. The model theorises that this serves to reduce the cohesiveness of the group, and some authors have suggested the need for mitigation of this through effective facilitation to promote a sense of group collaboration (Nicholson, Pugliese, & Bezemer, 2017; Nordberg & Booth, 2017). Thus, it could be that healthcare guideline groups would be particularly sensitive to

changes in group membership due to increased heterogeneity.

The Quantitative Data

Statistically reliable differences were found for increased participation only when committee members were asked to compare their experiences of the NGT method with informal methods used by the committee in other meetings. This supports increased participation as a key feature of informal consensus methods (Murphy et al., 1998).

No significant differences were found for the other questionnaire items. However, there was a trend in favour of NGT for feeling the meeting was more productive and feeling satisfied with the formal consensus method. There could be several reasons for this. Firstly, there was a trend in favour of the NGT, and it could be that the smaller sample size inflated the possibility of type II error. If this is the case, a larger sample might have yielded more reliable results. Secondly, it could be that the benefits of the NGT were not captured by the questionnaire items. This is possible since a validated questionnaire was not available, and so the questionnaire was develop without extensive testing and feedback, but rather only following a small pilot. Further research could use larger groups of user consultation and feedback to develop questions that meaningfully capture the difference in experiences of guideline groups using NGT. From the thematic analysis, it appears there are differences across professions in terms of what components of the NGT they find helpful and satisfying. It could be these differences that account for the wide spread in the data. For example, in response to increased free-flowing discussion, medical professionals might increase their ratings for decreased productivity. However, non-medical professionals might report the opposite, that of

increased productivity. These potential differences support the use of mixed methods, as was adopted by the present study, to better understand reasons for variability within the data.

Limitations

Although this research aimed to understand individual experiences of informal consensus and expectations of formal consensus using NGT, it did not provide a direct measure of these variables. Rather, it relied upon members' experiences of current informal consensus methods and recollections of past experiences of formal consensus methods. This could be subject to a variety of biases and heuristics, including social desirability (Levy, 1981), self-serving (Myers & Twenge, 2015), self-referencing (Rogers, Kuiper, & Kirker, 1977), and confirmation (Nickerson, 1998) biases. Nevertheless, the results are placed within the context of a wider literature, and encourage further research into the area.

One particular source of bias might have been in recruitment to the study. Participants were not offered compensation for their time, which could have resulted in the recruitment of motivated participants. Motivation could have biased participant responses due to being more likely to represent particularly strong negative or positive views. There were several attempts to mitigate this potential bias. Firstly, efforts were made to increase the convenience of interviews to encourage participation from as broad as sample as possible, including flexible times, location, and the potential for phone interviews. Secondly, the researcher attended multiple committee meetings so that they were familiar to the committee to engender trust and openness. Thirdly, the quantitative questionnaire was gathered from all committee members and attempted to corroborate the qualitative data. Despite these efforts,

there were still interviews that did not happen due to lack of engagement, and members who were not interviewed were not thoroughly followed up due to limitations on resources.

The background of participants might have impacted the results. The cultural background of the participants was overwhelmingly White European. It is unclear how the themes generated would have been different had the sample been from a broader range of cultural backgrounds. A difference between participants which was not included in the analysis was their previous experiences of formal consensus methods.

A significant limitation of the study was that participant prior experiences of NGT or other formal consensus methods were not accounted for in the analysis and presentation of the results. Participants reported a broad range of prior experiences, and these differences spanned across all professional backgrounds. Therefore, systematic differences in experience are less likely to have biased the data for any one particular profession. Previous NGT experience has not been systematically linked to particular attitudes towards it (Gresham, 1986). However, further research would be important to understand how the views and experiences of participants progress with exposure to formal consensus methods. This could be done using an alternative qualitative methodology to investigate changes over time in particular participants that use formal consensus methods. A narrative approach could be used to achieve this (Braun & Clarke, 2006). Investigating the impact of prior formal consensus method experience on participant expectations could help with the selection and training of committee members. The NGA could focus on recruiting members that are likely to make best use of formal consensus methods. It could be,

for example, that having prior experience of any type of formal consensus method leads to greater acceptance of the method and lower felt effort (e.g. differences in the theme prevalence of "credibility" and "effort and resource intensiveness"). In addition, it might be particularly important to understand whether the impact on participant expectations is specific to the type of prior experience. For example, does the previous experience need to be specific to the type of formal consensus being used (e.g. past experience of NGT only impacts expectations of using NGT) or is the impact is more general (e.g. past experience of any type of formal consensus method impacts expectations of using NGT).

This research is predominantly qualitative. Therefore, the data is not intended to be generalisable, and is limited to the particular context in which the data was gathered (Barker, Pistrang, Elliott, & Barker, 2002). For example, interviewees often made reference to the variation in group processes across guideline meetings, which might have been specific to this guideline group. Since all data was gathered from predominantly one guideline group, the themes may be specific to the particular committee. The aim of this study was to better understand NGT in the context of NICE guideline development, and thus it was appropriate to implement the research within the specific context of NICE.

Conclusions

The overall view from all professions is that NGT is helpful provided it is offered in the correct context with additional allocation of resources through proper training and support and preparation time. The benefits of NGT were potentially broad and viewed differently depending on professional background, with some individuals highlighting the robustness and transparency of the method, the ability to

use evidence from a variety of sources, or the complimentary balancing of anonymity and discussion. The area that was most endorsed and supported by the quantitative data was the management of expertise in terms of equal participation by NGT. This was identified as an important benefit of formal consensus methods in the literature, which highlights the method as encouraging the exchange of unshared expertise. Individuals had some concerns regarding whether certain expertise should be weighted as part of the process, and this requires more research.

Implications for Research and Practice

The present study highlights the importance of a mixed methods approach to better understand the complexities affect group dynamics in consensus development for healthcare guideline groups. There are many avenues for further extension of this research. These include further research into how social power might impact committees when making decisions. Further research could also help illustrate whether the views from the present research are present in other guideline committees, and whether perceptions are impacted by other variables, for example the cohesiveness of the group or presence of past experiences with formal consensus methods. Another area for further investigation is how the weighting of expertise might be of benefit to guideline groups. Finally, it would be helpful to evaluate whether the use of NGT improves not only guideline development but leads to better recommendations as assessed by organisations external to NICE .

Recommendations

Regarding NGT formal consensus

- Because individuals have a tendency to be cautious in the face of a new

method, extra time and training may be required upon initiation of formal consensus methods to provide people with explicit examples of its credibility.

- Consideration may be necessary in light of the additional resources required by formal consensus methods, the majority of which is experienced by the technical team. The explicit allocation of additional resources might help individuals feel more supported, for example via the production of a manual or supplementary support or supervision.
- The benefits of formal consensus methods are understood not only to be towards the process and committee but also to the wider perception of the credibility of a guideline to mitigate against potential feelings of suspicion regarding the credibility of guideline development particularly when there is limited good quality evidence.
- It could be important to manualise and clearly document formal consensus methods to preserve their transparency as an advantage. Furthermore, a manual is likely to benefit if it states the situations when formal consensus methods are to be used to make the best use of resources.

Regarding Methodology

- Individuals from different professional backgrounds may place emphasis on varied aspects of the benefits of formal consensus methods. Therefore, presenting the full range of potential benefits might be the best approach to promote engagement of a wide professional audience of committee members.

Regarding Group Processes

- More research is required to better understand how to manage the distribution of unshared knowledge.

- Individual chairs of the committees would benefit from training in formal consensus methods because of their potential to increase the effectiveness of the process. In particular, it could be helpful to encourage chairs to support committee members to exchange unshared information and experiences.
- There should be both anonymous and face-to-face elements to NGT since these hold different benefits. Anonymity reduces feelings of social pressure, whereas personal meetings preserve as sense of personal accountability and allow for inter-professional live discussions.
- Discussion should remain as an important element of the formal consensus process, which could support non-medical professionals in contributing to the process. However, discussion should also be limited and paired with individual rating and more structured interventions.
- Support could be focused more on allowing for non-medical professionals to share their expertise, which may be enhanced through formal consensus methods

Regarding the Continuity of Group Members

- Member consistency was considered as important from both a technical team and committee member standpoint in terms of conserving information across meetings. It is unclear whether formal consensus, due to its increased speed (Nemiroff, Pasmore, & Ford, 1976), could reduce the impact of the disruptions of changes in group members since tasks can be contained to within a single meeting. More research is necessary to better understand this.
- Increased group dissent due to heterogeneity of expertise could be increased by NGT because of the increased exchange of unshared expertise. This might result in reduced group cohesion and therefore could impact the tendency for

member consistency across group meetings. Facilitators could have a role in supporting groups to increase cohesion.

- Further research could investigate the views of committee members on consistent attendance, and the reasons for member drop-out.

References

Anson, R., Bostrom, R., & Wynne, B. (1995). An Experiment Assessing Group Support System and Facilitator Effects on Meeting Outcomes. *Management Science, 41,* 189–208.

Asch, S. E. (1952). Group forces in the modification and distortion of judgments. In *Social psychology* (pp. 450–501). Englewood Cliffs: Prentice-Hall, Inc.

Bahrami, B., Olsen, K., Bang, D., Roepstorff, A., Rees, G., & Frith, C. (2012). What failure in collective decision-making tells us about metacognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367,* 1350–1365.

Bales, R. F., & Strodtbeck, F. L. (1951). Phases in group problem-solving. *The Journal of Abnormal and Social Psychology, 46,* 485–495.

Ball, T., & Jones, S. (1977). The utilization of nominal groups as an aid in marketing organizations. *Proceedings: Southeast American Institute for Decision Science,* 63–65.

Bang, D., & Frith, C. D. (2017). Making better decisions in groups. *Royal Society Open Science, 4,* 1–22.

Barker, C., Pistrang, N., Elliott, R. (2002). *Research Methods in Clinical Psychology: An introduction for students and practitioners.* Chichester, UK: Wiley.

Beranek, M., Beise, C. M., & Niederman, F. A. (1993). Facilitation and group support systems. *Proceedings of the 26th Hawaii International Conference on*

Systems Sciences, Institute of Electrical and Electronics Engineers, 199–207.

Bini, S. A., & Mahajan, J. (2016). Achieving 90% Adoption of Clinical Practice Guidelines Using the Delphi Consensus Method in a Large Orthopedic Group. *The Journal of Arthroplasty, 31,* 2380–2384.

Black, N., Murphy, M., Lamping, D., McKee, M., Sanderson, C., Askham, J., & Marteau, T. (1999). Consensus Development Methods: A Review of Best Practice in Creating Clinical Guidelines. *Journal of Health Services Research & Policy, 4,* 236–248.

Bowman, J. M., & Wittenbaum, G. M. (2012). Time Pressure Affects Process and Performance in Hidden-Profile Groups. *Small Group Research, 43,* 295–314.

Braun, V., & Clarke, V. (2006). Using Thematic Analysis In Psychology. *Qualitative Research in Psychology, 3,* 77–101.

Brodbeck, F. C., Kerschreiter, R., Mojzisch, A., & Schulz-Hardt, S. (2007). Group decision making under conditions of distributed knowledge: The information asymmetries model. *Academy of Management Review, 32,* 459–479.

Brownstein, A. L. (2003). Biased predecision processing. *Psychological Bulletin, 129,* 545–568.

Camacho, L. M., & Paulus, P. B. (1995). The Role of Social Anxiousness in Group Brainstorming. *Journal of Personality and Social Psychology, 68,* 1071–1080.

Campbell, S. M., & Cantrill, J. A. (2001). Consensus methods in prescribing research. *Journal of Clinical Pharmacy and Therapeutics, 26,* 5–14.

Carron, A. V., & Brawley, L. R. (2000). Cohesion: Conceptual and Measurement Issues. *Small Group Research, 31,* 89–106.

Chen, E. T., Sicrar, S., Hwang, H.-G., & Hwang, H.-G. (1998). Use of GDSS and training of group members to improve decision-making: An empirical study. *Use of GDSS Journal of International Information Management,7,1,*2.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement, 20*, 37–46.

Coulter, I. D., Marcus, M., & Freed, J. R. (1998). Consistency across panels of ratings of appropriateness of dental care treatment procedures. *Community Dental Health, 15,* 97–104.

Dalkey, N., & Helmer, O. (1963). An Experimental Application of the Delphi Method to the Use of Experts. *Management Science, 9,* 458–467.

De Ruyter, K. (1996). Focus versus nominal group interviews: a comparative analysis. *Marketing Intelligence & Planning, 14,* 44–50.

Dennis, A. R. (1996). Information Exchange and Use in Group Decision Making: You Can Lead a Group to Information, but You Can't Make It Think. *MIS Quarterly, 20,* 433.

Diehl, M., & Stroebe, W. (1987). Productivity Loss In Brainstorming Groups: Toward the Solution of a Riddle. *Journal of Personality and Social Psychology, 53, 3,* 297-509.

Dopson, S., FitzGerald, L., Ferlie, E., Gabbay, J., & Locock, L. (2002). No magic targets! Changing clinical practice to become more evidence based. *Health*

Care Management Review, 27, 35–47.

Driskell, J. E., Radtke, P. H., & Salas, E. (2003, December). Virtual teams: Effects of technological mediation on team performance. *Group Dynamics, 7,* 297–323.

Durham, C. C., Locke, E. A., Poon, J. M. L., & McLeod, P. L. (2000). Effects of Group Goals and Time Pressure on Group Efficacy, Information-Seeking Strategy, and Performance. *Human Performance, 13,* 115–138.

Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71,* 5–24.

Elliott, R., Fischer, C. T., & Rennie, D. L. (1999). Evolving guidelines for publication of qualitative research studies in psychology and related fields. *The British Journal of Clinical Psychology, 38,,* 215–229.

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology, 7,* 113-136.

Eriksen, B.O., Almdahl, S.M., Hensrud, A., Jaeger, S., Kristiansen, I.S., Mürer, F.A., Nord, E., Pape, J.F., Robertsen, R., & Thorsen, G. (1996). Assessing health benefit from hospitalization. Agreement between expert panels. *International Journal of Technology Assessment in Health Care, 12,* 126–135.

National Institute for Health and Care Excellence. (2016). *Mental health problems in people with learning disabilities: prevention, assessment and management.* London: NICE.

Faulmüller, N., Mojzisch, A., Kerschreiter, R., & Schulz-Hardt, S. (2012). Do

You Want to Convince Me or to Be Understood?: Preference-Consistent Information Sharing and Its Motivational Determinants. *Personality and Social Psychology Bulletin, 38,* 1684–1696.

Fitch, K., Bernstein María, S. J., Aguilar, D., Burnand, B., Lacalle, J. R., Lázaro, P. van het Loo, M., McDonnell, J., Varder, J., & Kahan, J. P. (2001). The RAND/UCLA Appropriateness Method User's Manual. *Santa Monica, California: RAND Coorporation.*

Forbes, D. P., & Milliken, F. J. (1999). Cognition and Corporate Governance: Understanding Boards of Directors as Strategic Decision-Making Groups. *The Academy of Management Review, 24,* 489.

Galinsky, A. D., & Kray, L. J. (2004). From thinking about what might have been to sharing what we know: The effects of counterfactual mind-sets on information sharing in groups. *Journal of Experimental Social Psychology, 40,* 606–618.

George, J. F., Dennis, A. R., & Nunamaker, J. F. (1992). An experimental investigation of facilitation in an EMS decision room. *Group Decision and Negotiation, 1,* 57–70.

Gigone, D., & Hastie, R. (1997). The impact of information on small group choice. *Journal of Personality and Social Psychology, 72,* 132–140.

Graham, I. D., & Harrison, M. B. (2005). Evaluation and adaptation of clinical practice guidelines. *Evidence-Based Nursing, 8,* 68–72.

Graham, R., Mancher, M., Wolman, D. M., Greenfield, S., & Steinberg, E.

(2011). Clinical Practice Guidelines We Can Trust. In *Clinical Practice Guidelines We Can Trust.* Washington: National Academies Press.

Greitemeyer, T., & Schulz-Hardt, S. (2003). Preference-consistent evaluation of information in the hidden profile paradigm: Beyond group-level explanations for the dominance of shared information in group decisions. *Journal of Personality and Social Psychology, 84,* 322–339.

Gresham, D. J. (1986). *Expressed Satisfaction with the Nominal Group Technique Among Change Agents* (Unpublished doctoral dissertation). Graduate College of Texas A&M University, USA.

Grol, R., Dalhuijsen, J., Thomas, S., In ’t Veld, C., Rutten, G., & Mokkink, H. (1998). General practice Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. *British Medical Journal, 317,* 858–861.

Gruenfeld, D. H., Mannix, E. A., Williams, K. Y., & Neale, M. A. (1996). Group composition and decision making: How member familiarity and information distribution affect process and performance. *Organizational Behavior and Human Decision Processes, 67,* 1–15.

Guzys, D., Dickson-Swift, V., Kenny, A., & Threlkeld, G. (2015). Gadamerian philosophical hermeneutics as a useful methodological framework for the Delphi technique. *International Journal of Qualitative Studies on Health and Well-Being, 10,* 262-291.

Hackman, J. R., & Kaplan, R. E. (1974). Interventions into Group Process: An Approach to Improving the Effectiveness of Groups. *Decision Sciences, 5,* 459–

480.

Hall, J., & Watson, W. H. (1970). The Effects of a Normative Intervention on Group Decision-Making Performance. *Human Relations, 23,* 299–317.

Hall, J., & Williams, M. S. (1970). Group Dynamics Training and Improved Decision Making. *The Journal of Applied Behavioral Science, 6,* 39–68.

Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change, 78,* 1695–1704.

Henningsen, D. D., Cruz, M. G., & Miller, M. L. (2000). Role of social loafing in predeliberation decision making. *Group Dynamics: Theory, Research, and Practice, 4,* 168–175.

Higgs, J., Jensen, G., Loftus, S., & Christensen, N. (2008). *Clinical reasoning in the health professions.* Amsterdam: Butterworth-Heinemann.

Higgs, J., & Titchen, A. (1995). The Nature, Generation and Verification of Knowledge. *Physiotherapy, 81,* 521–530.

Hirokawa, R. Y. (1990). The Role of Communication in Group Decision-Making Efficacy. *Small Group Research, 21*, 190–204.

Hirokawa, R. Y., Ice, R., & Cook, J. (1988). Preference for procedural order, discussion structure and group decision performance. *Communication Quarterly, 36,* 217–226.

Hirokawa, R. Y., & Pace, R. (1983). A descriptive investigation of the possible communication-based reasons for effective and ineffective group decision

making. *Communication Monographs, 50,* 363–379.

Hoffman, L. R. (1979). Applying Experimental Research on Group Problem Solving to Organizations. *The Journal of Applied Behavioral Science, 15,* 375–391.

Hoffman, L. R., & Maier. (1959). Homogeneity of member personality and its effect on group problem-solving. *The Journal of Abnormal and Social Psychology, 58,* 27–32.

Hogg, M. A. (1992). *The social psychology of group cohesiveness : from attraction to social identity.* New York: New York Press.

Humphrey-Murto, S., Varpio, L., Gonsalves, C., & Wood, T. J. (2016). Medical Teacher Using consensus group methods such as Delphi and Nominal Group in medical education research Using consensus group methods such as Delphi and Nominal Group in medical education research. *Medical Teacher, 1,* 14–19.

Humphrey-Murto, S., Varpio, L., Wood, T. J., Gonsalves, C., Ufholz, L.-A., Mascioli, K., Wang, C., Foth, T. (2017). The Use of the Delphi and Other Consensus Group Methods in Medical Education Research. *Academic Medicine, 92,* 1491–1498.

Hutchings, A., Raine, R., Sanderson, C., & Black, N. (2006). A comparison of formal consensus methods used for developing clinical guidelines. *Journal of Health Services Research and Policy, 11,* 218–224.

Jarboe, S. (1988). A comparison of input-output, process-output, and input-process-output models of small group problem-solving effectiveness. *Communication Monographs, 55,* 121–142.

Jarrett, L. (2004*). A report on a study to evaluate patient/carer membership*

of the first NICE Guideline Development Groups; A report on a study to evaluate patient/carer membership of the first NICE Guideline Development Groups. London: NICE.

Jehn, K. A. (1995). A Multimethod Examination of the Benefits and Detriments of Intragroup Conflict. *Administrative Science Quarterly, 40,* 256.

Jessup, L. M., & George, J. F. (1997). Theoretical and methodological issues in group support systems research: Learning from groups gone awry. *Small Group Research, 28,* 394–413.

Jones, J., & Hunter, D. (1995). Consensus methods for medical and health services research. *British Medical Journal (Clinical Research Ed.), 311,* 376–380.

Kastein, M. R., Jacobs, M., van der Hell, R. H., Luttik, K., & Touw-Otten, F. W. M. M. (1993). Delphi, the issue of reliability: A qualitative Delphi study in primary health care in the Netherlands. *Technological Forecasting and Social Change, 44,* 315–323.

Kea, B., & Sun, B. C.-A. (2015). Consensus development for healthcare professionals. *Internal and Emergency Medicine, 3,* 373–383.

Kelly, J. R., & Karau, S. J. (1999). Group decision making: The effects of initial preferences and time pressure. *Personality and Social Psychology Bulletin, 25,* 1342–1354.

Kirchler, E., & Davis, J. H. (1986). The influence of member status differences and task type on group consensus and member position change. *Journal of Personality and Social Psychology, 51,* 83–91.

Klein, H. J., & Mulvey, P. W. (1995). Two Investigations of the Relationships among Group Goals, Goal Commitment, Cohesion, and Performance. *Organizational Behavior and Human Decision Processes, 61,* 44–53.

Klein, N., & Epley, N. (2015). Group discussion improves lie detection. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 7460–7465.

Kopp, I. B., Selbmann, H.-K., & Koller, M. (2007). Konsensusfindung in evidenzbasierten Leitlinien – vom Mythos zur rationalen Strategie. Zeitschrift Für Ärztliche Fortbildung Und Qualität Im Gesundheitswesen. *German Journal for Quality in Health Care, 101,* 89–95.

Kramer, M. W., Kuo, C. L., & Dailey, J. C. (1997). The Impact of Brainstorming Techniques on Subsequent Group Processes. *Small Group Research, 28,* 218–242.

Lanceley, A., Savage, J., Menon, U., & Jacobs, I. (2008). Influences on multidisciplinary team decision-making. *International Journal of Gynecologic Cancer, 18,* 215–222.

Landeta, J., Barrutia, J., & Lertxundi, A. (2011). Hybrid Delphi: A methodology to facilitate contribution from experts in professional contexts. *Technological Forecasting and Social Change, 78*, 1629–1641.

Larson, J., Christensen, C., Franz, T., & Abbott, A. (1998). Diagnosing groups: the pooling, management, and impact of shared and unshared case information in team-based medical decision making. *Journal of Personality and Social Psychology, 75,* 93–108.

Larson, J R, Christensen, C., Abbott, A. S., & Franz, T. M. (1996). Diagnosing groups: charting the flow of information in medical decision-making teams. *Journal of Personality and Social Psychology, 71,* 315–330.

Larson, James R., Foster-Fishman, P. G., & Keys, C. B. (1994). Discussion of shared and unshared information in decision-making groups. *Journal of Personality and Social Psychology, 67,* 446–461.

Levi, D. (2001). *Group dynamics for teams. In Group dynamics for teams.* California, US: Sage Publications, Inc.

Levy, S. J. (1981). Intepreting Consumer Mythology: A Structural Approach to Consumer Behavior. *Journal of Marketing, 45,* 49.

Lipman, T. (2000). Power and influence in clinical effectiveness and evidence-based medicine. *Family Practice, 17,* 557–563.

López, L. (2004). *Conflict Resolution and Group Decision-Making Exploring the dynamics of conflict resolution at the group level. 22rd International Conference of the System Dynamics Society.* Costa Rica: Graduate School of Business.

Maier, N. R. F., & Maier, R. A. (1957). An experimental test of the effects of "developmental" vs. "free" discussions on the quality of group decisions. *Journal of Applied Psychology, 41,* 320–323.

Maznevski, M. L. (1994). Understanding Our Differences: Performance in Decision-Making Groups with Diverse Members. *Human Relations, 47,* 531–552.

McCroskey, J. C. (1977). Oral Communication Apprehension: A Summary Of Recent Theory and Research. *Human Communication Research, 4,* 78–96.

Mcdougal, J. A., Brooks, C. M., & Albanese, M. (2005). Achieving Consensus on Leadership Competencies and Outcome Measures. *Evaluation & the Health Professions, 28,* 428–446.

Mercier, H., Trouche, E., Yama, H., Heintz, C., & Girotto, V. (2014). Experts and Laymen Grossly Underestimate the Benefits of Argumentation for Reasoning. *SSRN Electronic Journal, 21,* 341–355.

Miner, F. C. (1979). A Comparative Analysis of Three Diverse Group Decision Making Approaches. *The Academy of Management Journal, 22,* 1.

Minson, J. A., Liberman, V., & Ross, L. (2011). Two to Tango. *Personality and Social Psychology Bulletin, 37, 1325–1338.*

Moscovici, S., & Doise, W. (1994). *Conflict and consensus : a general theory of collective decisions.* Sage Publications: London.

Murphy, MK K, Black, N. A., Lamping, D. L., McKee, C. M., Sanderson, C. F., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health Technology Assessment, 2,* 1–88.

Myers, D. G., & Twenge, J. (2015). *Exploring social psychology (8th ed.).* New York: McGraw-Hill Education.

National Institute for Health and Care Excellence. (2014). *Developing NICE guidelines: the manual.* Process and Methods Guides, p. 245.

National Institute for Health and Clinical Excellence. (2007). *The guideline development process: an overview for stakeholders, the public and the NHS*. UK: NICE.

Nemeth, C. J., & Staw, B. M. (1989). The Tradeoffs of Social Control and Innovation in Groups and Organizations. *Advances in Experimental Social Psychology, 22,* 175–210.

Nemiroff, P. M., Pasmore, W. A., & Ford, D. L. (1976). The Effects Of Two Normative Structural Interventions On Established And Ad Hoc Groups: Implications For Improving Decision Making Effectiveness. *Decision Sciences, 7,* 841–855.

Nicholson, G. J., Pugliese, A., & Bezemer, P.-J. (2017). Habitual Accountability Routines in the Boardroom: How Boards Balance Control and Collaboration. *Accounting, Auditing & Accountability Journal, 30 2,* 222-246

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.

Nicollier-Fahrni, A., Vader, J.-P., Froehlich, F., Gonvers, J.-J., & Burnand, B. (2003). Development of appropriateness criteria for colonoscopy: comparison between a standardized expert panel and an evidence-based medicine approach. *International Journal for Quality in Health Care, 15*, 15–22.

Nordberg, D., & Booth, R. (2017). *Evaluating board effectiveness: A review and framework for evaluation of corporate boards,* presented at the British Academy of Management conference, Warwick, 2017. Warwick: The international journal of business in society.

Okhuysen, G. A. (2001). Structuring change: Familiarity and formal interventions in problem-solving groups. *Academy of Management Journal, 44*, 794–808.

Pagliari, C., Grimshaw, J., & Eccles, M. (2001). The potential influence of small group processes on guideline development. *Journal of Evaluation in Clinical Practice, 7*, 165–173.

Parks, C. D., & Cowlin, R. (1995). Group discussion as affected by number of alternatives and by a time limit. *Organizational Behavior and Human Decision Processes, 62,* 267–275.

Parks, C. D., & Cowlin, R. A. (1996). Acceptance of Uncommon Information into Group Discussion When That Information Is or Is Not Demonstrable. *Organizational Behavior and Human Decision Processes, 66,* 307–315.

Parks, C. D., & Sanna, L. J. (1999). *Group performance and interaction. In Group performance and interaction.* Boulder, CO, US: Westview Press.

Pavitt, C. (2009). A Sympathetic Reaction to the SM and DLCM as Group Communication Theories. *Human Communication Research, 35,* 272–278.

Poole, M. S., & Roth, J. (1989). Decision Development in Small Groups V Test of a Contingency Model. *Human Communication Research, 15,* 549–589.

Postmes, T., & Lea, M. (2000). Social processes and group decision making: anonymity in group decision support systems. *Ergonomics, 43,* 1252–1274.

Postmes, T., Spears, R., Sakhel, K., & De Groot, D. (2001). Social influence in computer-mediated communication: The effects of anonymity on group behavior. *Personality and Social Psychology Bulletin, 27,* 1243–1254.

Powell, C. (2003). The Delphi technique: myths and realities. *Journal of Advanced Nursing, 41,* 376–382.

Rescher, N. (1998). *Predicting the future : an introduction to the theory of forecasting.* New York: New York Press.

Roch, S. G., & Ayman, R. (2005). Group decision making and perceived decision success: The role of communication medium. *Group Dynamics, 9,* 15–31.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35,* 677–688.

Rowe, G., Wright, G, & Bolger, F . (1991). The Delphi Technique: A Re-Evaluation of Research and Theory. *Technological Forecasting and Social Change, 39, 3,* 235-251

Rycroft-Malone, J., Seers, K., Titchen, A., Harvey, G., Kitson, A., & McCormack, B. (2003). What counts as evidence in evidence-based practice? *Nursing and Health Care Management and Policy, 47,* 81–90.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal, 312,* 71–72.

Scheidel, T. M., & Crowell, L. (1964). Idea development in small discussion groups. *Quarterly Journal of Speech, 50,* 140–145.

Scholten, L., van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. W. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology, 43,* 539–552.

Schulz-Hardt, S., Brodbeck, F. C., Mojzisch, A., Kerschreiter, R., & Frey, D. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology, 91,* 1080–1093.

Schweiger, D. M., Sandberg, W. R., & Ragan, J. W. (1986). Group Approaches For Improving Strategic Decision Making: A Comparative Analysis of Dialectical Inquiry, Devil's Advocacy, and Consensus. *Academy of Management Journal, 29,* 51–71.

Shekelle, P G, & Schriger, D. L. (1996). Evaluating the use of the appropriateness method in the Agency for Health Care Policy and Research Clinical Practice Guideline Development process. *Health Services Research, 31,* 453–468.

Shekelle, Paul G., Kahan, J. P., Bernstein, S. J., Leape, L. L., Kamberg, C. J., & Park, R. E. (1998). The Reproducibility of a Method to Identify the Overuse and Underuse of Medical Procedures. *New England Journal of Medicine, 338,* 1888–1895.

Smith, C. M., Tindale, R. S., & Steiner, L. (1998). Investment Decisions by Individuals and Groups in `Sunk Cost' Situations: The Potential Impact of Shared Representations. *Group Processes & Intergroup Relations, 1,* 175–189.

Smith, G. F. (1998). Idea-Generation Techniques: A Formulary of Active Ingredients. *The Journal of Creative Behavior, 32,* 107–134.

Somech, A. (2006). The Effects of Leadership Style and Team Process on Performance and Innovation in Functionally Heterogeneous Teams. *Journal of Management, 32,* 132–157.

Souder, W. E. (1977). Effectiveness of Nominal and Interacting Group Decision Processes for Integrating R&D and Marketing. *Management Science, 23,* 595–605.

Stasser, G. (1992). Information salience and the discovery of hidden profiles by decision-making groups: A "thought experiment." *Organizational Behavior and Human Decision Processes, 52,* 156–181.

Stasser, G., & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology, 63,* 426–434.

Stasser, G., Stewart, D. D., & Wittenbaum, G. M. (1995). Expert Roles and Information Exchange during Discussion: The Importance of Knowing Who Knows What. *Journal of Experimental Social Psychology, 31,* 244–265.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48,* 1467–1478.

Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology, 53,* 81–93.

Stephenson, B. Y., Michaelsen, L. K., & Franklin, S. G. (1982). An Empirical Test of the Nominal Group Technique in State Solar Energy Planning. *Group & Organization Studies, 7,* 320–334.

Stetler, C. B., Brunell, M., Giuliano, K. K., Morsi, D., Prince, L., & Newell-

Stokes, V. (1998). Evidence-based practice and the role of nursing leadership. *The Journal of Nursing Administration, 28,* 45–53.

Stewart, D. D., & Stasser, G. (1998). The sampling of critical, unshared information in decision-making groups: the role of an informed minority. *European Journal of Social Psychology, 28,* 95–113.

Sutton, R. I., & Hargadon, A. (1996). Brainstorming groups in context: Effectiveness in a product design firm. *Administrative Science Quarterly, 41,* 685–718.

Sypher, B., Bostrom, R., & Seibert, J. (1973). Listening, Communication Abilities, and Success at Work. *The Journal of Business Communication, 26,* 293–303.

Tobacman, J. K., Scott, I. U., Cyphert, S., & Zimmerman, B. (1999). Reproducibility of measures of overuse of cataract surgery by three physician panels. *Medical Care, 37,* 937–945.

Valacich, J. S., Jessup, L. M., Dennis, A. R., & Nunamaker, J. F. (1992). A conceptual framework of anonymity in Group Support Systems. *Group Decision and Negotiation, 1,* 219–241.

Van de, A. H., & Delbecq, A. L. (1974). The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes. *Academy of Management Journal, 17,* 605–621.

van de Bovenkamp, H. M., & Trappenburg, M. J. (2009). Reconsidering patient participation in guideline development. Health Care Analysis : HCA . *Journal*

of Health Philosophy and Policy, 17, 198–216.

Van de Ven, A. H., & Delbecq, A. L. (1974). The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes. *Academy of Management Journal, 17,* 605–621.

Vinokur, A., Burnstein, E., Sechrest, L., & Wortman, P. M. (1985). Group decision making by experts: Field study of panels evaluating medical technologies. *Journal of Personality and Social Psychology, 49,* 70–84.

Volkema, R. J. (1983). Problem Formulation in Planning and Design. *Management Science, 29,* 639–652.

Washington, D. L., Bernstein, S. J., Kahan, J. P., Leape, L. L., Kamberg, C. J., & Shekelle, P. G. (2003). Reliability of Clinical Guideline Development Using Mail-Only versus In-Person Expert Panels. *Medical Care, 41,* 1374–1381.

West, A. (2011). Using the delphi technique: Experience from the world of counselling and psychotherapy. *Counselling and Psychotherapy Research, 11,* 237–242.

White, S. E., Dittrich, J. E., & Lang, J. R. (1980). The Effects of Group Decision-Making Process and Problem Situation Complexity on Implementation Attempts. *Administrative Science Quarterly, 25,* 428.

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2018). Problem-Solving Phase Transitions During Team Collaboration. *Cognitive Science, 42,* 129–167.

Winquist, J. R., & Larson, J. R. . J. (1998). Information pooling: When it impacts group decision making. *Journal of Personality and Social Psychology, 74,*

371–377.

Wittenbaum, G. M., Hubbell, A. P., & Zuckerman, C. (1999). Mutual enhancement: Toward an understanding of the collective preference for shared information. *Journal of Personality and Social Psychology, 77,* 967–978.

Wood, M., Ferlie, E., & Fitzgerald, L. (1998). Achieving clinical behaviour change: a case of becoming indeterminate. *Social Science & Medicine, 47,* 1729–1738.

Wortman, P. M., Vinokur, A., & Sechrest, L. (1988). Do consensus conferences work? A process evaluation of the NIH Consensus Development Program. *Journal of Health Politics, Policy and Law, 13,* 469–498.

Wortman, P., Smyth, J., Langenbrunner, J., & Yeaton, W. (1998). Consensus Among Experts and Research Synthesis: A Comparison of Methods. International *Journal of Technology Assessment in Health Care, 14, 109–122.*

Wu, T., & Seidmann, A. (2015). *Are groups better than individuals at making decisions? An experimental study on newsvendor problem*. Proceedings of the Annual Hawaii International Conference on System Sciences, 2015-March, 4252–4261. IEEE Computer Society.

Yang, H. (2013). *Determining the effectiveness of the delphi method for quantifying the drivers of demand for health and social care* (Unpublished doctoral thesis). University of Southampton, Southampton.

Yearwood, J., & Stranieri, A. (2010). Group Structured Reasoning for Coalescing Group Decisions. *Group Decision and Negotiation, 19, 1,* 77–105

Yousuf, M. I. (2007). Using Experts` Opinions Through Delphi Technique. *Practical Assessment, Research & Evaluation. 12,* 4.

Part 3: Critical Appraisal

Introduction

This critical appraisal comprises reflections regarding the process of conducting the present research project. A theme that runs across all sections of this thesis is that of participation and collaboration. This appears from the perspectives of formal consensus methods, inter-disciplinary guideline committees, and bridging the gap between research, policy, and practice. When I began this thesis, I was initially unsure of the place of clinical psychology in supporting guideline development. I wondered whether other disciplines would be better equipped in this arena. However, through the course of the research, I have come to appreciate how the skills of a psychologist can be applied to policy, particularly since inter-disciplinary participation is so important.

As psychologists, we are adept at encouraging a shared understanding and meaningful knowledge exchange between disparate individual worldviews, appreciating that perspectives are socially constructed and context dependent. We are also accustomed to holding on to a position of "not knowing", and allowing for uncertainty and curiosity as a space for the fostering of creative ideas (e.g. Anderson & Goolishian, 1992; Cecchin, 1987). This is particularly in our work with systems, which is essentially the case in the arena of guideline development. As my confidence grew, I felt an increasing interest in applying psychology on a macro-level. This critical appraisal describes some of my reflections during the research process, drawing upon a research journal that I kept throughout the work.

The policy context

Breuer (2000) notes that researcher influence can begin at the stage of topic selection. Researchers may choose a topic because of being particularly attracted to a role or environment, or because of emotional or intellectual reasons. Policy development has been an interest of mine for many years, and largely stems from my work as a manager in Improving Access to Psychological Therapies (IAPT). IAPT is a government initiative to increase access to psychological treatment specifically for depression and anxiety disorders. As a member of middle management, I became acutely aware of the challenges encountered when translating policy to practice in a meaningful way, and I wanted to better understand how guidelines were developed on a governmental level.

My prior experiences contributed to my feelings of suspicion towards guideline formation. For instance, I held assumptions that guideline development was quite far removed from the practical understanding of guideline implementation. I also imagined committee meetings to be biased towards the individuals that were most dominant in the room due to having more professional or academic prestige, and that most individuals who joined committees would do so for personal motivations.

Once I began working at the National Guideline Alliance (NGA), my beliefs shifted through the process of exposing myself to as many committee meetings as possible. I was impressed by the thoughtful inclusion of service users in meetings, which has been noted to require careful consideration to be effective (van de Bovenkamp & Trappenburg, 2009). I was also struck by how passionate the clinicians in the team were about concerning the implementation of guidelines.

However, it did feel as if the structure surrounding discussions were unclear, and not systematically managed. This created considerable variation in efficiency between guideline groups and meetings. Thus, although some of my beliefs were changed, my feelings grew in support of a more formalised method of discussion such as in formal consensus methods.

I noticed my assumptions becoming triggered periodically throughout the investigation, and it was helpful to become aware of this so I could remain reflexive and reflective. For example, I realised that I felt surprised at the interest and support that the committee members showed towards the project during recruitment. This was shown by the large majority volunteering to be part of the study despite not receiving compensation for their time. The result was contrary to my expectation that committee members would be resistant to change and thus would not want to engage with a new process.

Authors have suggested that psychologists are well placed to support public policy (McKnight, Sechrest, & McKnight, 2005). Maton, Perkins, & Saegert (2006) discuss the challenges of interdisciplinary working for community psychologists, including managing heterogeneous organisational "life worlds" of cultures, languages, and priorities. At the NGA "life worlds" not only represented the differences between the disciplines of committee members and technical staff that I was working with, but also between the responsibilities I occupied as a clinical psychology trainee.

Within my role at the NGA I felt competing priorities as an honorary staff member, UCL researcher, and clinician. For instance, as a clinician I wanted to develop guidelines based on all types of research evidence, including non-RCT

evidence and observational studies. As a member of the NGA, I realised that this was not possible since searching the whole research base was not manageable. Instead, the research evidence that was presented to the committee needed to be "good enough". Authors have discussed how an increased allocation of resources has not followed pressures to amplify the methodological and procedural rigour for guideline development. To manage this, they have suggested a negotiation of guideline procedures that balance validity and efficiency (Browman, Somerfield, Lyman, & Brouwers, 2015). I found it helpful to draw on the experience of colleagues to help me understand the expectations of quality versus efficiency when developing guidelines. I felt this echoed my work as a clinician when goal setting with clients for treatment, that it was important to balance expectations with resources to ensure goals were manageable.

Maton and colleagues (2006) have suggested methods to mitigate some of the barriers to interdisciplinary working. These are having a shared sense of values, fostering positive working relationships, and external factors such as the physical proximity of interdisciplinary colleagues. Keeping these factors in mind, I embedded myself into the team by attending meetings and spending time with colleagues at the NGA. I discussed with the team what their hopes for the project would be, and how it could fit into the values and vision of the organisation. Barker, Pistrang, &Elliott (2002) identify the importance of collaboration in the early stages of the project. This was done as early as the proposal phases of the project, with an aim that it would therefore feel more shared and helpful. I also offered written and verbal information about the timetabling and remit of my project, and identified key members of staff within the NGA that could act as "project leads" to allow for continuity of the research considering I was only available on non-placement days.

The literature review

I produced a systematic review as part of my NGA role for a real guideline in post-natal care, which I hoped to use for my thesis systematic review. This process offered experience of some of the limitations of guideline development. The review concerned postnatal interventions for emotional attachment in the first eight weeks following birth. Interventions that began in the antenatal period were excluded. The specified population were healthy women and babies, with no history of or current mental health problems, substance misuse problems, or learning disability. The review aimed to compare different interventions for emotional attachment as measured when babies were 12-18 months of age to identify which were more effective.

The search returned many hits, but only one paper was appropriate for inclusion. There were many reasons given for this dearth of finally included studies. Firstly, interventions are usually not applied to healthy individuals who appear adequately adjusted. Secondly, the construct of emotional attachment appears only in the context of separation anxiety, which is displayed in babies from 6 months of age. Therefore, research could only provide post-test data for emotional attachment which was unlikely to be conducted therefore by RCTs. Thirdly, interventions generally began during the antenatal period particularly to promote the retention of mothers in the study.

The above could be seen as an example of differences in approaches between governmental policy, economics and the research theory, also discussed by other researchers (Sanderson, 2004; Stevens, 2011). The view of stakeholders was that a guideline needed to focus on managing healthy individuals, but the research-base

centred more on supporting non-routine care. An economic viewpoint could offer a further alternative, since economically it is likely to be unfeasible to provide a cost-effective intervention for healthy individuals. This is because the intervention would not target a specific problem and is likely to be beneficial to a relatively small number of women.

These factors further highlighted the complexities of interdisciplinary working. In this case, the question that had been formulated was unsuitable for meaningful implementation. This had occurred because the guideline group had not sought comment from stakeholders and internal expertise for the purposes of correcting the question. What was essential was meaningfully and effectively integrating the information from those disciplines to create a whole that was greater than the sum of its parts. This included, for example, identifying when research questions and policy questions are not a good fit in the first instance to enable effective use of clinical expertise and the research base. These issues are at the heart of the ethos of formal consensus methods and the effective participation of members in group decision-making.

The implementation of the project

Ecologically valid research requires negotiation between the feasibility of implementation and generalisation of findings (Barker, Pistrang, & Elliott, 2002). The NGA did not have a detailed manual for the use of NGT, and the NGT procedure that was adopted deviated from the standard NGT technique. For example, recommendations from other guidelines were compiled and used as statements rather than drawing on empirical research to develop the consensus statements. This raised doubts within the committee about the usefulness of NGT, particularly because it

was not clear what evidence the statements were based on as their primary source. The NGA attempted to quality appraise the guidelines used by using the Appraisal of Guidelines for Research and Evaluation (AGREE-II Brouwers et al., 2010) criteria. However, the committee found this of limited use since it was drawing upon evidence that had already been interpreted by another committee.

Inconsistencies in the implementation and reporting between formal consensus studies including NGT has previously been raised as a limitation (see Humphrey-Murto et al., 2017). Through discussions with my supervisors and meetings with the NGA, it was identified that it would be most helpful to evaluate established practice and treat the research as a pilot study. A procedural manual could then be developed alongside practice to encourage replication. The aim would be to facilitate the ongoing evaluation of current practice and drive quality improvement over time.

There has been uncertainty at the NGA on when it is most appropriate to use NGT, with some committees opting for its use in the instance of broad and complex evidence rather than limited evidence. Encouraging the evaluation of practice could help justify deviations from the recommended use of NGT by the literature. This could be important to establish since it is common for practice to develop faster than the research literature. Indeed, there have been general recommendations for "practice-based research" to encourage collaboration between consumers and producers of the evidence-base for more effective research outcomes (Ammerman, Smith, & Calancie, 2014).

The majority of NGA and committee members volunteered to be interviewed. Interviewees spoke about finding the questions interesting, and many said that they

had not considered how formal processes could be applied in a healthcare guideline setting. Several members of the technical team and also of the committee asked for repeated reassurance that their utterances would remain confidential. I wondered whether this reflected a pressure to conform to current practices within NICE guidelines. These experiences emphasised the importance of qualitative work in a policy context, where individuals might feel unable to share their views openly.

The interpretation of data

Qualitative research involves subjective interpretation, which by necessity leads to a biased and partial account of data. Reflexivity has been suggested as a tool for the critical implementation of qualitative research (Finlay, 2002). It requires researchers to consider themselves as part of the social world that they are studying. My beliefs are likely to have impacted the results in many ways, from my interactions with the committee and NGA staff, to the development of the interview schedule questions.

Completing the literature review led me to feel aligned with the Nominal Group Technique (NGT), which was solidified by attending committee meetings that used informal consensus methods. When attending guideline committee groups prior to the commencement of the study, I felt frustration at the circular nature of some of the discussions, and thought that NGT could be a helpful tool to manage this. Through offering a presentation of the NGT method to the committee group, it may have been assumed that I was in agreement with introducing the technique. Indeed, I found that members would approach me generally to voice their support of it. The repeated experience of staff members approaching me to voice their support of the method further reinforced my positive NGT beliefs. I discussed my feelings of

discomfort at becoming more polarised towards NGT in supervision. I decided to attempt to achieve a more balanced perspective by seeking out individuals that are sceptical of the utility of formal consensus methods. Despite continuing to favour NGT, I believe opening up the conversation more widely with team members even those that were not included in the study, helped foster a broader perspective of NGT as situated within a larger structure of guideline development. These factors needed be considered as part of the research and some of these are outlined below.

The "subjectivity statement" aimed to describe potential biases by the author. The process of "bracketing" has been suggested to be helpful for increasing reflexivity (Speziale & Carpenter, 2011). Bracketing can be defined as the process of enhancing the awareness and therefore allowing for the suspension of pre-conceived attitudes towards the data by the researcher through their explicit identification. However, the interpretation of the data remains researcher-dependent, and so therefore "bracketing" assumptions could have limited impact on the subjectivity of the data analysis (J. A. Smith, Flowers, & Larkin, 2009). A process of triangulation therefore can help to mitigate potential biases by obtaining evidence from multiple methods, investigators, sources, and theories to identify corroborating evidence (Barker, Pistrang, & Elliott, 2002).

This research project introduced multiple perspectives into the process in order to reduce the potential for systematic biases. This was achieved through the inclusion of an external researcher in the design and implementation phases, such as data collection, analysis, and development of the interview schedule. The aim was not to arrive at an objective "truth", since this is not deemed possible given the relativistic nature of qualitative research. Rather, it was to encourage the explicit

sharing and discussion of interpretations to render them more conscious and thus increase the potential for research reflexivity. The addition of an external researcher also allowed for some of the interviews to be completed by a person who was unfamiliar to the committee. This could have encouraged openness in interviewee responses, particularly in relation to negative views about NGT. It is of note however that there were no systematic differences in the prevalence of themes found between investigators, which might have been expected if there were consistent differences in responses. An additional method of triangulation was that of using a mixed methods approach. This allowed for the corroboration of the themes found by taking into account the quantitative scoring of the views of the whole committee and comparing them to the subset that were interviewed. I found these triangulation elements helpful because of how transparent they allowed decision-making to be as codes were being developed. It managed my feelings of anxiety when codes were moving from being initially developed towards higher levels of abstract further removed for the data. In these instances it was particularly helpful to see that that there was a clear route to the data that had been corroborated and agreed upon by another researcher or by other data evidence.

One of the decisions that needed to be made was regarding the level of data interpretation at the analysis stage. A semantic rather than a latent analysis was chosen to remain as close to the data as possible rather than interpreting beyond participant utterances. There were several reasons for this decision. Firstly, a semantic approach seemed to fit the research question best, which was looking at expectations of NGT rather than the subjective experiences of participants. Secondly, a semantic approach mirrored what had been adopted by previous research in other fields, and so could be more easily cross-examined for consistency. This was the case

also for inter-researcher reliability when comparing themes, since they could be easily refuted or supported by the data. An alternative approach to use could have been framework analysis (Ritchie & Spencer, 1994) by using thematic categories developed from the literature review. It was thought that the subject matter of the literature review was too separate from the focus of the main research project for framework analysis to be helpful as a primary approach. Namely, the question of participant experiences of formal consensus methods was predicted to be insufficiently overlapping with participant expectations of using NGT. There was also no research identified by the search that focused on participant expectations of NGT, and so it was not clear how an appropriate framework could be derived. Nevertheless, the thematic analysis from the conceptual introduction did influence the development of data categories for the main research paper, and so the themes were in part grounded in previous literature of general experiences of formal consensus methods.

Limitations and Recommendations

Limitations and recommendations have been discussed throughout this thesis. Particular threats to external validity in the research concern the specific method for NGT as used by the NGA. Thus, caution should be taken when generalising the findings to NGT and the data should be considered as part of a feasibility project. However, the consistency of themes between the present study and the literature review more generally offers encouragement that the conclusions could be broadly relevant to NGT processes. Manualising the processes of the NGT method could help establish more consistent application of NGT to develop a better understand of how it can be most effective for guideline development.

A model that might be useful for developing NGT at the NGA could be that of participatory action research (PAR; Baum, MacDougall, & Smith, 2006). The ethos of PAR is that participants and researchers work together in a cycle of audit, collecting data and changing practice. This reduces power differentials between researcher and participant and encourages equal participation and ownership over improved consensus processes, which agrees with the aims of formal consensus methods more generally. Since guideline development requires the inclusion of a range of individuals, using a research process that could represent the maximum number of individuals in an inclusive way seems most appropriate.

Conclusions

The experience of conducting the research was a very rich learning opportunity for me, and I felt I was able to work in an area uncommonly occupied by clinical psychologists. This enhanced my appreciation for cross-disciplinary working for applied psychology. The views and experiences expressed in this research add to a novel area of understanding into best practice for the use of formal consensus methods in healthcare guideline development.

References

Ammerman, A., Smith, T. W., & Calancie, L. (2014). Practice-Based Evidence in Public Health: Improving Reach, Relevance, and Results. *Annual Review of Public Health, 35,* 47–63.

Anderson, H., & Goolishian, H. (1992). The client is the expert: a not knowing approach to therapy | Anglia Ruskin University Library. In S. McNamee & K. Gergen (Eds.), *Therapy as social construction.* London: Sage.

Barker, C., Pistrang, N., & Elliott, R. (2002). *Research methods in clinical psychology : an introduction for students and practitioners* (Third Edit). Chichester: J. Wiley.

Baum, F., MacDougall, C., & Smith, D. (2006). Participatory action research. *Journal of Epidemiology and Community Health, 60,* 854–857.

Breuer, F. (2000). Qualitative Methods in the Study of Biographies, Interactions and Everyday Life Contexts: The Development of a Research Style. *Qualitative Social Research, 1, 2,* 3

Brouwers, M. C., Kho, M. E., Browman, G. P., Burgers, J. S., Cluzeau, F., Feder, G., Fervers, B., Graham, I. D., Grimshaw, J., Hanna, S. E., Littlejohns, P., Makarski, J., Zitzelsberger, L. (2010). AGREE II: Advancing guideline development, reporting and evaluation in health care. *Journal of Clinical Epidemiology, 63, 12,* 1308-1311.

Browman, G. P., Somerfield, M. R., Lyman, G. H., & Brouwers, M. C. (2015). When is good, good enough? Methodological pragmatism for sustainable

guideline development. *Implementation Science, 10, 28*, 1-5.

Cecchin, G. (1987). Hypothesizing, Circularity, and Neutrality Revisited: An Invitation to Curiosity. *Family Process, 26,* 405–413.

Finlay, L. (2002). "Outing" the Researcher: The Provenance, Process, and Practice of Reflexivity. *Qualitative Health Research, 12,* 531–545.

Humphrey-Murto, S., Varpio, L., Wood, T. J., Gonsalves, C., Ufholz, L.-A., Mascioli, K., Wang, C., Foth, T. (2017). The Use of the Delphi and Other Consensus Group Methods in Medical Education Research. *Academic Medicine, 92,* 1491–1498.

Maton, K. I., Perkins, D. D., & Saegert, S. (2006). Community Psychology at the Crossroads: Prospects for Interdisciplinary Research. *American Journal of Community Psychology, 38,* 35–49.

McKnight, K. M., Sechrest, L., & McKnight, P. E. (2005). Psychology, Psychologists, and Public Policy. *Annual Review of Clinical Psychology, 1,* 557–576.

Ritchie, J., & Spencer, L. (1994). Qualitative Data Analysis for Applied Policy Research. In A. M. Huberman & M. Miles (Eds.), *The Qualitative Researcher's Companion* (pp. 305–329). California: SAGE Publications, Inc.

Sanderson, I. (2004). Getting Evidence into Practice. *Evaluation, 10,* 366–379.

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative phenomenological analysis : theory, method, and research.* London: Sage Publications.

Speziale, H. S., & Carpenter, D. R. (2011). *Qualitative research in nursing : advancing the humanistic imperative (Fifth Edit).* Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Stevens, A. (2011). Telling Policy Stories: An Ethnographic Study of the Use of Evidence in Policy-making in the UK. *Journal of Social Policy, 40*, 237–255.

van de Bovenkamp, H. M., & Trappenburg, M. J. (2009). Reconsidering patient participation in guideline development. Health Care Analysis : HCA . *Journal of Health Philosophy and Policy, 17,* 198–216.

Appendices

Conceptual Introduction Appendices

Appendix A: Table illustrating conceptual introduction themes at the early stages

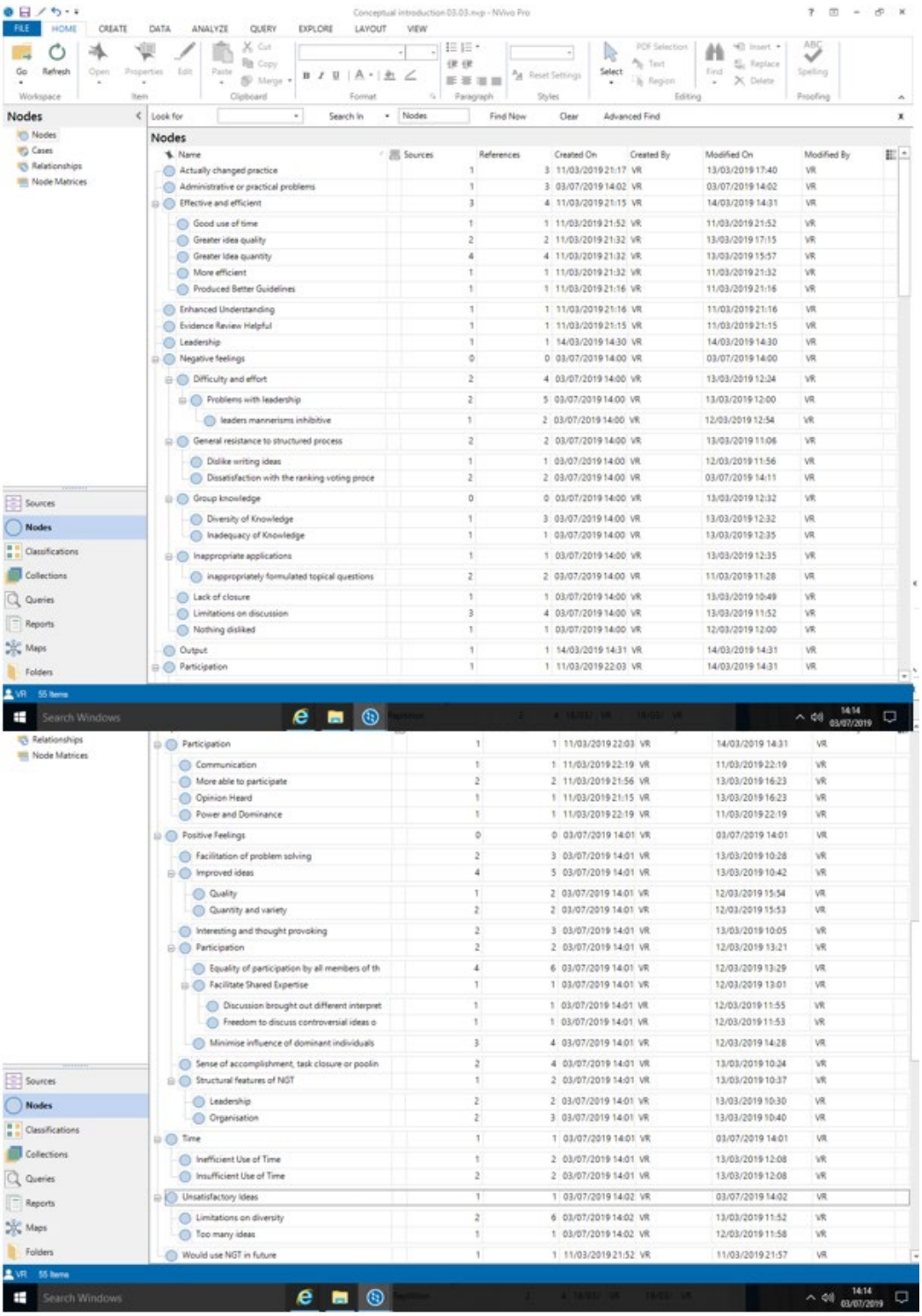
Formal consensus method	Group of themes	Themes	Subthemes
Delphi			
	Negative	Administration difficulties	
		Anonymity	
		Attributes of Delphi	
		Effortful	Time consuming
			Repetition
		Dislike of questions	
		Lack of closure	
		Lack of face to face	
	Positive	Anonymity	
		Efficient	
		Low Effort	
		Participation	Interesting knowledge exchange
			Sharing of expertise
		Positive feelings	
	Quantitative data	Ideas	Quantity and variety of ideas
		Effective and efficient	
NGT	Positive	Changes beyond the method	Actually changed

			practice Enhanced understanding Would use NGT in future
		Effective and efficient	Good use of time More efficient Produced better guidelines
		Evidence review helpful	
		Ideas	Greater idea quality Greater idea quantity
		Participation	Communication Opinion heard Power and dominance Equality of participation Facilitating shared expertise Freedom to discuss controversial ideas
		Facilitation of problem solving	
		Improved ideas	Improved quality Improved quantity and variety
		Positive feelings	Interesting and thought

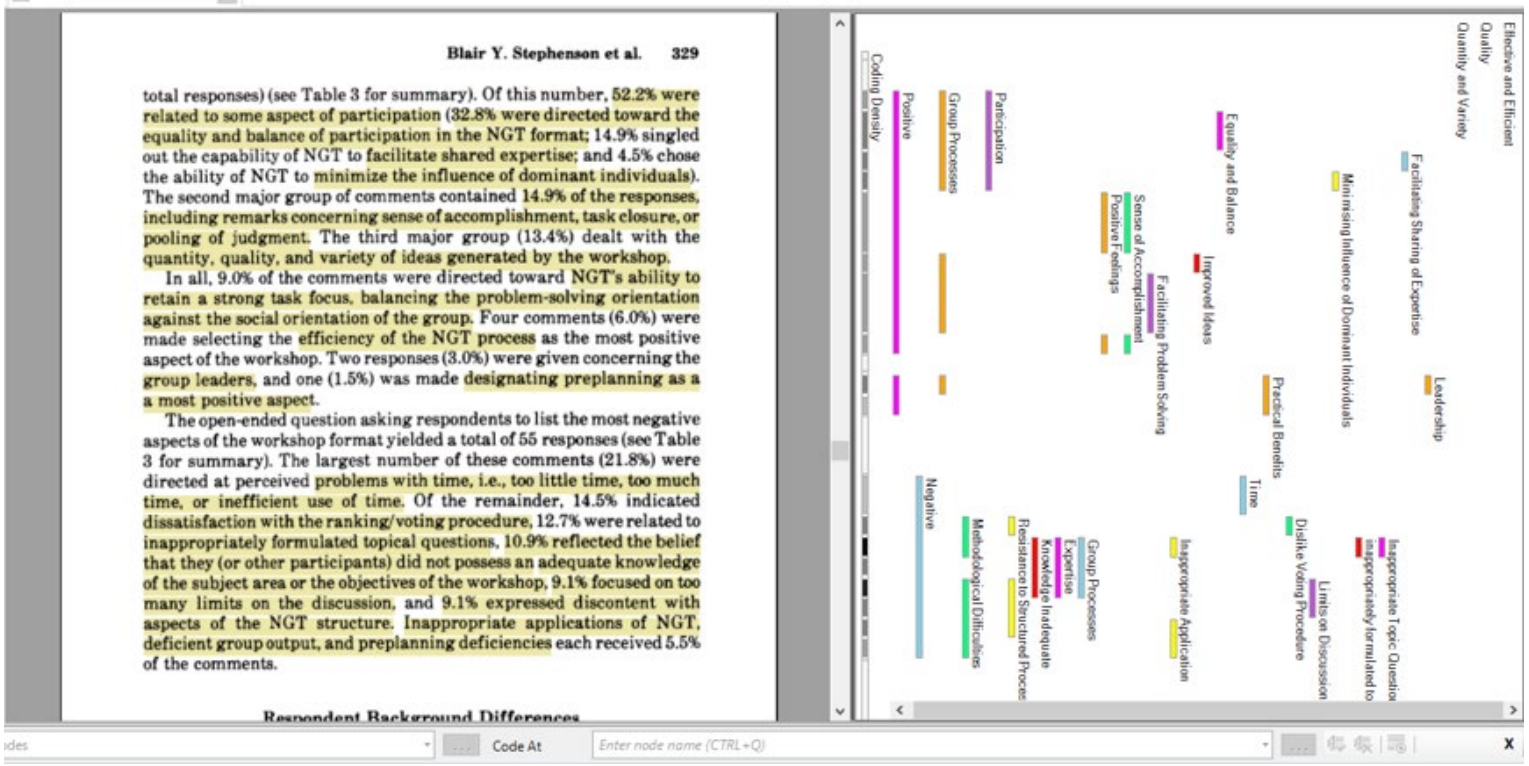
			provoking
			Sense of accomplishment
			Feeling of satisfaction
		Structural features	Organisation
			Leadership
	Negative	Administrative or practical problems	
		Difficulty and effort	Problems with leadership
		General resistance to structured process	Dislike writing ideas
			Dissatisfaction with the ranking or voting procedure
		Group knowledge	Knowledge too diverse
			Knowledge inadequate
		Inappropriate applications	Inappropriately formulated topic questions
		Lack of closure	
		Limitations on discussion	
		Nothing disliked	
		Time	Inefficient use of time
			Insufficient time
		Unsatisfactory ideas	Limitations on diversity

			Too many ideas
	Quantitative	Participation	Equality and balance
			Minimising influence of dominant individuals
		Improved ideas	Quantity and variety
			Quality of ideas
		Positive feelings	Sense of accomplishment
			Interesting and thought provoking
		Effective and efficient	
RAND	Positive	Participation	Sharing of expertise
		Ideas	Greater quantity and quality of ideas
		Positive feelings	Better method
	Negative	Greater effort	
	Quantitative data	Participation	Facilitated sharing of expertise
		Positive feelings	Sense of accomplishment
		Difficult and effortful	

Appendix B: Conceptual introduction later stage illustration of themes as displayed in NVIVO qualitative software



Appendix C: Example of analysis of one of the transcripts when coding during the literature review



Appendix D: Table presenting a more detailed account of the distribution of themes

Themes	Subthemes	NGT				Delphi									RAND	Hybrid	
Themes	Subthemes	NGT				Delphi									RAND	Hybrid	
		Qualitative				Quantitative						Qualitative			Quant	Quant	Quant
		V	St	G	K	B	V	St	Ru	G	K	V2	Y	W	V2	Sh	M	L
Positive	Participation																	
	Equality and Balance																	
	Minimising influence of dominant individuals																	
	Facilitated sharing of expertise																	
	Low Effort																	
	Anonymity																	
	Improved Ideas																	
	Quantity and Variety																	
	Quality of ideas																	
	Positive feelings from participants																	
	Sense of accomplishment																	
	Interesting and thought provoking																	
	Leadership																	
	Administrative and Practical Benefits																	
	Facilitation of problem-solving or achieving consensus																	
	Effective and Efficient																	

Themes	Subthemes		NGT										Delphi				RAND	Hybrid	
			Qualitative				Quantitative						Qualitative			Quant	Quant	Quant	
			V	St	G	K	B	V	St	Ru	G	K	V2	Y	W	V2	Sh	M	L
Negative	**Difficulty and Effort**																		
		Problems with Leadership																	
		Group and Individual Knowledge Too Diverse																	
		Group and Individual Knowledge Inadequate																	
		Repetitive																	
		Misunderstanding of questions																	
	Limits on discussion																		
	Lack of face to face																		
		Absence of real-time feedback																	
		Impersonal																	
		Lacking stimulation																	
	Unsatisfactory Ideas																		
		Limitations on diversity of ideas																	
		Too many ideas																	
	Resistance to structured process																		
		Recording of ideas																	
		Voting procedure																	
	Time																		
		Insufficient Use of Time																	
		Time consuming																	
	Inappropriate application																		
		Inappropriate Topic Questions																	
	Administrative and Practical Problems																		
		Problems with Attrition																	
	Lack of closure																		
	Nothing Disliked																		

Empirical paper appendices

Appendix E: Consent form for participants

UCL DIVISION OF
PSYCHOLOGY
AND LANGUAGE
SCIENCES
1-19 TORRINGTON
PLACE
WC1E 7HB

CONSENT FORM FOR COMMITTEE MEMBERS IN RESEARCH STUDIES

Please complete this form after you have read the Information Sheet and/or listened to an explanation about the research.

Title of Study: Implementation and evaluation of formal consensus methods of clinical guidelines

Department: Research Department of Clinical, Educational & Health Psychology

Name and Contact Details of the Researcher: Victoria Roberts
Telephone: [redacted]

Name and Contact Details of the Principal Researcher: Professor Stephen Pilling

Telephone: +44 [redacted]

Name and Contact Details of the UCL Data Protection Officer: __ Lee Shailer l.shailer@ucl.ac.uk____

This study has been approved by the UCL Research Ethics Committee: Project ID number: CEHP2018569

Thank you for considering taking part in this research. The person organising the research must explain the project to you before you agree to take part. If you have any questions arising from the Information Sheet or explanation already given to you, please ask the researcher before you decide whether to join in. You will be given a copy of this Consent Form to keep and refer to at any time.

I confirm that I understand that by ticking/initialling each box below I am consenting to this element of the study. I understand that it will be assumed that unticked/initialled boxes means that I DO NOT consent to that part of the study. I understand that by not giving consent for any one element that I may be deemed ineligible for the study.

		Tick Box
	I confirm that I have read and understood the Information Sheet for the above study. I have had an opportunity to consider the information and what will be expected of me. I have also had the opportunity to ask questions which have been answered to my satisfaction and would like to take part in: - a 20-30 minute interview about my experiences of group decision-making	
	I understand that I will be able to withdraw my data up to 4 weeks after interview	
	I understand that my data gathered in this study will be stored anonymously and securely on an audio recorder. It will not be possible to identify me in any publications. The file on the audio recorder will be transferred onto password protected computers on site. It will be transported via password-protected encrypted USB sticks. The file will be labelled using a pseudonym that will be linked to your name in a separate file stored in a separate place under password protection.	
	I understand that my information may be subject to review by responsible individuals from the University (Professor Steve Pilling) for monitoring and audit purposes.	
	I understand that my participation is voluntary and that I am free to withdraw at any time without giving a reason, without my legal rights being affected. I understand that if I decide to withdraw, any personal data I have provided up to that point will be deleted unless I agree otherwise.	
	I understand the potential risks of participating. In the unlikely event of becoming distressed during this research, I understand how to access the support that will be available to me.	
	I understand that although participating will not benefit me directly, it will allow me to anonymously contribute to a growing body of research on this area.	
	I understand that the data will not be made available to any commercial organisations but is solely the responsibility of the researcher(s) undertaking this study.	
	I understand that I will not benefit financially from this study or from any possible outcome it may result in in the future.	
	I agree that my anonymised research data may be used by others for future research. No one will be able to identify you when this data is shared.	
	I understand that the information I have submitted will be published as a report and I wish to receive a copy of it. Yes/No	
	I am aware of who I should contact if I wish to lodge a complaint.	
	I voluntarily agree to take part in this study.	
	I understand that my information will be stored anonymously under password protection. I would be happy for the data I provide to be archived at University College London (UCL) in password protected UCL computers. I understand that my data will be stored in a separate place to my name. I understand that other authenticated researchers will have access to my anonymised data.	

If you would like your contact details to be retained so that you can be contacted in the future by UCL researchers who would like to invite you to participate in follow up studies to this project, or in future studies of a similar nature, please tick the appropriate box below.

	Yes, I would be happy to be contacted in this way	
	No, I would not like to be contacted	

_____ _____

Name of participant Date

Signature

_____ _____

Researcher Date

Signature

Appendix F: Information sheet for participants

UCL DIVISION OF PSYCHOLOGY
AND LANGUAGE SCIENCES

Participant Information Sheet For Committee Members

UCL Research Ethics Committee Approval ID Number: CEHP2018569

YOU WILL BE GIVEN A COPY OF THIS INFORMATION SHEET

Title of Study: Implementation and evaluation of formal consensus methods of clinical guidelines

Department: Research Department of Clinical, Educational & Health Psychology

Name and Contact Details of the Researcher: Victoria Roberts

v.roberts@ucl.ac.uk

Telephone: [redacted]

Name and Contact Details of the Principal Researcher: Professor Stephen Pilling

Telephone: [redacted] or ext. [redacted]

1. **Invitation Paragraph**
   You are being invited to take part in a research project that aims to understand the experiences of committee members involved in formal and informal consensus methods of guideline decision-making. Before you decide it is important for you to understand why the research us being done and what participation will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part. Thank you for reading this.

2. **What is the project's purpose?**
   Decision-making in guideline committees generally relies on methods of informal consensus. These methods work best when there is research evidence that supports a certain decision so that the group can more easily and transparently reach agreement. However, in healthcare it is sometimes the case that there is research evidence that is too limited or too broad to answer the question at hand. When this occurs, the group can be vulnerable to being influenced by the opinion of a few members, and can make recommendations that are not grounded in research evidence. Formal consensus methods of decision-making offer a structured, guided approach that could support decision-making in the context of limited research evidence.

   This project aims to measure the experience of committee members when using formal and informal consensus methods. Experience will be measured in two ways. Firstly, interviews will be carried out with committee members asking them about their experience of group decision-making. Secondly, the study will give committee members questionnaires asking them to rate their experiences after having used formal and informal consensus methods in order to identify any differences.

3. **Why have I been chosen?**
   You have been chosen because you form part of a committee group in which formal and/or informal consensus methods have been used. You are amongst up to 100 participants that will be asked to participate in this study.

4. **Do I have to take part?**
5. It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and be asked to sign a consent form. You can withdraw at any time without giving a reason. if you decide to withdraw you will be asked what you wish to happen to the data you have provided up that point.

6. **What will happen to me if I take part?**
   If you are asked to be interviewed, a time will be arranged that is convenient to you in time and place. This can be done in person or over the phone and will last 20-30 minutes. We cannot pay for travel expenses, and so we will endeavour to find a way that is convenient for you.

   It has been agreed which question would be best suited to use a formal or informal consensus method. The informal consensus part of this study will not require anything different from the group decision-making process. However, for the question for which it has been decided to use the formal consensus method, a variation of a formal consensus method called the Nominal Group Technique (NGT) will be used. In this technique, statements are generated from the research evidence that is available. The group will then meet to discuss the statements. Following discussion, each member will rate their agreement with the statements privately. This will be done on a Likert scale as in the example below.

Services for people with PTSD should be accessible to people from different cultural and religious backgrounds

Strongly Agree Strongly Disagree

1	2	3	4	5	6	7	8	9	10

Comments:_____

Following a first round of rating, responses will be collected. The statements will be managed depending on their level of agreement in the group. Those will high level of agreement will be kept for the guideline, and those with a low level of agreement will be rejected. The statements which are agreed upon 60-80% of the time will be rewritten in accordance with the comments people have left. They will then be brought back to the group for discussion. After another discussion, group members then can re-rate their agreement with the new statements. This gives a final opportunity for the inclusion or exclusion of the statements from the guideline.

After both the formal and informal consensus methods, you will be given a brief questionnaire that asks you about your experience of the group on a Likert scale. An example of this is below.

To what extent did you feel free to participate and contribute your opinion in the meeting?

Very low high Very

1	2	3	4	5	6	7	8	9	10

Comments:_____

7. **Will I be recorded and how will the recorded media be used?**
   For the interview part of this research, the audio recordings of your interview will be used only for analysis and for illustration in conference presentations and lectures. No other use will be made of them without your written permission, and no one outside the project will be allowed access to the original recordings. The interviews will be recorded and transferred from the audio recorder to password protected computers on site. They will be transported via password protected encrypted USB sticks. Your data file will be stored under a pseudonym that will be linked to your name using a key in a separate data file that will also be password protected and stored separately. This information may be monitored for research purposes by UCL researchers Professor Steve Pilling, Patrice Carter, and other members of the research team.

   For the formal and informal consensus part of this research, the questionnaires will be removed from the consent forms upon receipt and will be linked via a numbered "key" system. This key will be stored in a password protected data file that will be stored separately to the consent forms and questionnaires so that your data cannot be identified. The consent forms and questionnaires will be kept in a locked filing cabinet at UCL. Only the investigators will have access, and the data will be destroyed after 20 years.

8. **What are the possible disadvantages and risks of taking part?**
   There are no reasonable foreseeable discomforts, disadvantages and risks directly related to taking part in this research. However, should the need arise, you are encouraged to bring any difficulties that arise as a result of this research immediately to the attention of any of the investigators involved.

9. **What are the possible benefits of taking part?**
   Although there are no immediate benefits to taking part in this research, you will be contributing to a growing body of research in this area of what is important for decision-making in clinical guidelines.

10. **What if something goes wrong?**
    In the event that you would like to raise a complaint or discuss an aspect of the research further, please contact Victoria Roberts (principal researcher) on v.roberts@ucl.ac.uk. If you are not satisfied with the response, you may contact the Chair of the UCL Research Ethics Committee at ethics@ucl.ac.uk

11. **Will my taking part in this project be kept confidential?**
    All the information that we collect about you during the research will be kept strictly confidential. You will not be able to be identified in any ensuing reports or publications.

12. Limits to confidentiality

Please note that assurances on confidentiality will be strictly adhered to unless evidence of potential harm is uncovered. In such cases the University may be obliged to contact relevant statutory bodies/agencies.

13. What will happen to the results of the research project?

The results of this study will be provided via a report for NICE and the NGA. You may request a copy of the report if you wish.

14. Data Protection Privacy Notice

Notice:
The data controller for this project will be University College London (UCL). The UCL Data Protection Office provides oversight of UCL activities involving the processing of personal data, and can be contacted at data-protection@ucl.ac.uk. UCL's Data Protection Officer is Lee Shailer and he can also be contacted at data-protection@ucl.ac.uk.

Your personal data will be processed for the purposes outlined in this notice. The legal basis that would be used to process your personal data will be the provision of your consent. You can provide your consent for the use of your personal data in this project by completing the consent form that has been provided to you.

Your personal data will be processed so long as it is required for the research project. We will anonymise the personal data you provide us with. We will also endeavour to minimise the processing of personal data wherever possible.

If you are concerned about how your personal data is being processed, please contact UCL in the first instance at data-protection@ucl.ac.uk. If you remain unsatisfied, you may wish to contact the Information Commissioner's Office (ICO). Contact details, and details of data subject rights, are available on the ICO website at: https://ico.org.uk/for-organisations/data-protection-reform/overview-of-the-gdpr/individuals-rights/

Personal data will not be transferred outside the European Economic Area.

15. Who is organising and funding the research?
University College London are funding the research as part of a Doctorate in Clinical Psychology.

16. Contact for further information
Should you want further information on the project, you may contact the Researcher Victoria Roberts using the following details:
Telephone [redacted]
Email: v.roberts@ucl.ac.uk

Thank you for reading this information sheet and for considering to take part in this research study. You will be given a copy of this information sheet to keep for your records.

Appendix G: Pre-qualitative interview question schedule

Pre-Qualitative interview questions

General interview guidelines:

Prompt if participant gives one word answer.

Prompt "anything else?" after the participant has given their response before moving on to the next question.

-Just to remind you that this call will be recorded

-everything you say will be kept confidential and anonymised

-are you still happy to continue?

We are interested in formal consensus methods. We want to find out something about how they can be used. It is important to understand your experience but also the wider view of them as a whole. We wouldn't want to use something in NICE that is unacceptable, so we would appreciate if you could be as candid as possible in your responses. Your responses will be anonymised for data analysis.

1. What is your view of how decisions in this committee are made?

2. What is your experience so far of decision making when evidence is limited?

3. Do you think a more structured approach to decision making could be helpful, particularly when evidence is limited?

4. Can you tell me more about why you think that?

5. Are you familiar with formal and informal consensus methods?
    a. (if no) Offer definition: formal consensus methods are a structured approach to making decisions, and include Delphi, Nominal Group Technique, and RAND. A researcher will formulate evidence statements prior to the meeting of the committee about a topic based on pre-determined sources of evidence. These statements are then presented to each member of the committee who has an opportunity to rate them. These ratings are compiled and analysed for agreement. The statements that have the least agreement are revised based on feedback from the individuals. Individuals are then given the opportunity to re-rate the revised statements. Informal consensus methods offer no formal approach to the development of consensus statements once evidence is presented
    b. (if yes) can you tell me what your understanding is of formal consensus methods?

6. What do you think are the things that guide a committee to use a particular method, formal or informal?
    a. Anything else you can think of?

7. Do you have any experiences of using formal consensus methods?
8. If so, what are your own experiences of using formal consensus methods?

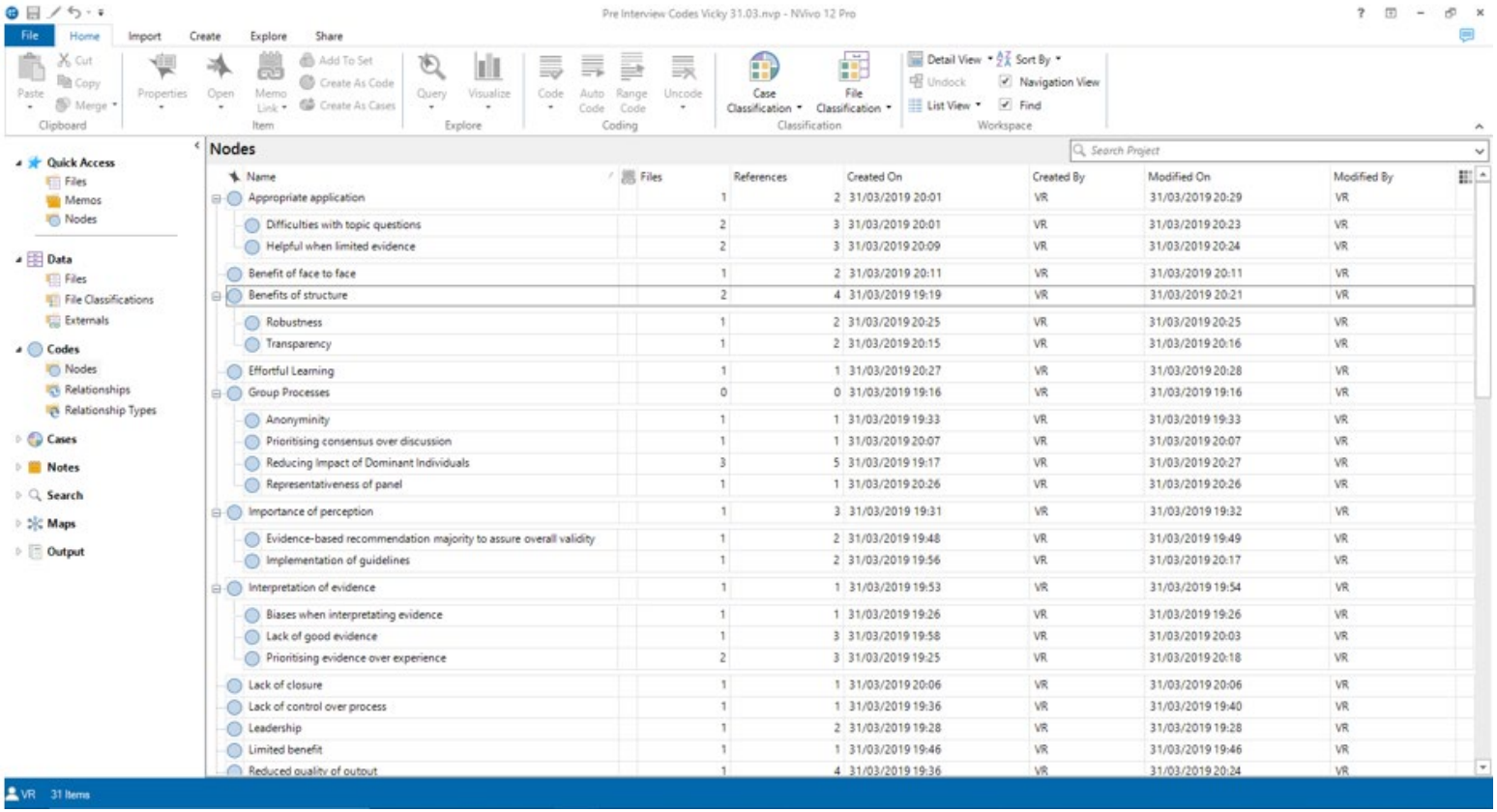
9. What do you think are the advantages and disadvantages of formal consensus methods?
10. What do you think is the impact of this (the disadvantage)?

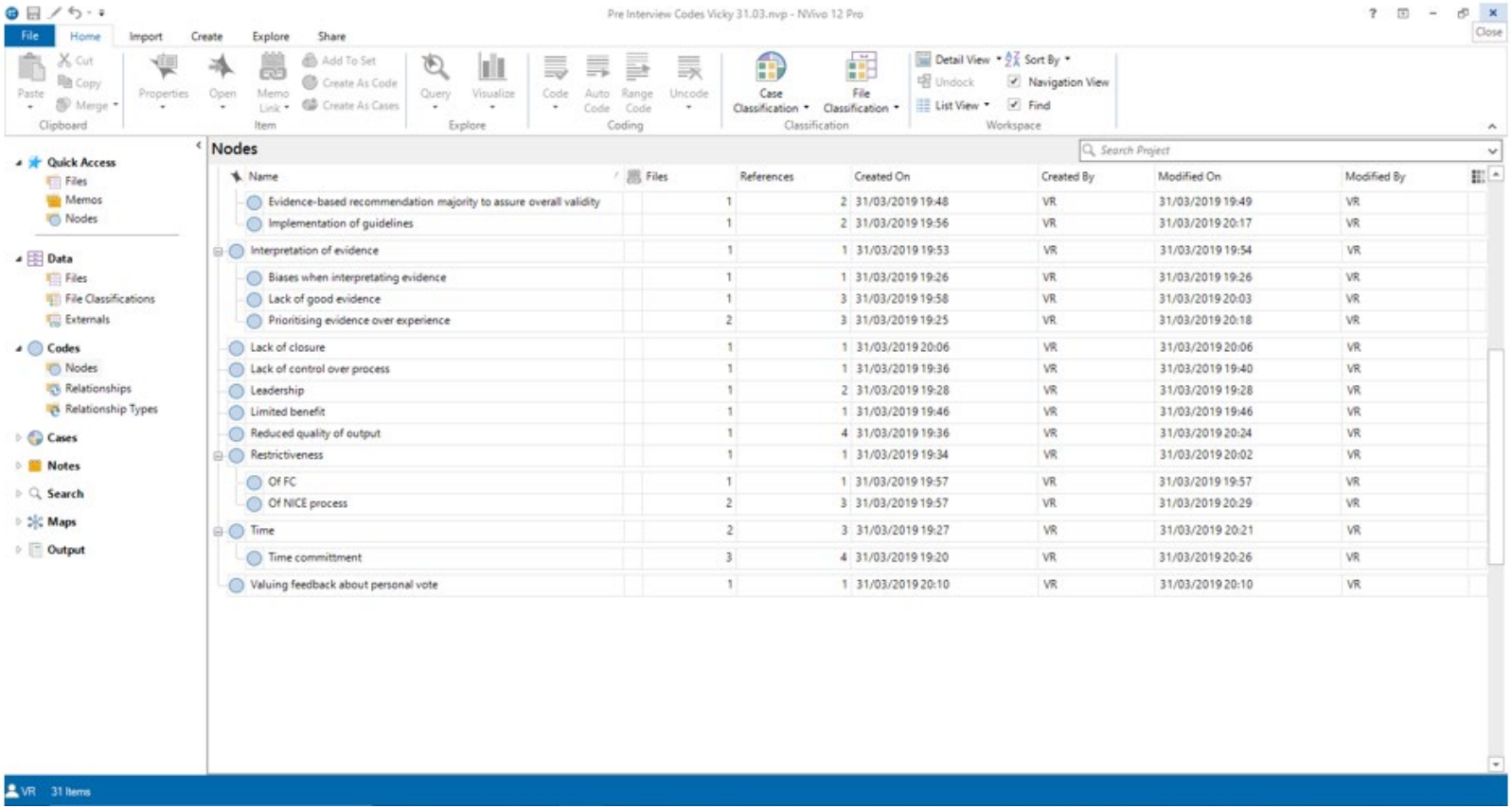
11. What is your own view of formal consensus methods in guidelines in general?

12. How do you think formal consensus methods might be viewed by your colleagues?

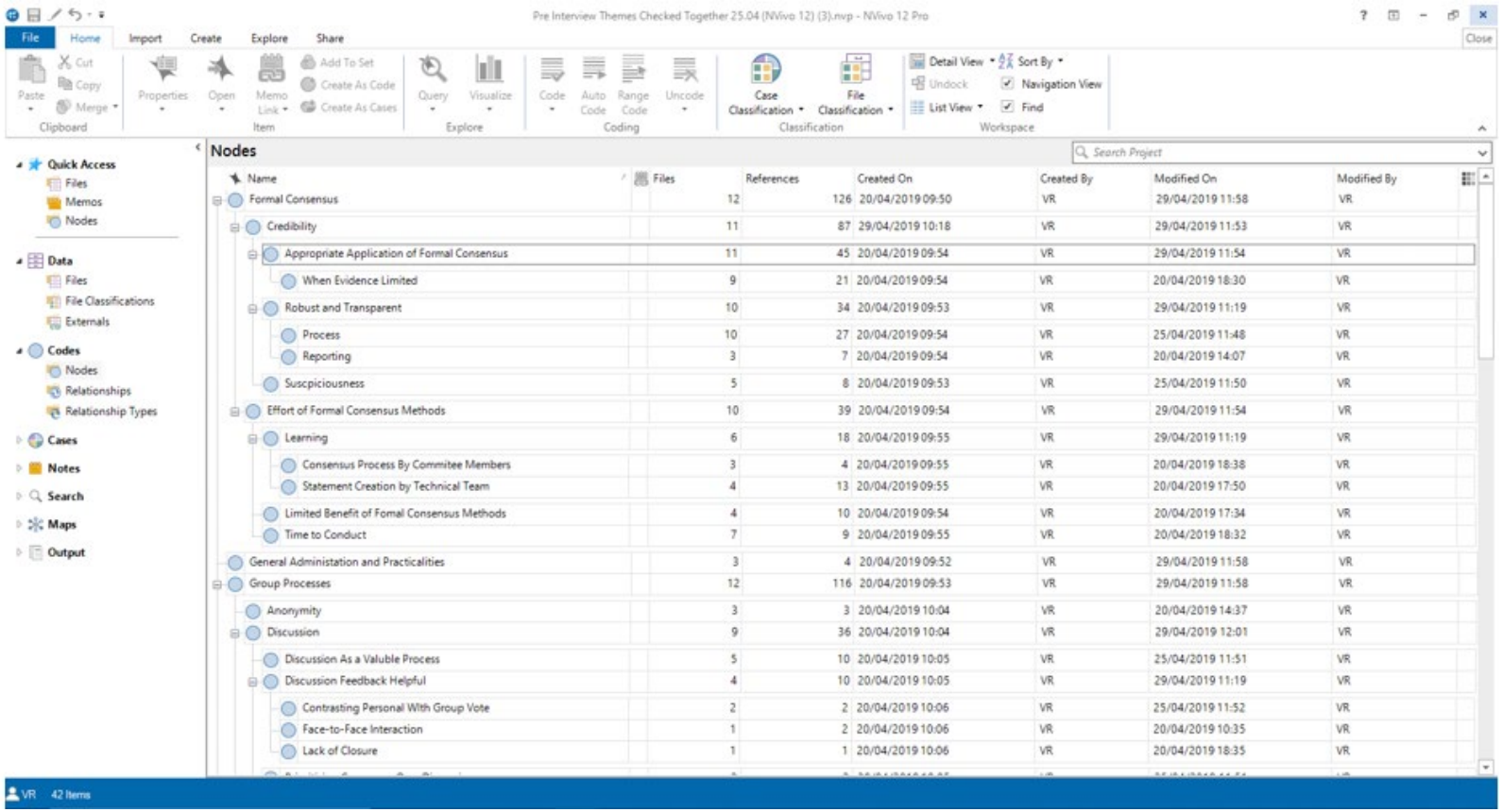
13. Do you think use of formal consensus methods will affect how the guideline is broadly perceived?

Appendix H: Screenshot of early stage themes of empirical paper as seen in NVIVO




Pre Interview Codes Vicky 31.03.nvp - NVivo 12 Pro
File
Home
Import
Create
Explore
Share
Close
Paste
Cut
Copy
Merge
Clipboard
Properties
Open
Memo Link
Add To Set
Create As Code
Create As Cases
Item
Query
Visualize
Explore
Code
Auto Code
Range Code
Uncode
Coding
Case Classification
File Classification
Classification
Detail View
Sort By
Undock
Navigation View
List View
Find
Workspace
Quick Access
Files
Memos
Nodes
Data
Files
File Classifications
Externals
Codes
Nodes
Relationships
Relationship Types
Cases
Notes
Search
Maps
Output
Nodes
Search Project
Name
Files
References
Created On
Created By
Modified On
Modified By
Evidence-based recommendation majority to assure overall validity 1 2 31/03/2019 19:48 VR 31/03/2019 19:49 VR
Implementation of guidelines 1 2 31/03/2019 19:56 VR 31/03/2019 20:17 VR
Interpretation of evidence 1 1 31/03/2019 19:53 VR 31/03/2019 19:54 VR
Biases when interpretating evidence 1 1 31/03/2019 19:26 VR 31/03/2019 19:26 VR
Lack of good evidence 1 3 31/03/2019 19:58 VR 31/03/2019 20:03 VR
Prioritising evidence over experience 2 3 31/03/2019 19:25 VR 31/03/2019 20:18 VR
Lack of closure 1 1 31/03/2019 20:06 VR 31/03/2019 20:06 VR
Lack of control over process 1 1 31/03/2019 19:36 VR 31/03/2019 19:40 VR
Leadership 1 2 31/03/2019 19:28 VR 31/03/2019 19:28 VR
Limited benefit 1 1 31/03/2019 19:46 VR 31/03/2019 19:46 VR
Reduced quality of output 1 4 31/03/2019 19:36 VR 31/03/2019 20:24 VR
Restrictiveness 1 1 31/03/2019 19:34 VR 31/03/2019 20:02 VR
Of FC 1 1 31/03/2019 19:57 VR 31/03/2019 19:57 VR
Of NICE process 2 3 31/03/2019 19:57 VR 31/03/2019 20:29 VR
Time 2 3 31/03/2019 19:27 VR 31/03/2019 20:21 VR
Time committment 3 4 31/03/2019 19:20 VR 31/03/2019 20:26 VR
Valuing feedback about personal vote 1 1 31/03/2019 20:10 VR 31/03/2019 20:10 VR
VR
31 Items

Appendix I: Screenshot of late stage themes of empirical paper as seen in NVIVO



Appendix J: Ethical approval email for empirical study

From: King, John
Sent: 17 September 2018 14:20
To: Pilling, Steve <s.pilling@ucl.ac.uk>
Subject: Fwd: Ethics Approval CEHP2018569 Pilling

From: King, John
Sent: Thursday, 13 September, 14:19
Subject: Ethics Approval CEHP2018569 Pilling
To: Pilling, Steve, VPRO.Ethics

Dear Steve,

I'm happy to approve your recent ethics application, "Questionnaire-based psychological research."

The approval reference is CEHP2018569 and as requested the expiry is 30/09/2023.

I have attached your documents to this email and copied in UCL REC for archiving. Please keep this email as your record of approval, and please use the amendment process as described on the PaLS Intranet Ethics pages to make any changes to the programme including the addition of new instruments or study samples.

Best Wishes,

John

--
Dr John King
Senior Lecturer, Research Department of Clinical, Educational and Health Psychology & Institute of Cognitive Neuroscience
Division of Psychology and Language Sciences
University College London
1-19 Torrington Place
London WC1E 7HB

Tel: +44 (0)20 7679 5993 (internal 45993)
Email: john.king@ucl.ac.uk
Web: https://iris.ucl.ac.uk/research/personal?upi=JAKIN44
--

Appendix K: Quantitative Questionnaire used in the empirical paper

Please fill out this questionnaire in comparison to other committee meetings you been in where informal consensus discussion has been had. We are interested in finding out whether this method was helpful or not along some dimensions. There might be some elements that are indistinguishable between meetings, which could be captured by a middle number. Please attempt to complete as many items as you can.

Please circle your role on the committee:
Chair / medical / non-medical / lay member / Other- please state: _____

Question	Much Less						Much More
	1	2	3	4	5	6	7
Your experience in the group discussion							
How much did you feel able to participate?							
To what extent did you feel your time was well spent?							
How would you rate your overall contribution?							
The recommendations decided by the group							
How clinically useful do you feel they are?							
How satisfied were you with them?							
Group discussion							
How unproductive would you say the discussion was?							
How focused was the discussion on the clinical question?							
How effectively were disagreements managed?							
Formal consensus method							
How satisfied were you with the consensus method?							
How effectively did the statements capture the important aspects of the review question?							
Did the use of statements worsen your understanding of the evidence?							
How well presented was the evidence?							

Space for additional comments to clarify any of your answers if desired:
_____
_____
_____
_____