Research guidelines for the Delphi survey technique

Felicity Hasson BA MSc
Research Associate, Centre for Nursing Research, University of Ulster

Sinead Keeney BA MRes
Research Associate, Centre for Nursing Research, University of Ulster

and Hugh McKenna RCN RMN DipN(Lond) BSc(Hons) Adv Dip Ed RNT DPhil FRCSI
Head of School of Health Sciences, University of Ulster, Newtownabbey, Northern Ireland

Accepted for publication 17 April 2000

INTRODUCTION

The ability to make effective decisions in situations where there is contradictory or insufficient information has led to an increased use of consensus methods, namely brainstorming, nominal group technique and the ‘Delphi’ survey technique (hereafter referred to as the Delphi).

Originally developed by the Rand Corporation for technological forecasting this technique was named after the famous oracle at Delphi. The approach has commonly been adopted in medical, nursing and health services research (Williams & Webb 1994, Kirk et al. 1996, Gibson 1998) and many differing forms are now in existence. These include, the ‘modified Delphi’ (McKenna 1994), the ‘policy Delphi’ (Crisp et al. 1997), and the ‘real-time Delphi’ (Beretta 1996). Due to the flexibility of this technique, the literature is replete with studies reporting its use but modified forms of the Delphi have been

Consensus methods such as the Delphi survey technique are being employed to help enhance effective decision-making in health and social care. The Delphi survey is a group facilitation technique, which is an iterative multistage process, designed to transform opinion into group consensus. It is a flexible approach, that is used commonly within the health and social sciences, yet little guidance exists to help researchers undertake this method of data collection. This paper aims to provide an understanding of the preparation, action steps and difficulties that are inherent within the Delphi. Used systematically and rigorously, the Delphi can contribute significantly to broadening knowledge within the nursing profession. However, careful thought must be given before using the method; there are key issues surrounding problem identification, researcher skills and data presentation that must be addressed. The paper does not claim to be definitive; it purports to act as a guide for those researchers who wish to exploit the Delphi methodology.

Keywords: consensus methods, Delphi survey technique, quasi-anonymity, sampling, data analysis, collection, resources, reliability, validity, interpretation, health services research, nursing
criticized for the lack of methodological rigour as far back as the mid-1970s (Sackman 1975). Regardless of what format is employed, the appropriate use of this technique requires a high degree of methodological precision (Peiro Moreno & Argelaguet 1993) and research rigour.

The authors have used the Delphi in a variety of research studies and have faced a number of dilemmas in its application, administration and reporting. This has resulted in methodological difficulties where judgements had to be made in the absence of good guidance; this has obvious implications for the trustworthiness of the findings. Such guidance is readily available for other data collection methods such as general surveys, focus groups and interviewing. Indeed, an extensive review of the Delphi literature identified that to date no universal guidelines exist. This paper aims to provide an understanding of the preparation and action steps to be taken when using the Delphi. Discussion will also focus on familiarizing the health professional with the basic skills needed to use the Delphi and provide examples to illustrate the difficulties inherent in its use.

ISSUES INVOLVED IN USING THE DELPHI TECHNIQUE

Problem identification

Problem classification seems somewhat obvious as all research begins with the identification of the research problem; here however, the researchers may be given a topic to research, in which case the decision is already taken for them. Nevertheless, in many cases researchers may decide upon a topic themselves. Consideration must be given to an array of factors including the resources available and the researchers’ competency and skills. It is at this initial stage that the circumstance or problem linked to group communication determines the use of the Delphi and not the nature of the technique itself (Linstone & Turoff 1975). For example, using the Delphi to identify health and social care professionals’ views on smoking cessation is more practical than using the technique to collect descriptive information on smoking cessation programmes. Therefore, decisions that lend themselves to the use of group involvement can be considered appropriate. Turoff (1970) outlined four research objectives that called for the use of the Delphi:

- to explore or expose underlying assumptions or information leading to differing judgments;
- to seek out information which may generate a consensus on the part of the respondent group;
- to correlate informed judgments on a topic spanning a wide range of disciplines; and
- to educate the respondent group as to the diverse and interrelated aspects of the topic.

The Delphi is only appropriate to investigate certain research problems, so careful consideration must be given to the nature of the problem before selecting this approach. Understanding the nature of the problem and the logistical considerations that arise from the topic need to be established before deciding upon its use. Attention must be given to other data collection methods, such as the postal questionnaire or interview schedule. Reid (1988) asserts that the decision by any researcher to employ this technique centres upon the appropriateness of the available alternatives.

Understanding the process

The Delphi is a group facilitation technique that seeks to obtain consensus on the opinions of ‘experts’ through a
series of structured questionnaires (commonly referred to as rounds). The questionnaires are completed anonymously by these ‘experts’ (commonly referred to as the panellists, participants or respondents). The term participants will be used in this paper. As a part of the process, the responses from each questionnaire are fed back in summarized form to the participants.

The Delphi is therefore an iterative multistage process designed to combine opinion into group consensus (McKenna 1994, Lynn et al. 1998). The initial questionnaire may also collect qualitative comments, which are fed back to the participants in a quantitative form through a second questionnaire. Alternatively, qualitative data can be collected through focus groups or interviews and used to inform a quantitative first round of the Delphi. As with all good surveys, pilot testing with a small group of individuals should precede implementation. After statistical analysis regarding group collective opinion, the results from the second questionnaire help in the formulation of the third quantitative questionnaire. This process is ongoing until consensus is obtained or the law of diminishing returns sets in. That is, responses are summarized between rounds and communicated back to the participants through a process of controlled feedback, this process is repeated until consensus is reached or until the number of returns for each round decreases.

The process gathers opinion without the need to bring panellists together physically. By using successive questionnaires, opinions are considered in a non-adversarial manner, with the current status of the groups’ collective opinion being repeatedly fed back. This informs the group members of the current status of their collective opinion (Goodman 1987) and helps to identify items that participants may have missed or thought unimportant. Therefore, the opportunity exists for participants to change their opinions (Couper 1984, McKenna 1994).

Finding a sample

Studies employing the Delphi make use of individuals who have knowledge of the topic being investigated, which McKenna (1994 p. 1221) defines as ‘a panel of informed individuals’; hence the title ‘experts’ being applied. For example, a study investigating the role of the health visitor would include health visitors who are knowledgeable about the subject under consideration (Lemmer 1998).

Controversial debate rages over the use of the term ‘expert’ and how to identify adequately a professional as an expert. The claim that one group represents valid expert opinion has been criticized as scientifically untenable and overstated (Strauss & Zeigler 1975). The commitment of participants to complete the Delphi process is often related to their interest and involvement with the question being examined. Therefore, a fine balance must be struck in selecting experts who will be relatively impartial so that the information obtained reflects current knowledge and/or perceptions (Goodman 1987), yet also have an interest in the research topic. Moreover, if individuals are to be affected directly by the decision to be made, they are more likely to become involved in the Delphi process. Thus, this technique is exposed both to researcher and to subject bias. As participants know the group’s responses, they may change their views in line with what others are saying. In contrast, this has also been perceived as an advantage of the Delphi in that this is what brings panellists towards group consensus.

The researcher must be aware of when to stop collecting data and what the definition of ‘consensus’ is in relation to the study’s findings (Williams & Webb 1994). If for example, only those opinions that received over 50% agreement in round two were fed back to respondents in round three, this may bias the range of opinions from successive rounds. Outside factors such as limited resources may also influence the level of consensus selected by the researcher.

The number and representativeness of participants will affect the potential for ideas as well as the amount of data to be analysed. To provide representative information, some studies have employed over 60 participants (Alexander & Kroposki 1999) while others have involved as few as 15 participants (Burns 1998). Obviously the larger the sample size, the greater the generation of data, which in turn influences the amount of data analysis to be undertaken. This will lead to issues of data handling and potential analysis difficulties, particularly if employing a qualitative first round approach.

Identification and selection of the sample

Often the selection of the sample of ‘experts’ involves non-probability sampling techniques, either purposive sampling or criterion sampling. Here participants are not selected randomly, so representativeness is not assured. Rather, they are selected for a purpose, to apply their knowledge to a certain problem on the basis of criteria, which are developed from the nature of the problem under investigation. Purposive sampling is based on ‘the assumptions that a researcher’s knowledge about the population can be used to handpick the cases to be included in the sample’ (Polit & Hungler 1997 p. 229). These assumptions are founded on criteria and as Patton (1990 p. 176) explains ‘the logic of criterion sampling is to review and study all cases that meet some predetermined criterion of importance’. Both techniques are similar to selective sampling procedures.

Once the sample is confirmed, the next stage involves the negotiation of access to information regarding the potential participants; and this is a process fraught with difficulties. Gatekeepers may need to be identified to help
pinpoint those individuals who will have knowledge of the topic under study. For example, this could involve a nurse manager consenting to give access or nominating nurses under his/her responsibility. Alongside this, secondary data may have to be accessed to aid identification of the sample. This brings further problems relating to access and authenticity, as databases containing addresses may be obsolete.

Once identified, the potential participants must be approached in an attempt to recruit them to the study. Some researchers target their sample ‘cold’ without any prior notices: this approach may influence the response rate. In his study McKenna (1994) found that it was more advantageous to employ face-to-face interviews in the first round as this helped to increase the response rates in that and subsequent rounds. Although direct contact with the sample is time consuming, the nature of this contact can affect the results obtained. It should be noted, however, that these are criticisms that could be aimed at any qualitative study. However, the Delphi, unlike other methods, requires a continued commitment from participants being questioned about the same topic over and over again, using a slightly modified questionnaire each time. Therefore, as the Delphi is heavily dependent upon the sample having the time to commit to the process, it is also important that those who have agreed to participate, maintain involvement until the process is completed (Buck et al. 1993).

**Informing the sample**

Preparing your sample is an important step, which if not carried out appropriately, could adversely affect response rates in ongoing rounds. When respondents have agreed to participate they need to be informed of exactly what they will be asked to do, how much time they will be expected to contribute and what use will be made of the information they provide. While this can be done verbally for small groups, Whitman (1990) found that alongside verbal instruction, written information accompanying the first round Delphi is effective. If the sample has an understanding of the study’s aims and the process, this helps to build a research relationship, which is important as the ongoing response from the second and third rounds is based on the premise of self-selection. Reminder letters or phone calls may be employed to try to enhance response rates but ultimately as with any research study, the response rate is based on the discretion of the respondent.

Traditionally, the Delphi technique has been paper based, which requires participants to be skilled in written communications. Increasingly, however, the use of electronic communications is employed, requiring participants to be computer literate. Ensuring that your target sample possesses the correct skills needs careful consideration before using this approach.

**Data collection and analysis**

Three issues guide this stage: the discovery of opinions; the process of determining the most important issues; and managing opinions — that is data analysis.

**The discovery of opinions**

First, discovering the opinions raises the question of how many rounds it takes to reach consensus. The number of rounds depends on the amount of time available, whether the researcher has indicated the Delphi sequence with one broad question or with a list of questions, and consideration of levels of sample fatigue. The literature demonstrates that the classic Delphi technique had four rounds (Young & Hogben 1978); however, more recent evidence appears to show that either two or three rounds are preferred (Proctor & Hunt 1994, Beech 1997, Green et al. 1999). Knowing when to stop is crucial — too soon will provide results that may not be meaningful, not soon enough may cause sample fatigue and may tax resources (Schmidt 1997).

Consideration must also be given to the level of consensus to be employed. A universally agreed proportion does not exist for the Delphi, as the level used depends upon sample numbers, aim of the research and resources. McKenna (1994) drawing on Loughlin & Moore’s work (1979) suggests that consensus should be equated with 51% agreement amongst respondents, Sumison (1998) recommends 70%, while Green et al. (1999) opted for an 80%. Alternatively, Crisp et al. (1997) questioned the value of using percentage measures, suggesting that the stability of the response through a series of rounds is a more reliable indicator of consensus.

**Process of determining issues**

The process to determine opinions begins with round one. Within the classical Delphi, round one begins with an open-ended set of questions that generates ideas and allows participants complete freedom in their responses. This helps to identify issues, which would be addressed in subsequent rounds (Gibson 1998). Participants are encouraged to donate as many opinions as possible so as to maximize the chance of covering the most important opinions and issues. However, this can generate large amounts of data, so many researchers have limited the number of opinions a participant can contribute. Schmidt (1997) recommends that participants be asked for at least six opinions, as several participants are likely to raise the same issue using different terms. Careful thought must be given to the initial round one question in order to ensure that it is well phrased and definitive. Some studies (Duffield 1993, Jerkins & Smith 1994) revised this approach. In round one they provide pre-existing information for ranking or response. It must be recognized that
this approach could bias the responses or limit the available options.

As previously stated, round two is made up of the analysis of the results of round one. Similarly, in round three participants are sent the results of the analysis of round two’s responses with statistical information presented to indicate items that have gained collective opinion. This process is ongoing until no further consensus can be reached.

Data analysis

Finally, data analysis involves the analysis and careful management of qualitative and quantitative data. Data from the first round of the Delphi are often qualitative and can be analysed using content analysis techniques; this may involve the use of qualitative software such as Nud*ist or Ethnograph (Pateman 1998). Data collected from this initial stage are analysed by grouping similar items together. Where several different terms are used for what appears to be the same issue, the researcher groups them together in an attempt to provide one universal description. These descriptions and grouping systems need to be verified to ensure that the data are fairly represented. When using a classic Delphi no items should be added during analysis and the wording used by participants, with minor editing, should be used as much as possible in listing items for round two. To help structured debate, this technique can be used in conjunction with informal literature reviews and/or meta-analysis.

Some studies report that infrequently occurring items can be omitted to keep the resulting list manageable (Whitman 1990, Green et al. 1999). However, this goes against the basic tenets of the Delphi technique. Participants themselves should judge items in terms of quality, not the researchers. The iterative building process, central to the Delphi, is difficult enough to advance, without the inappropriate intervention of the research team. Green et al. (1999) noted this when they employed the Delphi in a study of GPs’ (general practitioners) information requirements. They reordered and reduced the data generated in round one, which ultimately moved the statements included in later rounds further from the verbatim responses on which they were originally based. When investigating opinions that are not well established, this produces large amounts of data. To avoid introducing bias, this needs to be managed in terms of feedback to the sample. Therefore, the number of items carried over to subsequent rounds needs careful consideration, as too many can cloud consensus.

Subsequent rounds are analysed to identify convergence and change of respondents’ judgements or opinions. Ascertaining the level of collective opinion often entails the use of descriptive and inferential statistics. For example, round two requires the data from the ratings of the items to be analysed by producing statistical summaries for each item. Central tendencies (means, medians and mode) and levels of dispersion (standard deviation and the inter-quartile range) are computed to provide participants with information about collected opinion. This enables participants to see where their response stands in relation to that of the group. However, the level of measurement will influence the types of statistical tests undertaken, for example the standard deviation does not apply to ordinal or nominal data so returning such information back to participants is misleading.

Resources

The success of the Delphi technique relies upon the administrative skills of the researcher, which should never be underestimated. Devising a coding system to track respondents and their responses from the first to the final round, sending reminders out and analysing change of opinion all have to be undertaken. Further, creating filing systems, formulating mailing labels and establishing a mail base need to be developed and maintained. Attention must be given to the level of resources required; for example, physical storage space, envelopes, paper, labels and so on. These are issues which are rarely touched upon in the literature, but upon which the smooth implementation of the Delphi is based.

Ethical considerations

Within the Delphi, participants do not meet with each other face to face and therefore they can present and react to ideas unbiassed by the identities and pressures of others (Goodman 1987). Reviews of this method claim that anonymity is one of the features which characterize it from other consensus methods (nominal group technique). In order to maintain the rigour of this technique, a response rate of 70% is suggested by Sumsion (1998) for each round: to achieve this the researcher must know the identity of respondents, and non-respondents must be pursued. Therefore, the pursuit of true anonymity presents problems. The term ‘quasi-anonymity’ may be used to indicate that respondents will be known to the researcher and even to one another, but their judgements and opinions remain strictly anonymous (McKenna 1994).

Reliability and validity

When undertaking any research study, consideration must be given to issues of reliability and validity. Reliability is the extent to which a procedure produces similar results under constant conditions on all occasions. There is no evidence of the reliability of the Delphi method: in other words if the same information were given to two or more panels, would the same results be obtained? To overcome this dilemma, Lincoln & Guba’s (1985) criteria for qualit-
ative studies could be applied to help ensure that credible interpretations of the findings are produced. The criteria are based on four major issues, namely credibility (truthfulness), fittingness (applicability), auditability (consistency) and confirmability.

The Delphi is based upon the assumption of safety in numbers (i.e. several people are less likely to arrive at a wrong decision than a single individual). Decisions are then strengthened by reasoned argument in which assumptions are challenged, thus helping to enhance validity. Threats to validity arise principally from pressures for convergence of predictions (Hill & Fowles 1975) which undermines the Delphi's forecasting ability. However, the use of participants who have knowledge and an interest in the topic may help to increase the content validity of the Delphi (Goodman 1987) and the use of successive rounds of the questionnaire helps to increase the concurrent validity. Nonetheless, it has to be stated that the validity of results will be ultimately affected by the response rates.

Reporting results

In Delphi surveys there exists no consistent method for reporting findings (Schmidt 1997) and a review of the literature showed that a number of approaches have been used. These include graphical representation (Malhotra et al. 1994), and the textual presentation of statistical results outlining central tendencies, variance and ranks (Woff et al. 1996, Chocholik et al. 1999). The following diagram attempts to outline those sections that researchers should report upon when using the Delphi. This will help readers to judge the reliability of the method and the results obtained.

Reporting on each round separately illustrates clearly the array of themes generated in round one and gives an indication of the strength of support for each round. The presentations of findings are important and findings from subsequent rounds should be reported in a summarized format to indicate the relative standing of each of the opinions. When reporting statistical tests the reader must be informed of how to interpret the results and how to digest the findings in relation to the emphasis being placed upon them.

Interpretation of findings

The Delphi has been criticized, as it is perceived to force consensus and is weakened by not allowing participants to discuss the issues raised. There is also no opportunity for participants to elaborate on their views (Walker & Selfe 1996, Goodman 1987). Further, there is a danger that greater reliance will be placed on the results than might be warranted. Therefore, it is important to note that the existence of a consensus does not mean that the correct answer, opinion or judgement has been found. Instead, the method and results should be used as a means for structuring group discussion and as a means of raising issues for debate. It merely helps to identify areas that one group of participants or ‘experts’ considers important in relation to that topic. Findings from a Delphi survey help to streamline work, which can be used as an adjunct to meetings, thus allowing the involvement of more individuals and enhancing the reliability and validity of the results. It may therefore be most useful in gathering opinions from large numbers of people and as a ‘heuristic device’ (Fischer 1978) rather than as a means of predicting the future.

CONCLUSION

The Delphi technique is becoming increasingly popular in health and social research. However, modification to the classic version of the Delphi is also increasing and this may cause methodological difficulties (McKenna 1994). Used correctly and rigorously, the Delphi can contribute significantly to broadening knowledge within the nursing profession. Nonetheless, careful thought must be given before using this method as no universal guidelines exist. This paper has aimed to address this gap in the literature.

The conclusions from this paper can be summarized in the form of a checklist of issues, which could be used by researchers undertaking this technique.

Checklist

- Clarify the research problem, remember the Delphi technique is a group facilitation technique and as such only lends itself to group involvement.
- Identify the resources available and skills of the researcher in analysis, administration and relationship building.
- Understand the technique’s process and decide upon which medium to use (electronic or written communication).
- Decide on the structure of the initial round (either qualitative or quantitative) and the number of rounds to employ.
- Determine the criteria and the definition of ‘expert’ and the meaning of ‘consensus’ in relation to the studies aims.
- Give careful thought to the criteria employed, the justification of a participant as an ‘expert’, the use of non-probability sampling techniques, either purpose or criterion methods;
- Give attention to issues which guide data collection: the discovery of opinions, the process of determining the most important issues referring to the design of the initial round, and the management of opinions,
analysis and handling of both qualitative and quantitative data.

- Consider how to present the final results in either graphical and/or statistical representations with an explanation of how the reader should interpret the results, and how to digest the findings in relation to the emphasis being placed upon them.

- Finally, address issues of ethical responsibility, anonymity, reliability, and validity issues in an ongoing manner throughout the data collection process.

This paper does not claim to be definitive; it aims only to act as a guide for those researchers wishing to undertake the Delphi method by providing structure where arbitrary practice is standard practice. If used appropriately, the technique certainly has much to offer in terms of gaining opinions from a wide range of individuals on a specific topic.

References


