Considering the implication of variations within delphi research

Miles Thompson

Family Practice

The definitive version is published in Family Practice
http://dx.doi.org/10.1093/fampra/cmp051

NB: E-mail addresses in the manuscript may no longer be valid. Miles Thompson can be contacted via this web site: www.mvdct.org.uk

Thompson, M. (2009). Considering the implication of variations within delphi research. Family Practice. 26, 420-424.
http://dx.doi.org/10.1093/fampra/cmp051

Considering the implication of variations within Delphi research

Running Title: Variations within Delphi research

Miles Thompson, DClinPsy. Clinical Psychologist *

Bath Centre for Pain Services

Royal National Hospital for Rheumatic Diseases

Bath, UK

*Correspondence to:

Miles Thompson

Bath Centre for Pain Services

RNHRD

Upper Borough Walls

Bath. BA1 1RL. UK.

Tel: +44 1225 465941 Ext 346

Fax: +44 1225 473461

E-mail: miles.thompson@rnhrd.nhs.uk

Word count: including title page and references (3,404)

Considering the implication of variations within Delphi research

Abstract

Background: Delphi research is an increasingly popular method within medical, social and psychological research. It tends to be employed where established theory or knowledge are lacking but where “experts” are thought to hold relevant information. It consists of developing and administering sequential questionnaires that seek to move towards a position of relative consensus. Although the original authors of the technique established a specific method, the literature reveals modifications in the way this is applied. Variations include: 1) restricting the ability of experts to respond to the original question, 2) changing or varying the expert groups used and 3) the point at which the research ends. This paper provides an overview of the technique, explains these variations and their implications before highlighting possible ways forward.

Keywords: Delphi Variations Methodology Research

Introduction

Delphi research is an example of a consensus method used in medical, social and psychological research. Consensus methods, like Delphi research, offer systematic ways of gathering information from individuals deemed to hold relevant knowledge (often known as “experts”). Such methods tend to be employed when ‘published information is inadequate or non-existent’ (1). However, establishing a consensus does not mean that an objective truth has been found; such answers are necessarily relative and tied to the time and contexts of the experts consulted (2).

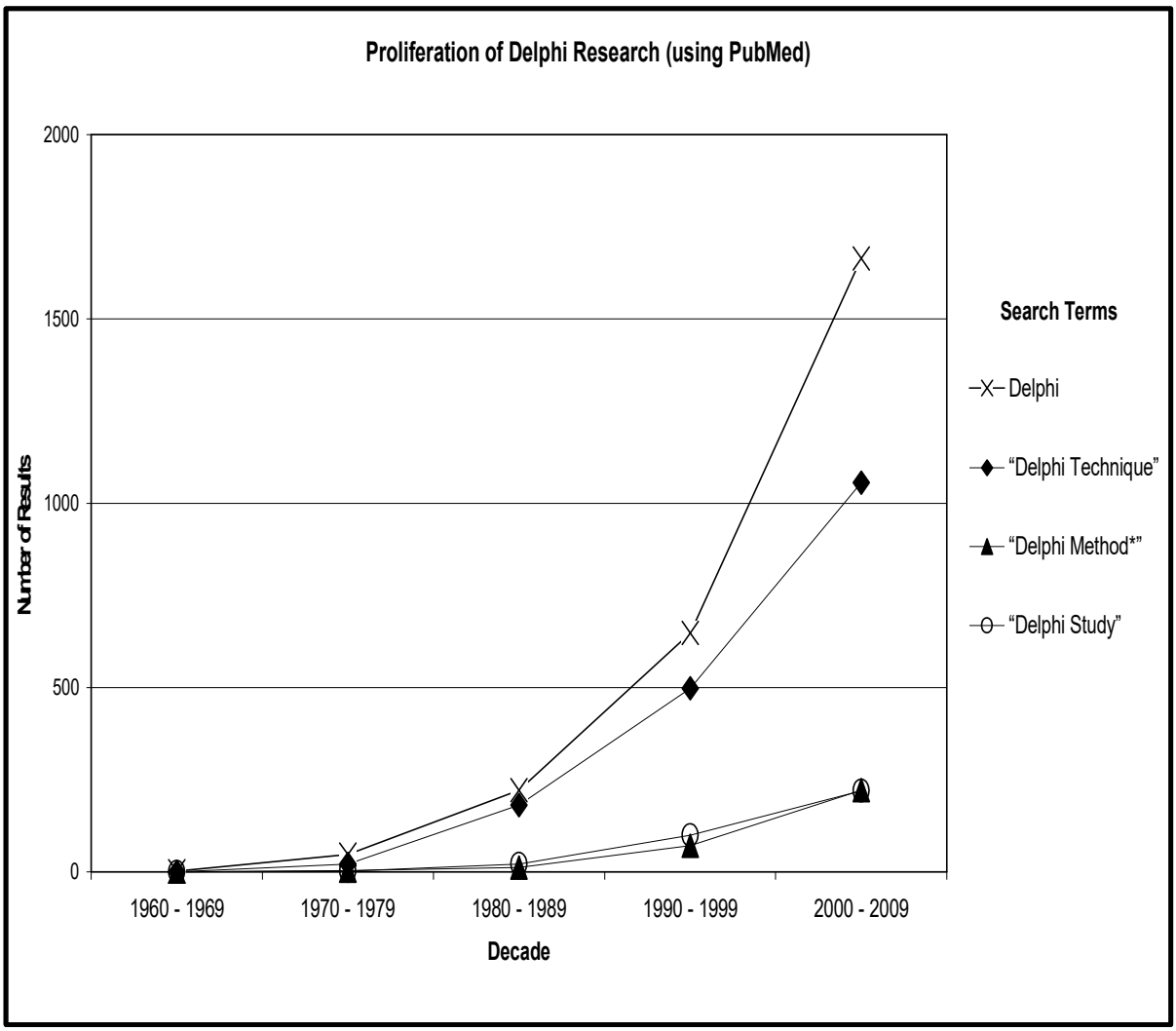
Other consensus methods, such as the nominal group technique (NGT) are also used. The NGT utilises a structured, face-to-face format that gathers information from experts. However experts can be located over a wide area, even world wide, and this can often mean that the

NGT is impractical or too expensive. Delphi research, on the other hand, can more easily and cost effectively embrace a geographically dispersed sample of experts (3). It is also more confidential. In this way each participant is able to make their unique contribution without undue hindrance or influence from others (4-7).

Delphi research

Delphi research was devised by Dalkey and Helmer (8). It involves the development and administration of a series of sequential questionnaires that seek to move towards a position of relative consensus. The first phase (D1) allows each expert to independently contribute any information they deem pertinent to the research question. Open questions are often used to gather wide ranging responses. Following the collection of this data, 'most researchers summarize, edit, categorize, and eliminate redundant answers' (4). In the second phase (D2), the summarised information is fed back to the original expert group in a reduced questionnaire format. Here, experts are asked to rate the edited statements from D1 (6), often in terms of agreement, usefulness or relevance. Researchers then collect the data, typically in the form of Likert scores, and statistics such as median values or inter-quartile ranges are calculated for each item (9). In traditional Delphi research the questionnaire statements are then fed back to the expert group to be re-rated (D3). This time with median (or other) values displayed alongside each item. This stage can be repeated (D4, 5, 6…) until researchers believe consensus has been reached. Each successive round provides experts with the opportunity to modify their judgements in line with new average scores (5). Researchers claim that data gathered using this methodology is superior to opinions gathered from individuals (6, 10).

Increasing Popularity and Variance

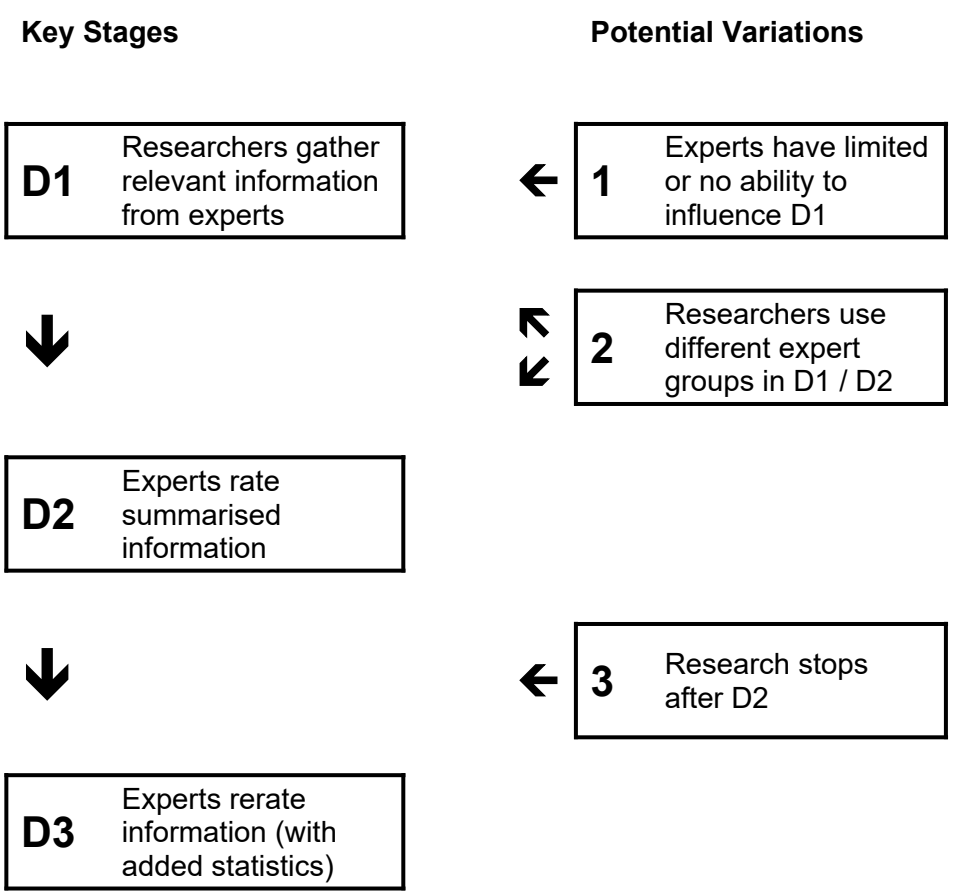


The graph above suggests that Delphi research has increased in popularity since its invention in the 1960's. Searches for the graph were conducted on the 7th of June 2009 (accordingly the graph utilises incomplete data for this year). The four terms were entered into the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed/). No restrictions were placed on where the term could be found in order to maximise search results. The phrase "Delphi Method*" was included with a wildcard to capture both the term "method" and "methodology". Terms which contained more than one word were searched for within quotation marks to both minimise search results and to increase the likelihood that the article was referring to the research methodology discussed in this article and not the word "Delphi" in some other context. The graph demonstrates an increasing number of results for Delphi research related search terms over recent decades, particularly since the turn of this century.

Delphi research has been used recently in both primary care (11, 12) and this journal (13, 14).

Whilst the proliferation of a research method is often something to be lauded, difficulties can arise if the technique is used inconsistently. A brief examination of some the research indicated from the results above quickly reveals marked methodological differences. While many studies continue to employ the exact methodology described earlier, others significantly deviate. Although authors comment that Delphi research can be used in a variety of different ways (15), do these methodological variations impact the integrity of the method? This paper will explore three variations that occur in the literature: 1) restricting the ability of experts to respond to the original question, 2) changing or varying the expert groups and 3) the end point of the research. It will use examples of research from the literature to illustrate it points. Please note, this paper is not suggesting that these variations are the only ones that occur in the literature: other significant deviations may exist. Indeed researchers and consumers of research should be alert to other variations and the implications of these.

Illustration 1. Key Stages and Potential Variations in Delphi Research



1. Restricting the ability of experts to respond to the original question

Example 1: Neimeyer and Diamond investigated historical and anticipated future directions of counselling psychology (16). The authors developed a questionnaire through reviewing the literature themselves (D1). When completed, it was sent to the directors of 72 Institutional Members of the Council of Counselling Psychology Training Programmes. Directors were asked to rate how likely each of the 28 items were to occur over the course of the next 10

years (D2). The results were collected, collated and then the same experts re-rated the items (D3). The authors then discussed the relative importance given to statements.

In example 1, Neimeyer and Diamond began their study by reviewing the literature and producing a questionnaire themselves (D1). By adopting this procedure, the researchers impose a rigid starting point for their participants. In this way, expert thinking is severely constrained by the confines of the pre-designed questionnaire. Indeed, it appears that the only influence experts can have is over the final order in which the items from the original literature review appear. As such, one might wonder why the authors needed recourse to the experts at all. The research detailed above might be better described as a rating exercise. Although this might be considered an acceptable research methodology, it should be labelled as is, and not as Delphi Research. The implication of situations such as this is discussed in the “ways forward” section of this paper.

Other Delphi research also utilises methods other than open-ended questions to begin their studies, including the use of statements that have been designed by the research authors (17), lists that have been based upon literature reviews and clinical experience (18), lists that have been designed in pilot groups (19), focus groups (20) or lists generated in previous studies (21).

Researchers note that Delphi research methodology becomes weaker if too restrictive a process is imposed on the participants (22). Although authors also suggest ways to mediate this (22) there still remains a risk that when experts are presented with strict and restrictive criteria at the beginning of a study their ability to freely express their opinions is reduced. Limitations of this kind potentially restrict one of the core ideas behind Delphi research.

2. Changing or varying the expert group

Example 2: Kennedy and Llewelyn (23) set out to establish expert's perception of the future direction of UK clinical psychology training. The authors sent open-ended questionnaires to key staff in the 26 UK clinical psychology training programs (D1). Participants were asked to indicate what they thought would be the most likely elements of training in the first decade of the 21st century. The authors reduced these predictions to 151 representative items and sent them to three different groups (D2). These individuals where asked to rate the statements on a three point scale according to likelihood that the issue would be relevant to training. The authors discussed the predictions that were rated as being most likely and most unlikely using median data.

The above research begins by consulting one group of experts (D1), before getting three different groups of individuals to rate the information (D2). Other Delphi research shows similar departures between D1 and D2. Graham and Milne explicitly use two different groups, one known as 'the Delphi group' (D1), the other known as 'respondents' (D2) (24). The literature reveals that this is not unusual. For example, the results of D1 have been simultaneously considered by: 3 different groups each containing different populations - teachers, parents and health professionals (25), 5 different groups of people from the same geographical area (21), 7 groups of the same profession (psychiatrists) from different countries (19).

Traditional Delphi research would use the same expert pool in both D1 and D2. It is unclear whether using a different expert group or even a non-expert group for D2 could be detrimental to the results of the research. However, any instance where information is being rated by individuals other than the original experts may well detract from the validity of responses. At the very least, if D2 no longer involves the original expert group the conclusions of the study can not be said to derive from them. It is of course possible and sometimes useful to get other groups to rate the data, but this does not preclude the experts involved in D1 also rating it.

3. The end point

The literature suggests that many examples of Delphi research reach their end point before that laid out in the exemplar methodology described above. For example, many studies stop after D2 (see example 2). In these instances experts only rate the statements once (if at all) and although mean or median data about how the statements are initially rated is often used in research reports this data is often not re-presented to expert or other groups alongside the statements for re-rating in further rounds (D3, 4, 5…).

In some ways it is unclear why experts should need to repeatedly re-rate the research statements unless researchers feel a point of total agreement needs to be reached. Taking such a stance might suggest that a position of absolute truth is being sought (or is even possible) when, in reality, Delphi research is most likely to produce a time limited snapshot of expert opinion with areas of agreement, overlap and divergence seems. Indeed, it could be argued that any change in expert opinion which does take place between D2 and D3 (or 4, 5, 6…) would be a result of experts being influenced by the opinions of others. This might seem odd when the lack of influence of this kind is one of the unique selling points of Delphi research. Of course, there is probably a marked difference between being influenced by the opinions of colleagues within the context of a nominal group technique (when they might be sat opposite you at a table) and being influenced by the anonymous opinions of others whose views you can consider privately and in your own time.

Another reason for legitimately ending Delphi research after D2 is to reduce the task demands placed on experts, which may become arduous if successive rounds take place (4, 22). Indeed, in instances where D3 (4, 5, 6…) are carried out, it is possible that an apparent consensus might actually be a sign of an attritional regression towards the mean (2).

This paper does not suggest that latter rounds of Delphi research should be abandoned, only that they should be balanced against the demands put on the expert group and that

researchers consider and discuss possible reasons for any changes that are recorded. Indeed repeatedly rating items can be particularly useful if Delphi research is being used to develop consensus practice standards or indicators of quality. In these instances, repeated rounds can be used to uncover which statements can be agreed by all (or most) experts (i.e. which can help identify lowest standards) and which can only be agreed by some (i.e. identifying exemplars).

Ways forward

The three areas of divergence outlined above indicate that different research methods co-exist under the overall umbrella title of Delphi research. Worryingly, it seems possible that in certain instances somewhat weaker research may escape scrutiny because it shelters under the protection of this title. Legitimate arguments can, of course, be made for using different techniques – for example for stopping the research after D2 (variation 3), however the implication of imposing rigid starting criteria on Delphi research is more serious (variation 1). In these instances it is possible that the authors have had a much greater influence on the study than the experts on which they call and the limiting influence this places on the experts needs to be highlighted so that future researchers, reviewers and readers are aware of it.

The implication of using different groups in D1 and D2 also warrants discussion (variation 2). There is a significant difference, between only wanting to know what a single group of experts think about an issue, and wanting to know what a different group think about what the original group thought. It would perhaps be useful to label Delphi research accordingly. For example, studies which use the same groups in D1 and D2 could be said to have an internal focus whilst those which use a different group in D2 could be said to have an external focus. Even in this latter case, it may still be useful to get the original experts to rate the material they created.

One final solution would be to phase out the use of the term ‘Delphi research’ in academic papers. In this case, authors would be obliged to transparently describe the methods that they used without hiding behind the apparent strength of the title ‘Delphi research’. It is, of course, possible that such an approach is already taking place, but neither reviewers nor readers will be aware of this, as the work will not refer to Delphi research. Indeed, this is the approach taken by the author of this article who carried out Delphi research, but described the methods used without referring to the term at any point (27).

Perhaps in the past reviewers have assumed that because Delphi research methodology is referred to, this indicates that one, established, method is being followed. The distinctions explored above suggest that this is not so. It is important to understand the pitfalls in assessing all Delphi research studies on a like for like basis, or assuming that all studies follow the same protocol. Instead, the methodology of each Delphi research study needs to be individually examined and understood before any judgement can be made about the claims of its research. Similarly, research papers that claim to use a “modified” Delphi research design need to be clear about how it has been modified so that the readers of the research can assess the impact of these changes accordingly.

Acknowledgements

With sincere thanks to Alex Westcombe and Natalie Hogg who commented on an earlier version of this work and, more recently, for the help of Emma Dunford and Kevin Vowles. The author is also grateful for the helpful comments of the anonymous reviewers.

Captions to Figures

Graph 1. Proliferation of Delphi Research (using PubMed)

Illustration 1. Key Stages and Potential Variations in Delphi Research

References

1. Jones J, Hunter D. Qualitative Research: Consensus methods for medical and health services research. BMJ. 1995;311(7001):376-80.
2. Fish LS, Busby DM. The Delphi method. In: Sprenkle DH, Moon SM, editors. Research methods in family therapy. New York, NY: Guilford Press; 1996. p. 469-82.
3. Jeffery D, Ley A, Bennun I, McLaren S. Delphi survey of opinion on interventions, service principles and service organisation for severe mental illness and substance misuse problems. Journal of Mental Health. 2000 Aug;9(4):371-84.
4. Dawson MD, Brucker PS. The utility of the Delphi method in MFT research. American Journal of Family Therapy. 2001 Mar-Apr;29(2):125-40.
5. Delbecq A, Van de Ven A, Gustafson D. Group Techniques for Program Planning: a Guide to Nominal Group and Delphi Processes. New York: Foresman;1975.
6. Linstone H, Turoff M. Delphi Method. Addison Wesley Longman Publishing Co; 1975.
7. Melpignano M, Collins ME. Infusing youth development principles in child welfare practice: Use of a Delphi survey to inform training. Child & Youth Care Forum. 2003 Jun;32(3):159-73.
8. Dalkey N, Helmer O. An Experimental Application of the Delphi Method to the Use of Experts. Management Science. 1963;9(3):458-67.
9. Jenkins DA, Smith TE. Applying Delphi methodology in family therapy research. Contemporary Family Therapy: An International Journal. 1994 Oct;16(5):411-30.
10. Ascher W. Forecasting. Baltimore: Johns Hopkins University Press; 1978.
11. Gensichen J, Vollmar HC, Sönnichsen A, Waldmann UM, Sandars J. E-learning for education in primary healthcare - turning the hype into reality: a Delphi study. European Journal of General Practice. 2009;15(1):11-4.
12. Burke S, Martyn M, Stone A, Bennett C, Thomas H, Farndon P. Developing a curriculum statement based on clinical practice: genetics in primary care. British Journal of General Practice. 2009 Feb;59(559):99-103.
13. Cronin E, Campbell S, Ashworth M, Hann M, Blashki G, Murray J, Tylee A. A tale of two systems: perceptions of primary care for depression in London and Melbourne. Family Practice. 2009 Jun;26(3):210-20.
14. Engels Y, Campbell S, Dautzenberg M, van den Hombergh P, Brinkmann H, Szécsényi J, Falcoff H, Seuntjens L, Kuenzi B, Grol R; EPA Working Party. Developing a framework of, and quality indicators for, general practice management in Europe. Family Practice. 2005 Apr;22(2):215-22.
15. Clayton MJ. Delphi: A technique to harness expert opinion for critical decision-making tasks in education. Educational Psychology. 1997 Dec;17(4):373-86.
16. Neimeyer GJ, Diamond AK. The anticipated future of counselling psychology in the United States: A Delphi poll. Counselling Psychology Quarterly. 2001 Mar;14(1):49-65.
17. Kim J, Lee Y-s, Suen H, Lee GS. A Delphi Study of Young Children's Readiness in Korea: Challenges and Implications for Early Childhood Schooling. Educational Research and Evaluation. 2003 Dec;9(4):345-55.
18. Turner GH, Weiner DK. Essential components of a medical student curriculum on chronic pain management in older adults: Results of a modified Delphi process. Pain Medicine. 2002 Sep;3(3):240-52.
19. Ferri C, Chisholm D, Van Ommeren M, Prince M. Resource utilisation for neuropsychiatric disorders in developing countries: A multinational Delphi consensus study. Social Psychiatry and Psychiatric Epidemiology. 2004 Mar;39(3):218-27.
20. O'Brien AP, O'Brien AJ, Hardy DJ, Morrison-Ngatai E, Gaskin CJ, Boddy JM, et al. The New Zealand development and trial of mental health nursing clinical indicators- A bicultural study. International Journal of Nursing Studies. 2003 Nov;40(8):855-61.
21. Schopper D, Ammon C, Ronchi A, Rougemont A. When providers and community leaders define health priorities: The results of a Delphi survey in the canton of Geneva. Social Science & Medicine. 2000 Aug;51(3):335-42.
22. Wilson A, Averis A, Walsh K. The influences on and experiences of becoming nurse entrepreneurs: A Delphi study. International Journal of Nursing Practice. 2003;9(4):236-45.

23. Kennedy P, Llewelyn S. Does the future belong to the scientist practitioner. The Psychologist. 2001;2:74-8.
24. Graham LF, Milne DL. Developing basic training programmes: A case study illustration using the Delphi method in clinical psychology. Clinical Psychology & Psychotherapy. 2003 Jan-Feb;10(1):55-63.
25. Meuleners L, Binns C, Lee AH, Lower A. Perceptions of the quality of life for the adolescent with a chronic illness by teachers, parents and health professionals: A Delphi study. Child: Care, Health and Development. 2002 Sep;28(5):341-9.
26. Kessler MR, Nelson BS, Jurich AP, White MB. Clinical decision-making strategies of marriage and family therapists in the treatment of adult childhood sexual abuse survivors. American Journal of Family Therapy. 2004 Jan-Feb;32(1):1-10.
27. Thompson M. Exploring the trainees' view of a socio-political approach within UK clinical psychology. Journal of Community & Applied Social Psychology. 2007;17(1):67-83.