Continuing Medical Education

Classic Delphi Method

Pushpa L Jayawardana

No affiliation (Former Professor, Department of Public Health, Faculty of Medicine, University of Kelaniya, Sri Lanka)

*Correspondence: pushpaljayawardana@gmail.com   https://orcid.org/0000-0002-5064-1506

Introduction

Classic Delphi Method is one of the techniques used to assess consensus, which is defined as a "generally accepted opinion or decision among a group of people" (1). This technique is based on the assumption that having a group of experts will lead to more valid outcomes than the decisions made by a single expert (2-8). Reaching consensus is important in situations where empirical evidence is scarce or controversial (2-4, 9-10) for forecasting (11-12) and to elicit content validity of research tools (5, 13-14). Modified Delphi (MDT) (15) and Nominal Group (NGT) (16) Techniques are two other methods used for consensus assessment. The main difference of these techniques from Classic Delphi is the presence of structured interaction (face-to-face dialogue) between the panel members.

Delphi is an ancient city of Greece where the shrine/oracle of the Greek God Apollo was located. A priestess from the oracle was possessed by God Apollo whose utterances/decisions were communicated to lay people, through interpretations made by mastery of a group of expert priests (3, 9, 17). The name of the technique originated from this.

Features of Classic Delphi

Delphi Technique is a structured group process

where several rounds of questionnaires developed based on the information of the previous iterations are circulated, which enables the panellists to arrive at a decision (2, 18-19). It has four unique characteristics: 1) anonymity, 2) iterations (repetitions), 3) controlled feedback and 4) statistical group response (4, 7, 9, 12, 14, 20-21). Controlled feedback refers to the procedure where a researcher analyses qualitative comments and statistical measures, summarizes and shares those with the panellists/experts after each round/iteration (4, 22-24).

Delphi process

The activities (9, 11, 21, 23, 25-26) that occur in each round are described below (Figure 1).

Activity A: Determining the aim of the study which is to guide a group towards consensus

Activity B: Collection of information by conducting a thorough literature search on the research problem (26)

Activity C: Identification and recruitment of experts with consent and briefing them on the aim of the study (19, 26)

Activity D: Development of a questionnaire with a series of open-ended questions (close-ended

questionnaires will not generate new ideas) related to the topic (26)

This will help to obtain specific information regarding the research problem, discernible beyond literature survey (21).

Activity E: Describes a three-round Delphi process as an example with the aims of individual rounds (Figure 2)

- **Round one (R1)** – The aim of R1 is to circulate the questionnaire to experts/panellists. They are expected to complete it by brainstorming ideas (which should be prioritized) and indicating the rationale for individual responses against each question (18). Outcome of this is a series of statements (23, 27). Once the questionnaires are returned, the opinions indicated should be analysed, classified into themes and synthesized/created (6, 28). The round one comments/opinions should be collated to develop the structured survey instruments with Likert-type scales which are distributed during subsequent rounds (25). The questions should be designed to clarify specific issues.

- **Round two (R2)** – The aim of R2 is to achieve convergence of opinions. Aggregated feedback and the R2 questionnaire should be circulated among the panellists. They should be requested to evaluate their agreement regarding each statement, by rating/ranking from this round onwards (21, 25, 29). They should be offered the opportunity to change their previous responses based on group judgements (10). With the return of R2 questionnaire, statistical analysis of the responses should commence, details of which should be provided to the panellists.

- **Round three (R3)** – The aim of R3 (as well as any additional rounds) is to determine consensus where panellists are expected to re-rate each item. It is the same as R2 questionnaire, but generally includes a three-point Likert scale with the options of strong,

moderate and minimal endorsements to be ranked with each statement (23). The Round 3 responses are summarized, and if pre-determined consensus/stability level has been reached, the process is discontinued. If not, it should be conducted until the above has been reached (3, 25, 28). This may be delivered either via ordinary post or electronic mail (16, 21, 30).

Activity F: Statistical analysis

This activity involves both qualitative and quantitative analyses (3, 18-19, 21, 25, 27). Qualitative analysis will evaluate the open-ended questions using exploratory, thematic or content-based analytic methods (3, 27, 29), which is scientifically more robust (26). From the second round onwards, the quantitative analysis with statistical feedback should take place (18, 29).

- **Assessing consensus/agreement** – Consensus refers to the extent to which each panellist agrees on the issue discussed, and the agreement among individual panellists. Quantitative analysis is performed to accomplish the desired level of consensus. Most simple and commonly used method is percentage agreement on a given item (29). Convergence of opinion reflected by median is considered a better indicator of consensus rather than a unanimous decision (21, 27).

Other methods considered are descriptive statistics which include central tendency (mean, median and mode) and dispersion (standard deviation and IQR) (3-4, 20). Regarding central tendency, median and mode are preferred (21, 27) because Likert scales consist of ordinal data. Coefficient of variation is another recommended measure (4, 20, 31-32) although not applicable for ordinal data (20). Some studies have suggested triangulation, which is applying more than one statistical measure (e.g., SD and IQR) to determine consensus (4, 12), yet establishing consensus does not mean that the correct answer has been elicited (9, 27). The results of the analysis may

be presented using frequency tables for qualitative data or the latter and other graphical methods such as histograms and frequency polygons for quantitative data (9, 20).

- **Assessing stability** – Stability is the extent of consistency of experts' responses over consecutive iterations on the same item (10, 20, 30-33). Stability is of two types, namely group and individual. When group stability is reached, statistical difference of two separate rounds will not be significant (20, 27). Of the two, individual stability (non-significant difference in the individual panellist's responses to an item in two rounds) is considered more relevant than group stability (20, 33). The focus of Delphi studies is on stability rather than on consensus, for terminating the process (20, 33-34). It is stated therefore, that "consensus is valueless if stability has not been elicited prior to it" (12, 22).

- **Selecting statistical tests -** Although parametric tests namely Pearson Correlation Coefficient (20, 32) and intra-class coefficient (20) have been recommended in the literature to test stability, the more appropriate ones are non-parametric tests (32) applied relevant to the type of data analysed (e.g., Wilcoxon matched-pairs signed rank test for paired

ordinal data) (12, 20)**.** A simple robust measure recommended to determine stability is assessment of the median and its IQR across rounds (12). McNemar test is another measure proposed for paired qualitative data (20, 32). Although Chi-squared Test has been utilized (30-31), it is not appropriate to test individual stability, as what it measures is independence of rounds (12)**.** Kappa coefficient is another test described in the literature (12, 20, 33)**,** however it is applicable only for nominal data (12).

- **Stopping/closing criteria -** There are both statistical and practical criteria described for termination of the Delphi process. For consensus, what is more accepted is convergence than percentage agreement. Though used frequently, consensus could occur before reaching stability. As the latter is statistically more robust, it is considered as a more appropriate closing criterion (22). As mentioned above, assessment of individual than group stability is more pertinent (27). Regarding practical measures, number of iterations decided *a priori* is considered invalid because consensus/stability may not have been reached, when defined number of rounds have been completed (10)**.** A practical measure which is widely accepted is "point of diminishing returns" (no new information is gathered) (3, 25, 28).
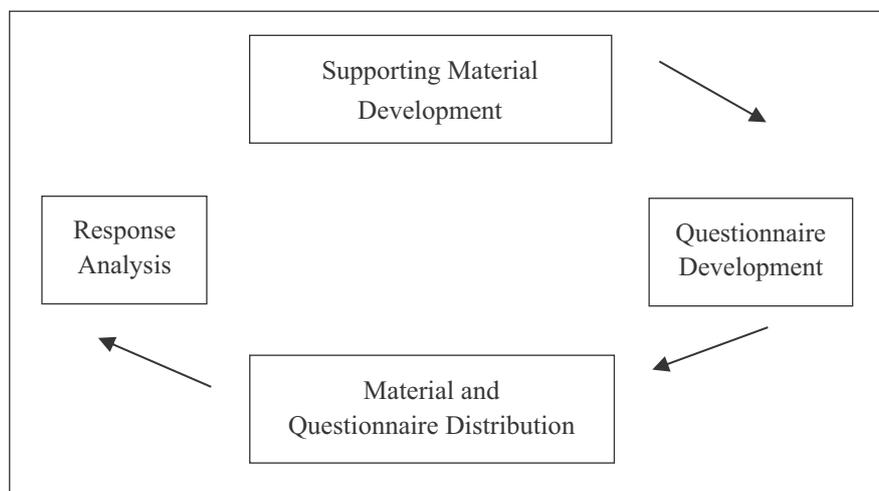

Figure 1: Delphi Round Cycle (9)

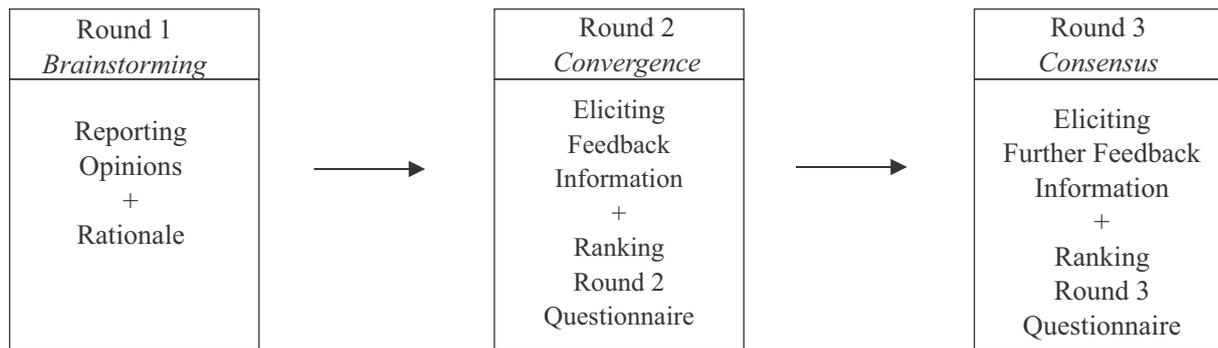| Round 1	Round 2	Round 3
Brainstorming	*Convergence*	*Consensus*
Reporting Opinions + Rationale	Eliciting Feedback Information + Ranking Round 2 Questionnaire	Eliciting Further Feedback Information + Ranking Round 3 Questionnaire

Figure 2: Aims of three-round Delphi Process (9)

- **Validity -** Incorporating several experts as panellists with knowledge is likely to ensure content validity (3, 14). Content validity index is another measure to confirm this (13, 34). A threat to validity is likely to occur if panellists are pressurized to converge regarding judgements (3). However, validity of Delphi technique is debated as the measures used for its assessment do not conform to standard scientific techniques (9, 28).

Activity G: Reporting results

There are no universal guidelines for report writing. It should consist of a detailed account on each of the rounds (11), as described below (Table 1).

Conclusions

Classis Delphi technique is a useful and effective method in determining consensus and stability to help arrive at conclusions regarding unresolved issues despite having a few limitations. It should, however, be conducted adhering to the rules set out to harvest successful and unbiased results.

References

1. Cambridge English Dictionary. Available from: https://dictionary.cambridge.org/dictionary/english/consensus.

2. Niederberger M & Spranger J. Delphi Technique in health sciences: a Map. *Front Public Health* 2020; 8: 457. DOI: 10.3389/fpubh.2020.00457.

3. Hasson F, Keeney S, McKenna H. Research Guidelines for the Delphi Survey Technique. *J Adv Nurs* 2000; 32(4): 1008-1015.

4. Giannarou L & Zervas E. Using Delphi technique to build consensus in practice. *Int J Bus Sci Appl Manag* 2014; 9(2): 65-82.

5. Hasson F & Keeney S. Enhancing rigour in the Delphi technique research. *Technol Forecast Soc Change* 2011; 78: 1695-1704. DOI: 10.1016/j.techfore.2011.04.005.

6. Grime MM & Wright G. *Delphi method.* Wiley Stats Ref: Statistics Reference Online 2016. DOI: 10.1002/9781118445112.stat07879.

7. Dalkey NC. *The Delphi Method: An experimental study of group opinion.* RAND Corporation, Santa Monica, Califonia 90406, 1969.

8. Maceviciute E & Wilson TD. A Delphi investigation into the research needs in Swedish librarianship. *Inf Re* 2009; 14(4): 1-24. Available from: http://informationr.net/ir/14-4/paper419.html.

9. Jones J & Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; 311: 376-380. DOI: 10.1136/bmj.311.7001.376.

10. Barrios M, Guilera G, Nuno L, Gomez-Benito J. Consensus in the Delphi method: what makes a decision change? *Technol Forecast Soc Change* 2021; 163: 1-10. DOI: 10.1016/j.techfore.2020.120484.

11. Davidson P. *Delphi method.* University of Phoenix Research Hub: 1-8. Available from: https://research.phoenix.edu/content/research-methodologygroup/delphi-method.

12. Trevelyan EG & Nicola Robinson N. Delphi methodology in health research: how to do it? *Eur*

Table 1: Suggested Classic Delphi Technique Report

No.	Areas	Information Included
1.	Research problem	Clearly defined
2.	Research topic	Concise
3.	Rationale	Justification
4.	Literature review	Summary of articles reviewed
5.	Methods	Expert selection with details of characteristics
		Number of rounds/iterations – with details of each
		Psychometric properties (validity, reliability/stability)
		Type of statistical analysis
		Ethical responsibilities
6.	Results	Response rate for each round
		Round 1 – total number of issues generated
		Round 2 – report the strength of support with dissensions
		Further rounds – presentation of results
7.	Discussion	Issue of consensus
8.	Conclusions	Interpretation of consensus gained/not gained
		Future research
9.	Appendix	Questionnaires used in each round

Source: (3)

J Integr Med 2015; 7(4): 423-428. DOI: 10.1016 /j.eujim.2015.07.002.

13. Bruce JC & Lack ML. Using subjective judgement to determine the validity of a tutorial performance-evaluation instrument. *Health SA Gesondheid* 2009; 14(1): 1-6. DOI: 10.4102/14i1.409.

14. Goodman CM. The Delphi technique: a critique. *J Adv Nurs* 1987; 12: 729-734.

15. Eubank BH, Mohtadi NG, Lafave MR, Wiley JP, Bois AJ, Boorman RS et al. Using the modified Delphi method to establish clinical consensus for the diagnosis and treatment of patients with rotator cuff pathology. *BMC Med Res Methodol* 2016; 16(56): 1-15. DOI: 10.1186/s12874-016-0165-8.

16. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm* 2016; 38: 655-662. DOI: 10.1007/s11096-016-0257-x.

17. Marchais-Roubelat A & Roubelat F. The Delphi method as a ritual: inquiring the Delphic Oracle. *Technol Forecast Soc Change* 2011; 78: 1491-1499. Available from: https://www.academia.edu/ 5746565/The_Delphi_method_as_a_ritual_Inquiri ng_the_Delphic_Oracle.

18. Keeney S, Hasson F, McKenna H. Consulting the oracle: ten lessons from using the Delphi technique in nursing research. *J Adv Nurs* 2006; 53(2); 205-212.

19. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS ONE* 2011; 6(6): e20476. DOI: 10.1371/journal.pone.0020476.

20. Heiko A. von der Gracht. Consensus measurement in Delphi studies review and implications for future quality assurance. *Technol Forecast Soc Change* 2012: 79(8): 1525-1536. DOI: 10. 1016/j.techfore.2012.04.013.

21. Hsu CC & Sandford BA. The Delphi Technique: making sense of consensus. *Pract Assess Res Evaluation* 2007; 12(10): 1-8. DOI: 10.7275/pdz9-th90.

22. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its

appropriateness. *World J Methodol* 2021; 11(4): 116-129. DOI: 10.5662/wjm. v11.i4.116.

23. Sekayi D & Kennedy A. Qualitative Delphi Method: a four-round process with a worked example. *Qual Rep* 2017; 22(10): 2755-2763. DOI: 10.46743/2160-3715/2017.2974.

24. Dufresne K. *The Delphi Technique.* Available from: https://s4be.cochrane.org/blog/2017/ 11/15/the-delphi-technique/.

25. Ludwig B. Predicting the future: have you considered using the Delphi methodology? *J Ext* 1997; 35(5): 1-3. Available from: https://archives.joe.org/joe/1997october/tt2.php.

26. Iqbal S & Pipon-Young L. The Delphi method. *Psychologist* 2009; 22: 598-601. Available from: https://thepsychologist.bps.org.uk/volume-22/edition-7/delphi-method.

27. 33Keeney S, Hasson F, McKenna H. *The Delphi Technique in Nursing and Health Research.* Chichester, West Sussex, United Kingdom: Wiley-Blackwell, 2011. ISBN: 978-1-405-18754-1. DOI: 10.1002/9781444392029.

28. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines

for use. *Am J Public Health* 1984; 74(9): 979-983. DOI: 10.2105/ajph.74.9.979.

29. Niederberger M, Ko¨berich S & DeWiss Network. Coming to consensus: the Delphi technique. *Eur J Cardiovasc Nurs* 2021; 20: 692-695. DOI: 10.1093/eurjcn/zvab059.

30. Day J & Bobeva M. A generic toolkit for the successful management of Delphi studies. *Electron J Bus Res Methods* 2005; 3(2): 103-116. Available from: www.ejbrm.com.

31. Dajani JS, Sincoff MZ, Talley WK. Stability and agreement criteria for the termination of Delphi studies. *Technol Forecast Soc Change* 1979; 13: 83-90.

32. Yu-Nu Yang. Methodology for testing the stability of experts' opinions between successive rounds of Delphi studies. *ERIC.* ED47216. Available from: https://files.eric.ed.gov/fulltext/ED472166.pdf.

33. Chaffin WW & Talley WK. Individual stability in Delphi studies. *Technol Forecast Soc Change* 1980: 16: 67-73. DOI: 10.1016/0040-1625(80)90074-8.

34. Yusoff MSB. ABC of content validation and content validity index calculation.*Med Educ* 2019; 11(2): 49-54. DOI: 10.21315/eimj2019.11.2.6.