SURVEYS · TYPES AND APPLICATIONS

Delphi Surveys: Building Expert Consensus

The Delphi method surveys a panel of experts in several rounds, feeding back the group's answers between rounds so that judgements can converge. This article explains how Delphi studies are run, how consensus is defined, what the evidence says about their accuracy and how to avoid common weaknesses.

ModQu research article · 1,632 words · 18 references (APA 7th edition)

Some questions cannot be answered with data, because the data do not yet exist or the question concerns the future. When will a new technology become widely used? Which quality indicators best measure hospital care? Which skills will a profession need in ten years? In such cases, the best available evidence may be the judgement of experts. The **Delphi method** is a structured way of collecting and combining that judgement through a series of questionnaires.

Origins

The method was developed at the RAND Corporation in the 1950s, initially to forecast the effects of technology on warfare. Dalkey and Helmer (1963) published the first account of an experimental application. The name refers to the oracle of ancient Delphi. The method spread in the 1960s and 1970s to technological forecasting, policy analysis and planning (Linstone & Turoff, 1975), and later became widely used in health care, education and management research.

How a Delphi study works

A Delphi study has four defining features (Rowe & Wright, 1999):

- **Anonymity.** Experts answer individually and do not know who gave which answer, so that no one is swayed by the status or personality of others.
- **Iteration.** The questionnaire is repeated over several rounds, usually two to four.
- **Controlled feedback.** Between rounds, each expert sees a summary of the group's answers, such as the median and range of estimates, and often the reasons given by others.
- **Statistical aggregation.** The final result is a summary of the group's judgements, such as the median, rather than an agreement reached in discussion.

Figure 1 shows the sequence of rounds.

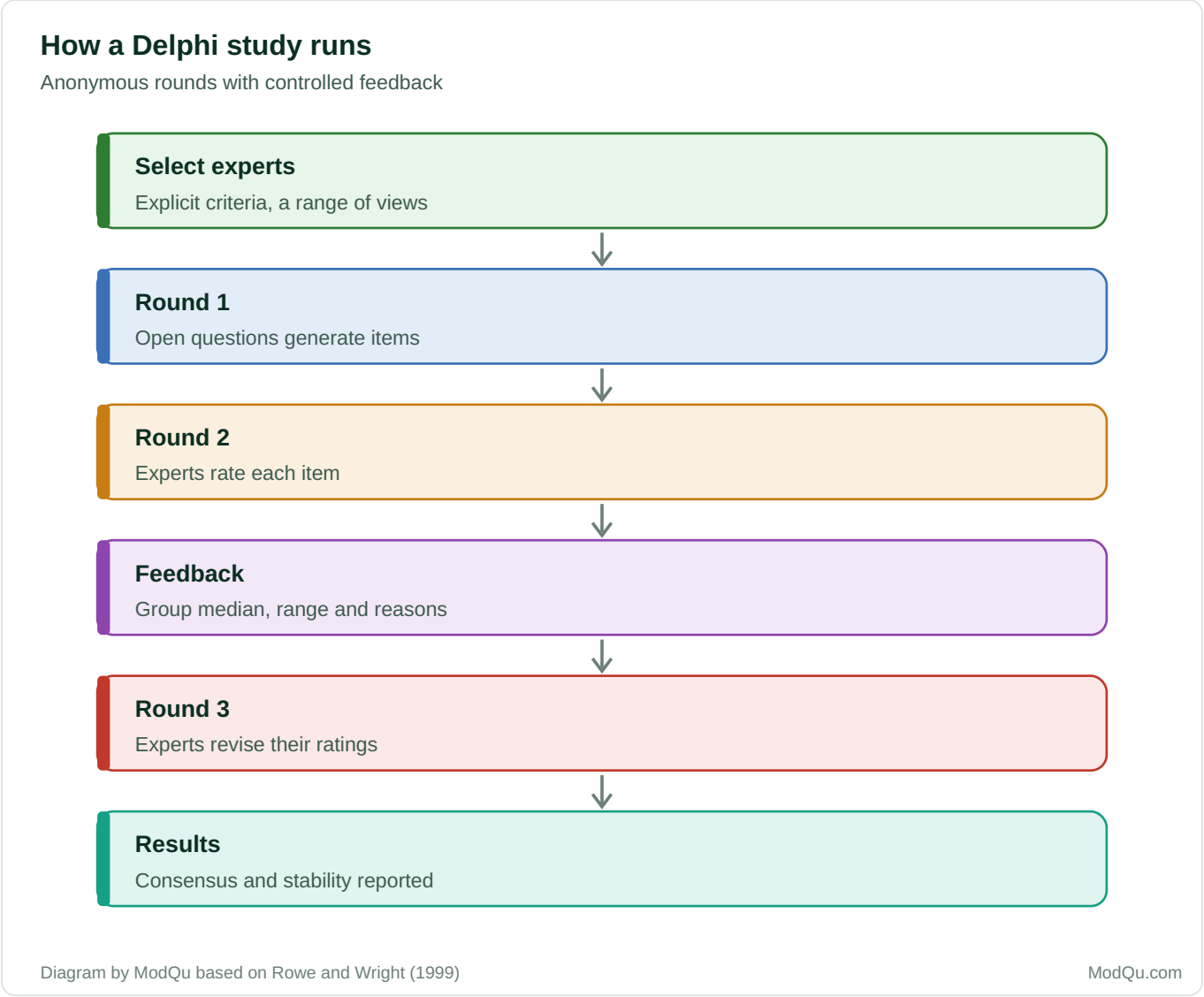


Figure 1. The sequence of rounds in a typical Delphi study. Diagram by ModQu.

The first round is often open: experts are asked to list the issues, factors or items they consider relevant. The researchers combine these into a list, which experts then rate in the second round, for example on a scale of importance or likelihood. In later rounds, experts see the group's ratings and their own previous answer, and may revise their ratings. Experts whose answers lie far from the group's are often asked to explain their reasons, which are shared with the group in the next round.

Selecting the panel

The quality of a Delphi study depends above all on its experts. Unlike a sample survey, a Delphi panel is not meant to represent a population statistically. It is chosen for knowledge and experience, and its composition should be described and justified. Researchers usually set explicit criteria, such as years of experience, publications, professional roles or membership of particular groups, and then identify people who meet them through professional bodies, literature searches or recommendations from other experts.

Diversity matters as much as expertise. A panel drawn entirely from one profession, one country or one school of thought is likely to agree quickly, but its agreement may reflect shared assumptions rather than sound judgement. Many studies therefore include several groups, such as practitioners, researchers, regulators and service users, and report results for each group as well as overall. Including people with direct experience, such as patients in a study of health care priorities, can bring perspectives that professionals miss (Keeney et al., 2011).

Panel size involves a trade-off. Larger panels reduce the influence of any single expert and allow comparisons between groups, but they are harder to manage and keep. Because experts are busy, attrition across rounds is common, and those who drop out may differ from those who stay. Researchers should report how many experts took part in each round.

Why not just hold a meeting?

Committees and meetings are the traditional way of combining expert judgement, but group discussion has well-known problems. Dominant individuals may carry the group, people may defer to those of higher status, and pressure for agreement can suppress dissent, a pattern known as groupthink (Janis, 1982). Delphi aims to keep the benefits of combining several experts' knowledge while avoiding these social pressures. Anonymity allows experts to change their minds without losing face, and feedback allows them to learn from others' reasoning.

Defining consensus

Many Delphi studies aim to reach **consensus**, but there is no single definition of what that means. A systematic review found that studies used many different definitions, most commonly a set percentage of experts agreeing, such as 70 or 75 per cent rating an item as important, and that many studies did not define consensus in advance (Diamond et al., 2014). Other definitions use the spread of answers: for example, consensus is reached when the interquartile range of ratings falls below a set value. Figure 2 shows how the spread of estimates typically narrows over rounds.

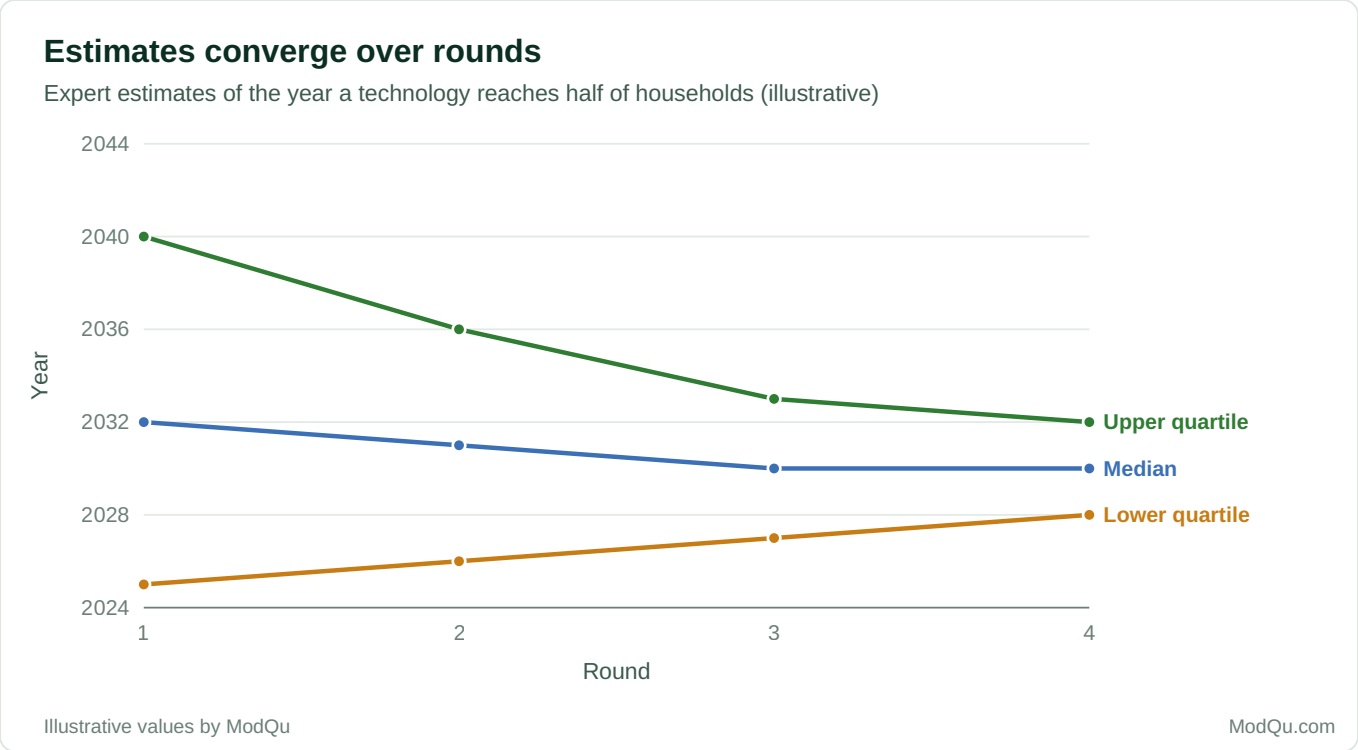


Figure 2. Median and quartiles of expert estimates over four Delphi rounds (illustrative values). Chart by ModQu.

Convergence does not necessarily mean that experts have been persuaded by better arguments. Some may simply move towards the group's view to conform. For this reason, many researchers recommend reporting **stability**, whether answers have stopped changing between rounds, as well as agreement, and treating persistent disagreement as a finding in its own right (Hasson et al., 2000; von der Gracht, 2012).

Accuracy

Does the Delphi method produce better judgements than other methods? Rowe and Wright (1999) reviewed experimental studies and found that Delphi groups were generally more accurate than simple averages of individual judgements made once, and more accurate than unstructured groups. Improvements came from experts with less accurate first-round answers moving towards better answers. However, the evidence came mainly from laboratory studies using students rather than real experts, and early critics argued that the method's claims were not well supported (Sackman, 1975; Woudenberg, 1991). The method's value depends heavily on how it is carried out (Rowe & Wright, 2001).

Designing the questionnaires

Each round's questionnaire must be clear enough for experts to answer quickly and consistently. Items generated in an open first round need careful editing: duplicates must be merged, vague items made specific and each item stated so that it contains a single idea. Rating scales should be labelled clearly, and many studies use a nine-point scale for importance or appropriateness, with the ends and middle labelled. Space for comments allows experts to explain their ratings, and these comments are often the most useful feedback for the next round.

Feedback should be presented neutrally. Showing the median and the range or interquartile range for each item, together with the expert's own previous rating, allows each expert to see where they stand without being told what to think. Summaries of the reasons given for high and low ratings help experts weigh the arguments rather than simply moving towards the middle. Researchers should avoid editing reasons in ways that favour one view.

Limitations

Delphi studies have several limitations. Their results depend on who is selected, and different panels can reach different conclusions on the same question. The feedback process can push experts towards the group view without any new evidence, so that convergence partly reflects conformity. Anonymity reduces social pressure but also removes the chance to resolve misunderstandings through discussion. The method is often used for questions with no correct answer, which makes its accuracy hard to test. Finally, the quality of published Delphi studies varies widely, and many do not report how experts were selected or how consensus was defined (Boulkedid et al., 2011; Diamond et al., 2014).

Variants

- **Policy Delphi** seeks to identify and explore the range of views on a policy question, rather than to reach consensus (Turoff, 1970).
- **The RAND/UCLA appropriateness method** combines a Delphi round with a structured face-to-face meeting, and is widely used to decide when medical procedures are appropriate (Fitch et al., 2001).
- **Real-time Delphi** replaces separate rounds with an online questionnaire in which experts see others' answers as they are submitted and can revise their own at any time (Gordon & Pease, 2006).

Running a good Delphi study

Good practice includes (Okoli & Pawlowski, 2004; Hsu & Sandford, 2007; Keeney et al., 2011):

- **Select experts carefully**, with explicit criteria, and include a range of perspectives. Panels commonly have between 10 and 50 members.
- **Define consensus before the study begins.**
- **Keep questionnaires manageable**, because experts are busy and attrition across rounds is common.
- **Give clear feedback**, including the distribution of answers and reasons for divergent views.
- **Limit the number of rounds** to what is needed; three is common.

- **Report the method fully**, including how experts were chosen, response rates in each round, the definition of consensus and how it was reached. Reporting guidelines have been developed for this purpose (Jünger et al., 2017).

Uses

Delphi studies are used to forecast technological and social change, to set research priorities, to develop clinical guidelines, to select quality indicators and to define competencies and curricula. In health care, they are widely used to choose indicators of the quality of care, although a review found wide variation in how such studies were designed and reported (Boulkedid et al., 2011).

An example

A professional association wants to identify the skills its members will need in ten years. It recruits 30 experts from industry, universities and regulators. In round 1, experts list 64 skills, which the researchers combine into 28. In round 2, experts rate each skill's importance on a scale of 1 to 9. In round 3, they see the group's median and their own previous rating and may revise. Consensus, defined in advance as at least 75 per cent of experts rating a skill 7 or higher, is reached for 12 skills. Twenty-six experts complete all three rounds (illustrative example).

Conclusion

The Delphi method collects expert judgement through anonymous, repeated questionnaires with controlled feedback. It avoids many of the social pressures of meetings and can improve the accuracy of group judgements. Its results depend on careful selection of experts, clear definitions of consensus and full reporting. Used well, it provides a disciplined basis for decisions where data alone cannot settle the question.

References

Boulkedid, R., Abdoul, H., Loustau, M., Sibony, O., & Alberti, C. (2011). Using and reporting the Delphi method for selecting healthcare quality indicators: A systematic review. *PLoS ONE, 6*(6), Article e20476. https://doi.org/10.1371/journal.pone.0020476

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science, 9*(3), 458–467. https://doi.org/10.1287/mnsc.9.3.458

Diamond, I. R., Grant, R. C., Feldman, B. M., Pencharz, P. B., Ling, S. C., Moore, A. M., & Wales, P. W. (2014). Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *Journal of Clinical Epidemiology, 67*(4), 401–409. https://doi.org/10.1016/j.jclinepi.2013.12.002

Fitch, K., Bernstein, S. J., Aguilar, M. D., Burnand, B., LaCalle, J. R., Lázaro, P., van het Loo, M., McDonnell, J., Vader, J. P., & Kahan, J. P. (2001). *The RAND/UCLA appropriateness method user's manual*. RAND Corporation.

Gordon, T., & Pease, A. (2006). RT Delphi: An efficient, "round-less" almost real time Delphi method. *Technological Forecasting and Social Change, 73*(4), 321–333. https://doi.org/10.1016/j.techfore.2005.09.005

Hasson, F., Keeney, S., & McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing, 32*(4), 1008–1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Hsu, C.-C., & Sandford, B. A. (2007). The Delphi technique: Making sense of consensus. *Practical Assessment, Research, and Evaluation, 12*, Article 10.

Janis, I. L. (1982). *Groupthink: Psychological studies of policy decisions and fiascoes* (2nd ed.). Houghton Mifflin.

Jünger, S., Payne, S. A., Brine, J., Radbruch, L., & Brearley, S. G. (2017). Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: Recommendations based on a methodological systematic review. *Palliative Medicine, 31*(8), 684–706. https://doi.org/10.1177/0269216317690685

Keeney, S., Hasson, F., & McKenna, H. (2011). *The Delphi technique in nursing and health research*. Wiley-Blackwell. https://doi.org/10.1002/9781444392029

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications*. Addison-Wesley.

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42*(1), 15–29. https://doi.org/10.1016/j.im.2003.11.002

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting, 15*(4), 353–375. https://doi.org/10.1016/S0169-2070(99)00018-7

Rowe, G., & Wright, G. (2001). Expert opinions in forecasting: The role of the Delphi technique. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 125–144). Kluwer Academic. https://doi.org/10.1007/978-0-306-47630-3_7

Sackman, H. (1975). *Delphi critique: Expert opinion, forecasting, and group process*. Lexington Books.

Turoff, M. (1970). The design of a policy Delphi. *Technological Forecasting and Social Change, 2*(2), 149–171. https://doi.org/10.1016/0040-1625(70)90161-7

von der Gracht, H. A. (2012). Consensus measurement in Delphi studies. *Technological Forecasting and Social Change, 79*(8), 1525–1536. https://doi.org/10.1016/j.techfore.2012.04.013

Woudenberg, F. (1991). An evaluation of Delphi. *Technological Forecasting and Social Change*, *40*(2), 131–150. https://doi.org/10.1016/0040-1625(91)90002-W