'More of an art than a science'? The development, design and mechanics of the Delphi Technique

AUTHOR(S)

Sarah Drumm, Catriona Bradley, Frank Moriarty

HANDLE

10779/rcsi.15105411.v1

'More of an art than a science'? The development, design and mechanics of the Delphi Technique.

Sarah Drumm [a], Catriona Bradley [a], Frank Moriarty [b]

[a] Irish Institute of Pharmacy, Royal College of Surgeons in Ireland

[b] School of Pharmacy and Biomolecular Sciences, Royal College of Surgeons in Ireland

Abstract

The Delphi Technique is a group judgement method which is typically used to reach agreement from a group of people with expertise in a particular area. It is an iterative process where panel members complete questionnaires over several rounds, often rating their agreement/disagreement against a statement, with changes made in later rounds based on the feedback received. It has been used widely in pharmacy-related studies relevant to education, research and practice. This paper provides a critical analysis of the various design choices which researchers may consider when planning a Delphi namely the panel of participants, the use of the Likert scale, the effect of feedback, what constitutes consensus and the number of rounds. It also gives an overview of the development and origins of the Delphi, and discusses the advantages and disadvantages of the technique. Advantages include that the Delphi can be conducted with panel members in different geographical locations in their own time, however the technique can therefore take longer to conduct and lacks face-to-face discussion. Patient experts may be less comfortable participating in a relatively complex survey, however the anonymous nature of the process can be more inclusive in allowing participants to feedback candidly. This paper shows the importance of careful planning of the design choices to ensure the reliability and validity of the Delphi.

Keywords: Delphi, Consensus, Group judgement, Expert, Study design.

Introduction

Consensus or group judgement methods are used to obtain opinions from a group of people with expertise in a particular area, and can be used for predicting future patterns, determining priorities, generating ideas and solving problems.[1,2] Such methods of achieving consensus on an issue or problem are most valuable where evidence is equivocal and thus judgement must be relied upon, for example in defining potentially inappropriate prescribing or diagnostic criteria,[3,4] or professional competencies.[5] There are different methods for this process, one of which is the Delphi Technique or Delphi Method (hereafter referred to as the Delphi). The aim is to reach agreement or a convergence of opinion, and the structured process allows for the effective amalgamation of information.[6,7]

Linstone and Turoff characterise the Delphi as 'a method for structuring a group communication process so that the process is effective in allowing a group of individuals, as a whole, to deal with a complex problem' (p 3).[8] Although the Delphi was traditionally used as a tool for forecasting, it has now been embraced for uses beyond predictive purposes, becoming widely used as a tool to aid decision making by gathering expert opinion.[9–12] The most recent published pharmacy-related study that used the Delphi for forecasting (specifically the impact of internet supply of medicines on pharmacy) dates from 2002.[13]

The Delphi is an iterative process involving the completion of questionnaires over several rounds with respondents often indicating their agreement/disagreement against a statement, typically by using a Likert scale with changes made in later rounds based on the feedback received. These statements, items or other inputs are most often based on a review of the literature, a first round consisting of open-ended questions, or both. With the Delphi there is no necessity for participants to meet face-to-face.[2,14] Figure 1 depicts the typical format and features of the Delphi.[8,15,16] Feedback on the results and comments submitted by respondents is usually circulated on an anonymous basis, allowing respondents to consider and potentially revise their stance, and rate the same statements again in the next round.[4,8,14,17,18] In many studies, the respondents are not known to each other which means that there is a double level of anonymity. Jones and Hunter point out that there are two types of agreement in a Delphi – the first is where respondents rate their agreement against the statement, and the second is the level to which they agree with one another.[19]

McMillan et al note that "While consensus methods are commonly used in health services literature, few studies in pharmacy practice use these methods" (p. 655)[2]. Examples of the use of a Delphi

relevant to pharmacy education, research and practice, and design features of each, are outlined in Table 1.[20–27] The Delphi has wide applicability in pharmacy, and more broadly, health services research. The Delphi has been described as "more of an art than a science",[8] and although evidence is equivocal on the optimal approach to the Delphi, a substantial literature exists to inform such design choices. This paper provides an overview of design aspects/choices of the Delphi, the origins of the method, and advantages and disadvantages.
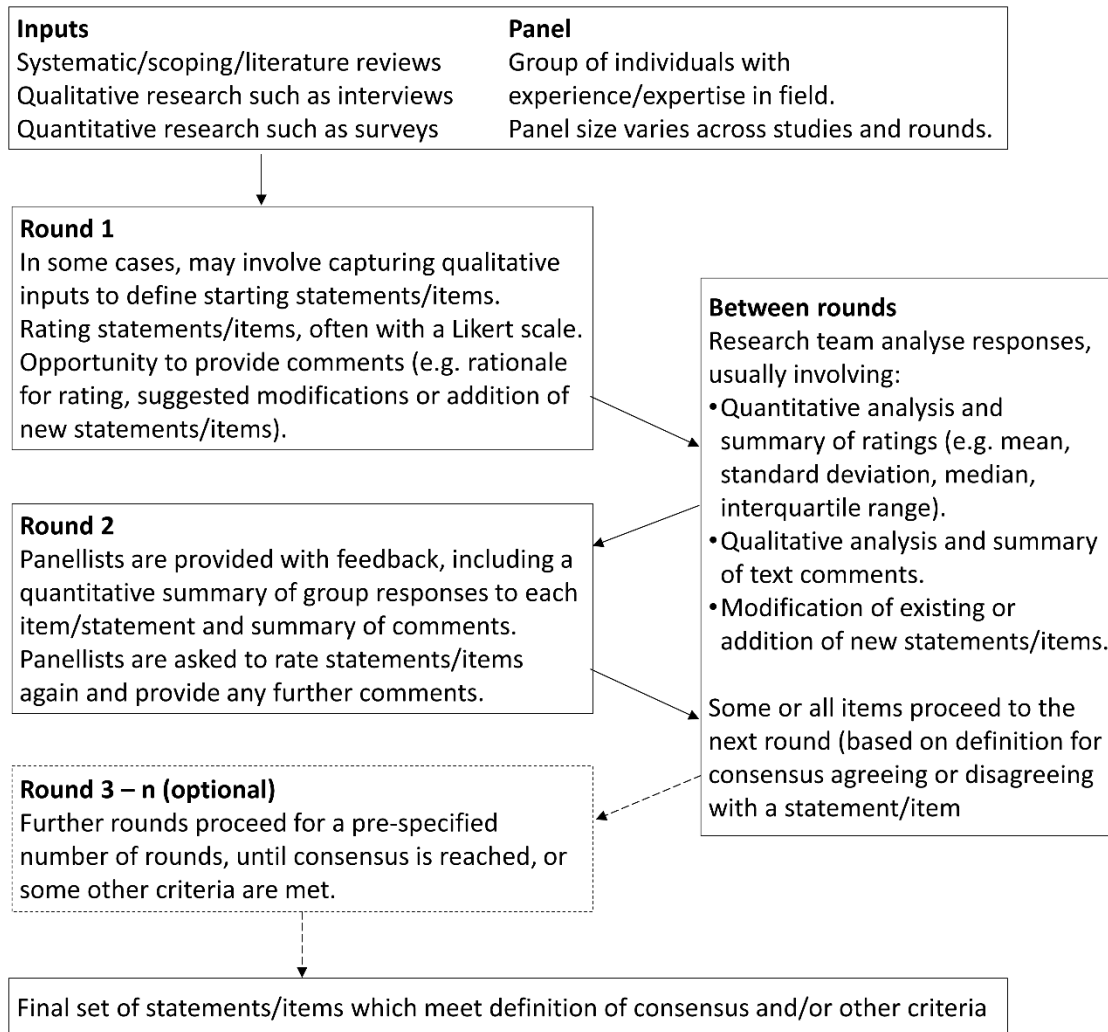
Inputs	Panel
Systematic/scoping/literature reviews	Group of individuals with
Qualitative research such as interviews	experience/expertise in field.
Quantitative research such as surveys	Panel size varies across studies and rounds.

Round 1
In some cases, may involve capturing qualitative inputs to define starting statements/items.
Rating statements/items, often with a Likert scale.
Opportunity to provide comments (e.g. rationale for rating, suggested modifications or addition of new statements/items).

Between rounds
Research team analyse responses, usually involving:
- Quantitative analysis and summary of ratings (e.g. mean, standard deviation, median, interquartile range).
- Qualitative analysis and summary of text comments.
- Modification of existing or addition of new statements/items.

Round 2
Panellists are provided with feedback, including a quantitative summary of group responses to each item/statement and summary of comments.
Panellists are asked to rate statements/items again and provide any further comments.

Some or all items proceed to the next round (based on definition for consensus agreeing or disagreeing with a statement/item

Round 3 – n (optional)
Further rounds proceed for a pre-specified number of rounds, until consensus is reached, or some other criteria are met.

Final set of statements/items which meet definition of consensus and/or other criteria

Figure 1. Typical format and features of the Delphi.

Mechanics of the Delphi

The Panel

Typically, the panel chosen to complete a Delphi comprises those who have experience or expertise in a particular field - sometimes referred to as experts or specialists, or stakeholders who form a

representative response group and who are affected by, or who work in the area.[8,28] Based on those selected, a certain level of knowledge about the field is presupposed and one would imagine that these would be the most appropriate respondents in a study. This may not always be the case however as Linstone and Turoff noted that 'a specialist is not necessarily the best forecaster' as they can focus on the smaller details of something without being able to consider the bigger picture and that in a drive for homogeneity and conformity, there is a risk that the 'tyranny of the majority' can overwhelm a respondent whose views may be in opposition to the group but who in fact may have superior insight (p 566 - 567).[8] Therefore it is important to ensure the composition of the panel reflects the desirable balance of expertise, however as Delphis are typically anonymous, the identity or credentials of respondents do not influence the group judgement. Examples from the pharmacy and prescribing literature illustrate panels being selected based on their expertise in the research topic area, typically including relevant healthcare professionals working in research, educational and/or patient care roles (Table 1). However, patients are also an important stakeholder group that may be included in panels as experts by experience. For example a Delphi to identify core outcomes for trials aiming to improve appropriateness of polypharmacy in older people in primary care included a large number of older adults on its panel.[26]

There is no defined agreement on what constitutes the optimum panel size, with examples ranging from five to more than one thousand respondents.[5,15,18] Loo noted that

> The careful selection of a relatively small panel according to a set of relevant criteria for the
> particular study (i.e. purposive selection) can yield valuable data for management or policy
> decision-making. Large, random samples are not the only option for such purposes (p 767).[29]

The important consideration is that the correct expertise is represented on the panel, rather than just filling it for the sake of reaching an arbitrary number. A systematic review identified a majority of Delphis included ≤25 respondents,[18] while in the Delphi carried out by Dalkey and Helmer seven respondents were used.[30]

Table 1. Examples of recent studies (2016-2020) relating to pharmacy/prescribing education, research, or practice that used the Delphi, and their key characteristics

Study	Topic	Panel	Scale	Consensus	Rounds
Covvey and Ryan 2018[20]	Course objectives in a pharmacy curriculum	87 pharmacy educators in the US who were engaged or interested in global health education	5-point Likert scale from 1 (not at all important) to 5 (extremely important)	At least 85% of the panel selecting Extremely Important or Very Important	3
Drumm et al. 2020[21]	An accreditation framework for continuing education activities for pharmacists	Representatives of seven international accreditation organisations for pharmacist continuing education	5-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree)	At least five of seven panellists selecting Agree or Strongly Agree	4 (with a face-to-face meeting before the final round)
Penm et al. 2019[22]	Definition of medication reconciliation, and essential components of medication reconciliation	24 international experts in the area (leadership in publications, education, professional interest and participation in the area of medication reconciliation)	4-point Likert scale from Strongly Disagree to Strongly Agree	At least 80% of the panel selecting Agree or Strongly Agree	3
Watson et al. 2019[23]	Pharmacists' roles in disasters	15 national and international experts on health aspects of disaster management, advancing practice of pharmacy, and knowledge of pharmacists' roles in disasters	5-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree), switching to a 4-point scale after round 1 with the neutral mid-point option removed	At least 80% of the panel selecting Agree or Strongly Agree	3
Gibbins et al. 2017[24]	Strategies for managing non-prescription combination analgesics containing codeine misuse/dependence in a community pharmacy setting	40 experts within the fields of pharmacy and drug misuse and/or dependence in Australia	6-point Likert scale from 1 (Strongly Disagree, or Very Ineffective) to 6 (Strongly Agree or Very Effective)	Median greater than the 4th point on the 6-point Likert scale, a maximum interquartile deviation (interquartile range/2) of 1, and ≥80% of panellists selecting agreement or effectiveness (highest 3 groups on the Likert scale)	3 (the first asking open-ended questions which formed the basis of items included in rounds 2 and 3)
Jebara et al. 2020[25]	A framework for the potential	33 health professionals (Director level)	6-point Likert scale from Strong	At least 70% of the panel selecting Agree or Strongly	2

	development and implementation of pharmacist prescribing in Qatar	involved in prescribing, policymakers, leading administrators, senior educators and regulators, patient safety and quality assurance managers in Qatar.	Disagree to Strongly Agree	Agree and less than 15% selecting Disagree or Strongly Disagree	
Rankin et al. 2018[26]	Core outcomes for trials aiming to improve appropriateness of polypharmacy in older people in primary care	111 international healthcare experts and a public participant panel of 41 older people from Northern Ireland	9-point Likert scale from 1 (limited importance) to 9 (critical importance)	At least 75% of the panel rating an outcome as critical (7 to 9)	3
Barry et al. 2016[27]	Indicators of prescribing appropriateness in children	18 GPs, pharmacists and paediatricians from the United Kingdom and Ireland	5-point Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree)	At least 75% of the panel selecting Agree or Strongly Agree	2

A search of PubMed for papers including "Delphi" and either "pharmacy" or "prescribing" in their title/abstract over a five-year period 2016-2020 identified 196 papers. Examples were selected to reflect education, research, and practice related studies.

Likert Scale

A Likert scale is commonly used in a Delphi as a tool to measure attitudes and opinions. It is a collection of Likert items, where the respondent is provided with a statement and they indicate on a scale their level of, for example, agreement or disagreement with it.[31–37] It is a quantitative instrument, and characteristics of a classic Likert scale include that it has several items arranged horizontally in an ordinal manner around a neutral middle moving from one end of a continuum to the other with the items approximately equidistant from each other.[31,32,38,39]

The language used in the statement must be coherent, avoiding double-barreled questions, leading questions and double negatives.[31,38,40] The items in the scale typically range from 'strongly disagree' to 'strongly agree' with the respondent selecting a corresponding number e.g. strongly disagree = 1 and strongly agree = 5.[32,41] Scales can be arranged in the opposite order, with the agreement label at the lower scoring end. This is used particularly when negatively worded statements are used.[41] However, empirical evidence suggests using ascending or descending order does not impact on the validity of responses.[42]

Scale design choices include the number of points included, whether there is a midpoint, and labelling of points. Scales often consist of 4-9 points. An odd number of points is more common, with the central point indicating that the respondent neither disagrees nor agrees. A smaller number of points may be unreliable when a respondent does not fit precisely into a category e.g. 'moderately agree', and they can therefore be inconsistent with their responses if surveyed on multiple occasions.[43,44] A greater number of items can allow respondents to select a response closer to their actual opinion, although including similarly-named items such as 'moderately agree' and 'slightly agree' may make it difficult for them to determine which accurately represents their opinion.

Inclusion of a central point allows for a neutral response for indecisive or ambivalent participants.[12,40,41,45] The 'central tendency bias' describes how respondents prefer to avoid the extreme ends of a scale, preferring instead to remain in a more neutral central position (which is variously referred to as 'neither agree nor disagree', or 'undecided'), encouraging respondents to engage in satisficing.[38,40,46–48] Despite suggestions that omitting the mid-point is advantageous in pushing respondents into a more definitive position,[8] evidence suggests forcing a choice may evoke a negative response when they are truly ambivalent, and skew responses to the negative.[49] Published examples of Likert scales more generally,[49] and within Delphis in the pharmacy field (Table 1), have tended to use odd numbers of points which include a neutral point.

Using labels only for the extreme ends of the scale can reduce the risk of central tendency bias, but this measure can introduce another bias where the respondents choose a label they understand rather than one without a label where they may be unclear on what the item is.[48] A 5-point scale reduces the risk of respondent uncertainty as with the extremes labelled, and the central point representing a neutral stance, there are just 2 points where the respondent is required to infer what the items are. A recent study evaluated how different types of Likert scales (3, 5 and 9 point) can impact the level of consensus in a Delphi study among patients to identify treatment goals, and identified no difference in reliability between.[50] Overall, evidence is equivocal on the optimal scale design, as different approaches will typically reduce risk of some bias(es) at the expense of increased risk of others, and so the best choice is dependent on the context and purpose of the scale.[49] However a proposed framework based on several experimental studies recommends that for opinion measurement a 5 or 7-point scale be used (the latter for respondents with high cognitive and verbal abilities and familiar with questionnaires), and fully labelled where responses are to be summarised directly using means and frequencies.[49]

Effect of feedback

The provision of feedback is another key component. It is often termed 'controlled feedback' as it is the facilitator of the Delphi who determines what feedback will be given to respondents.[14,15,51–54] Feedback can take the form of a numerical summary of ratings, narrative comments including rationale for ratings, or both.[55] The effect of feedback can vary, as respondents have several options when selecting their response depending on where their original response fell in comparison to the group. The first option is to keep their original response which may or may not be within the range of the group's responses. The second is that having viewed the feedback, they disagree with the majority and choose a response at the other end of the scale. The third is that should their initial response be an outlier from the group, the feedback might convince them to choose a response that will move the group closer to achieving consensus.[56] Evidence suggests there is more movement amongst those whose responses were furthest from the mean after the first round, and with participants receiving feedback, scores move towards the mean.[4,57] Rowe et al. observed that the reason that the Delphi 'works' is due to a combination of the iterative process, which allows respondents to reflect on their answers, and the effect of the feedback particularly where a rationale or further information is provided rather than just statistical data.[16]

Consensus

Prior to beginning the Delphi, a decision must be made on what constitutes consensus, e.g. the point when a percentage of the votes fall within a prescribed range.[58] The specifics of this vary from study to study.

There are several ways in which one can analyse data to determine whether consensus has been reached. These include evaluating interval scale statistics, include the mean, median or mode (measures of central tendency), or using standard deviation and interquartile range (measures of dispersion).[47,51,59] Where responses are skewed, the median should be used as outliers can have a disproportionate effect on the mean.[14,47] Nominal scale statistics, i.e. the percentage or frequency of responses,[36] can also be used as an alternative or a compliment. For instance to interpret the data from Likert scales, Likert proposed using a table with percentages and sigma values (standard deviation) with the sigma deviation taken from the mean.[31] Using an average when analysing the results of ordinal Likert scales can be problematic, as it does not account for variability nor provide answers to make actionable decisions, as it would be erroneous to interpret an average result as "Strongly Agree and a half".[46] To display the data, histograms, Likert charts, heat maps or scatterplots can be used.

A systematic review of Delphi studies identified the most common definition for consensus was percent agreement (25 of 100 studies), with 75% being the median threshold to define consensus (range 50-97%),[18] with many examples in the pharmacy literature using similar (Table 1). It has been suggested that the consensus level should be determined by the importance of the topic, so for serious or life or death issues, 100 percent may be required whereas for something relating to establishing a preference, a simple majority (i.e. 51 percent) might be the appropriate level.[60] Several studies propose using the stability of responses over rounds as a more reliable alternative to using percentages, while others caution that stability of responses should not be confused with consensus. [15,47,52,56] The standard deviation and interquartile range as measures of the spread of responses can also be used, with low spread indicating respondents are largely in agreement.[14,17,47,51] Consensus can be defined as responses meeting several conditions, as a single measure may not capture consensus.[59] For example, a Delphi to achieve consensus on management strategies for misuse/dependence of non-prescription combination analgesics containing codeine used three parameters to define consensus, including the span of the interquartile range, the median, and ≥80% of respondents.[24]

Number of rounds

There is no set number of rounds for a Delphi but as Linstone and Turoff observed:

> …in all early forecasting Delphis…a point of diminishing returns is reached after a few rounds. Most commonly, three rounds proved sufficient to attain stability in the responses; further rounds tended to show very little change and excessive repetition was unacceptable to participants (p 223).[8]

The number of rounds for some Delphis can be linked to the determination of consensus for example where rounds are held until consensus is achieved, or where the last rounds carried out show no significant difference.[4,14,47,61] The optimal number of rounds appears to be 2 or 3, with larger numbers inducing participant fatigue.[15,47,61] Retention of respondents over successive rounds is important, and number of rounds and other design choices should be considered to minimise loss to follow up. As mentioned, in some studies, the first round of the Delphi is an exploratory phase using a qualitative methodology e.g. open-ended questions or interviews, before moving to a quantitative phase which solicits feedback on these items. Other Delphis begin directly with the quantitative survey element.[62]

Origins of the Delphi

In July 1962 a memorandum was produced for the Research and Development, or Rand, Corporation, by Norman Dalkey and Olaf Helmer entitled *An Experimental Application of the Delphi Method to the Use of Experts.*[30] The report was a revised version of a classified report, *The Use of Experts for the Estimation of Bombing Requirements* (1951)*,* with the aim of making the report suitable for a wider audience.[63] The original report remains classified so this abridged version is still used today.[64] The report described:

> an experiment in the use of the so-called "DELPHI" [sic] method, which was devised in order to obtain the most reliable opinion consensus of a group of experts by subjecting them to a series of questionnaires in depth interspersed with controlled opinion feedback (p v).[63]

The report described how an 'experiment' which was described to participants as a study, had been carried out 10 years earlier with the aim or goal of obtaining expert opinion on the optimal locations in the US for bombing, by asking experts to imagine themselves to be a Soviet strategic planner, and to consider the number of A-bombs required. The report detailed the process of the Delphi with questions centered on a problem, requiring the respondent to provide a rationale for their answers, in addition to considering what information they required to arrive at a more confident answer. Dalkey and Helmer believed that this 'controlled interaction' avoided the pitfalls of more conventional forms of discussions such as round-tables, which could potentially inhibit independent thought and lead to hasty answers. The feedback element allowed the respondent to consider factors that they may have overlooked previously and correct any misconceptions they may have had.[63] The concept of using opinion, which can be controversial, was something that Dalkey expanded upon proposing that opinion falls between knowledge at one end of the information spectrum as it is verifiable and evidence-based, and speculation at the other end as it not generally based on evidence.[65]

The name Delphi does not appear after the first page of Dalkey and Helmer's report and is capitalised when used. The capitalisation suggests an acronym, although as the project was carried out for the US Air force, that might reflect how air force projects are designated. The name Delphi refers to the oracle at Delphi in Greece where the Pythia, the high priestesses of the Temple of Apollo, acted as a medium for Apollo. The priestesses were, appropriately for studies involving pharmacy, potentially under the influence of ethylene drifting from the intersection of two fault

lines on which the temple was built, and which produces an altered state of consciousness, similar to descriptions of the Pythia while in their mantic sessions.[66,67] They enigmatically translated the prophecies, or oracles, of Apollo for the pilgrims who sought assistance. Much has been written about how the rationale for choosing the name was due to the Oracle's skills at forecasting, interpretation and insight although as noted by Marchais-Roubelat and Roubelat, Dalkey and Helmer, the originators of the Delphi, did not and have not explained the origin or rationale for the name.[9,15,19,68] The name appears to have been first used as a joke at Rand due to the idea of forecasting and Dalkey and Helmer were not very fond of it, particularly as it evoked the image of a priestess disseminating obscure information.[59,69] Dalkey noted that

> In some ways, it [the name] is unfortunate – it connotes something oracular, something smacking a little of the occult – whereas as a matter of fact, precisely the opposite is involved (p 6).[64]

Advantages and Disadvantages

Advantages to using the Delphi include that it can be completed by respondents in different geographical locations in their own time which allows for reflection, and it is cost effective.[8,53] The structured method encourages participation from all group members while avoiding the issue of heterogeneity or of a dominant member or manipulation, as can sometimes be seen with focus groups.[2,8,14,51] Participants in Delphis have commented that it is inclusive, comprehensive, rigorous, systematic and efficient while giving them a personal stake in a project and allowing them to feel involved in the development of the research.[9,17,62]

Greason notes further advantages to the Delphi as 'Using a qualitative Delphi approach to explore ethics and policy in LTC [long-term care] resulted in deeper, more rich findings on ethics, an area seldom empirically, or qualitatively, explored' (p. 2) and as it

> iteratively aims to establish consensus on "ways forward" and produce tangible and applicable outcomes in underexplored areas, it involves up to stage three of the knowledge transition portion of Graham's Knowledge to Action-Ethics cycle, which "represents the process of knowledge creation and its translation into practice and policy" (p. 7).[28]

The use of multiple rounds which allow participants the time and space to present their views, and reflect on their opinions of others as well as their own 'sets a strong foundation for further knowledge translation and mobilization' (p. 9).[28]

The Delphi is not without its critics. Sackman reviewed a number of Delphi studies against the American Psychological Association 'Standards for Educational and Psychological Tests and Manuals' (1966).[70] He commented that the 'conventional Delphi is basically an unreliable and scientifically unvalidated technique in principle and probably in practice' (p vi) noting approximately two hundred potential drawbacks to the Delphi including sampling; the concept of the expert; group versus individual opinion; poor questionnaire design; whether the consensus is authentic; whether the questions, responses and results are meaningful; the presentation of descriptive results; the reliability and validity; whether a systematic review has been performed ahead of time; the dropout rate; the use of central tendency and the analysis and interpretation of findings. The Delphi's use of multiple rounds mean that the process can take longer to complete than other methods, but it also asks the respondents to spend longer participating than other methods do which can lead to attrition.[2,51,62] Some participants in Delphis missed the chance for face-to-face group interaction which can allow for lively debate and discussion.[4,17]

Sackman declared that 'The future is far too important for the human species to be left to fortune tellers using new versions of old crystal balls. It is time for the oracles to move out and for science to move in' (p 73).[70] His claim that the Delphi neglects standard experimental guidelines for scientific methods can be contrasted with the opinion of Linstone and Turoff who note that 'in its design and use Delphi is more of an art than a science' (p 6).[8] In spite of Sackman's 'vehement attack' (p 146),[57] the process is still in use which perhaps indicates that with the correct preparation, many of the issues as described by Sackman can be mitigated. It is worth noting that the review carried out by Sackman took place just 10 years after the publication of the details of the Delphi process and the number, and potentially the quality, of the studies was significantly less than is available today.

The Nominal Group Technique, another consensus method, is one that 'appears to be used more commonly with lay people than the Delphi Technique, although the reason why is not clear. Lay people may feel more comfortable participating in a face-to-face meeting, than in a relatively complex survey.'[2] Additionally the use of the word "expert" when describing the panel of a Delphi could potentially be seen as exclusionary by some who may perceive it as referring to for example qualified health professionals, rather than people with experience in the area under discussion

which may include patients, patient advocates and other "lay" stakeholders. Pharmacy-related Delphi studies have included patients as part of their panel, and may be appropriate for many types of research objectives e.g. identifying core outcomes for trials in a certain clinical area.[26] However, an advantage of the Delphi is its anonymity which may promote inclusion where there is a mixed panel of health professional and patient experts or other form of traditional hierarchy. While anonymity is hugely important in studies allowing participants to feedback candidly, it is particularly important in the health sciences for a number of reasons, including ensuring patient comfort in providing details on their experiences, and to guarantee that no potentially identifying patient information is included in the study.

Conclusion

The nebulous beginnings of the Delphi and the fact that the original report is still classified lends a certain air of intrigue to its development. However since this, the method has been extensively used and refined, with some degree of variation in aspects of its design. This suggests it is widely applicable, and flexible for use for a variety of purposes. The Delphi's geographical neutrality and web-based format has advantages over methods that require face-to-face meetings. Certainly with the Covid-19 situation, techniques such as the Delphi offer a viable and pragmatic approach to attaining consensus.

As discussed in this paper it is important that decisions on the design choices are made at the planning stage to ensure the reliability and validity of the Delphi. Whether or not the Delphis that are undertaken today remain true to that first conceived of by Dalkey and Helmer is another matter, but as it has evolved it has taken a shape that hopefully provides answers clearer than those of the original Oracle at Delphi.

References

1. Fitch K, Bernstein SJ, Burnand B, et al. *RAND/UCLA Appropriateness Method User's Manual*. RAND Corporation; 2001.

2. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharm*. 2016;38(3):655-662. doi:10.1007/s11096-016-0257-x

3. O'Mahony D, O'Sullivan D, Byrne S, O'Connor MN, Ryan C, Gallagher P. STOPP/START criteria for potentially inappropriate prescribing in older people: version 2. *Age Ageing*. 2015;44:213-218. doi:10.1093/ageing/afu145

4. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol*. 2003;56(12):1150-1156. doi:10.1016/S0895-4356(03)00211-7

5. Janke KK, Kelley KA, Sweet B V., Kuba SE. A modified Delphi process to define competencies

for assessment leads supporting a doctor of pharmacy program. *Am J Pharm Educ*. 2016;80(10). doi:10.5688/ajpe8010167

6.   Pope C, Mays N. Qualitative Research: Reaching The parts other methods cannot reach: An introduction to qualitative methods in health and health services research. *BMJ*. 1995. doi:10.1136/bmj.311.6996.42

7.   Di Zio S, Castillo Rosas JD, Lamelza L. Real Time Spatial Delphi: Fast convergence of experts' opinions on the territory. *Technol Forecast Soc Change*. 2017;115:143-154. doi:10.1016/j.techfore.2016.09.029

8.   Linstone HA, Turoff M. *The Delphi Method - Techniques and Applications*. Massachusetts, Reading: Addison Wesley; 1975.

9.   Gupta UG, Clarke RE. Theory and applications of the Delphi technique: a bibliography (1975-1994). *Technol Forecast Soc Change*. 1996;53(2):185-211. doi:10.1016/S0040-1625(96)00094-7

10.   Rowe G, Wright G. The Delphi technique as a forecasting tool: Issues and analysis. *Int J Forecast*. 1999;15(4):353-375. doi:10.1016/S0169-2070(99)00018-7

11.   Landeta J. Current validity of the Delphi method in social sciences. *Technol Forecast Soc Change*. 2006;73(5):467-482. doi:10.1016/j.techfore.2005.09.002

12.   Ecken P, Gnatzy T, von der Gracht HA. Desirability bias in foresight: Consequences for decision quality based on Delphi results. *Technol Forecast Soc Change*. 2011;78(9):1654-1670. doi:10.1016/j.techfore.2011.05.006

13.   Holmes ER, Tipton DJ, Desselle SP. The Impact of the Internet on Community Pharmacy Practice: A Comparison of a Delphi Panel's Forecast with Emerging Trends. *Health Mark Q*. 2002;20(2):3-29. doi:10.1300/J026v20n02_02

14.   von der Gracht HA. 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

15.   Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs*. 2000;32(4):1008-1015. doi:10.1046/j.1365-2648.2000.t01-1-01567.x

16.   Rowe G, Wright G, McColl A. Judgment change during Delphi-like procedures: The role of majority influence, expertise, and confidence. *Technol Forecast Soc Change*. 2005;72(4):377-399. doi:10.1016/j.techfore.2004.03.004

17.   Hanafin S, Brooks A. *The Delphi Technique: A Methodology to Support the Development of a National Set of Child Well-Being Indicators*. Dublin: National Children's Office; 2004.

18.   Diamond IR, Grant RC, Feldman BM, et al. Defining consensus: A systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol*. 2014;67(4):401-409. doi:10.1016/j.jclinepi.2013.12.002

19.   Jones J, Hunter D. Qualitative Research: Consensus methods for medical and health services research. *BMJ*. 1995;311(7001):376. doi:10.1136/bmj.311.7001.376

20.   Covvey JR, Ryan M. Use of a Modified Delphi Process to Determine Course Objectives for a Model Global Health Course in a Pharmacy Curriculum. *Am J Pharm Educ*. 2018;82(8):6358. doi:10.5688/ajpe6358

21.   Drumm S, Moriarty F, Rouse MJ, Croke D, Bradley C. The Development of an Accreditation

Framework for Continuing Education Activities for Pharmacists. *Pharmacy*. 2020;8(2):75. doi:10.3390/pharmacy8020075

22. Penm J, Vaillancourt R, Pouliot A. Defining and identifying concepts of medication reconciliation: An international pharmacy perspective. *Res Soc Adm Pharm*. 2019;15(6):632-640. doi:10.1016/j.sapharm.2018.07.020

23. Watson KE, Singleton JA, Tippett V, Nissen LM. Defining pharmacists' roles in disasters: A Delphi study. *PLoS One*. 2019;14(12). doi:10.1371/journal.pone.0227132

24. Gibbins AK, Wood PJ, Spark MJ. Managing inappropriate use of non-prescription combination analgesics containing codeine: A modified Delphi study. *Res Soc Adm Pharm*. 2017;13(2):369-377. doi:10.1016/j.sapharm.2016.02.015

25. Jebara T, Cunningham S, MacLure K, et al. A modified-Delphi study of a framework to support the potential implementation of pharmacist prescribing. *Res Soc Adm Pharm*. 2020;16(6):812-818. doi:10.1016/j.sapharm.2019.09.005

26. Rankin A, Cadogan CA, Ryan C, Clyne B, Smith SM, Hughes CM. Core Outcome Set for Trials Aimed at Improving the Appropriateness of Polypharmacy in Older People in Primary Care. *J Am Geriatr Soc*. 2018;66(6):1206-1212. doi:10.1111/jgs.15245

27. Barry E, O'Brien K, Moriarty F, et al. PIPc study: Development of indicators of potentially inappropriate prescribing in children (PIPc) in primary care using a modified Delphi technique. *BMJ Open*. 2016;6(9). doi:10.1136/bmjopen-2016-012079

28. Greason M. Connecting Findings to Meaningful Change: The Benefits of Using Qualitative Delphi in Empirical Ethics and Policy Research in Long-Term Care. *Int J Qual Methods*. 2018;17(1). doi:10.1177/1609406918803271

29. Loo R. The Delphi method: A powerful tool for strategic management. *Policing*. 2002;25(4):762-769. doi:10.1108/13639510210450677

30. Dalkey N, Helmer O. *An Experimental Application of the Delphi Method to the Use of Experts*. California: Rand Corporation; 1962. doi:10.1287/mnsc.9.3.458

31. Likert R. A technique for the measurement of attitudes. *Arch Psychol*. 1932;140:44-53.

32. Uebersax J. Likert scales: Dispelling the confusion. http://john-uebersax.com/stat/likert.htm. Accessed November 15, 2020.

33. Carifio J, Perla RJ. Ten Common Misunderstandings, Misconceptions, Persistent Myths and Urban Legends about Likert Scales and Likert Response Formats and their Antidotes. *J Soc Sci*. 2007;3(3):106-116. doi:10.3844/jssp.2007.106.116

34. Rattray J, Jones MC. Essential elements of questionnaire design and development. *J Clin Nurs*. 2007;16(2):234-243. doi:10.1111/j.1365-2702.2006.01573.x

35. Johns R. SQB Methods Fact Sheet 1: Likert Items and Scales. *Surv Res Netw*. 2010;1(March):1-11. https://www.ukdataservice.ac.uk/media/262829/discover_likertfactsheet.pdf.

36. Brown J. Likert items and scales of measurement? *Shiken JALT Test Eval SIG Newsletter*. 2011;15:10-14. http://hosted.jalt.org/test/bro_34.htm.

37. Claveria O. A new metric of consensus for Likert-type scale questionnaires: an application to consumer expectations. *J Bank Financ Technol*. 2021. doi:10.1007/s42786-021-00026-5

38. Krosnick J, Presser S. Question and Questionnaire Design. In: Marsden P, Wright J, eds.

Handbook of Survey Research. 2nd edn. Bingley: Emerald Publishing; 2010:263-313.

39.   Louangrath PI, Sutanapong C. Validity and Reliability of Survey Scales. *Int J Res Methodol Soc Sci*. 2018;4(3):99-115.

40.   Johns R. Survey Question Bank: SQB methods fact sheet 1: Likert items and scales. https://www.sheffield.ac.uk/polopoly_fs/1.597637!/file/likertfactsheet.pdf. Published 2010. Accessed November 15, 2020.

41.   Warmbrod JR. Reporting and Interpreting Scores Derived from Likert-type Scales. *J Agric Educ*. 2014;55(5):30-47. doi:10.5032/jae.2014.05030

42.   Maeda H. Response option configuration of online administered Likert scales. *Int J Soc Res Methodol*. 2015;18(1):15-26. doi:10.1080/13645579.2014.885159

43.   Kuncel RB. Response processes and relative location of subject and item. *Educ Psychol Meas*. 1973;33(3):545-563. doi:10.1177/001316447303300302

44.   Kuncel RB. The subject-item interaction in itemmetric research. *Educ Psychol Meas*. 1977;37(3):665-678. doi:10.1177/001316447703700309

45.   Meshkat B, Cowman S, Gethin G, et al. Using an e-Delphi technique in achieving consensus across disciplines for developing best practice in day surgery in Ireland. *J Hosp Adm*. 2014;3(4):1. doi:10.5430/jha.v3n4p1

46.   Bishop PA, Herron RL. Use and Misuse of the Likert Item Responses and Other Ordinal Measures. *Int J Exerc Sci*. 2015;8(3):297-302.

47.   Trevelyan EG, Robinson N. Delphi methodology in health research: How to do it? *Eur J Integr Med*. 2015;7(4):423-428. doi:10.1016/j.eujim.2015.07.002

48.   Douven I. A Bayesian perspective on Likert scales and central tendency. *Psychon Bull Rev*. 2018;25(3):1203-1211. doi:10.3758/s13423-017-1344-2

49.   Weijters B, Cabooter E, Schillewaert N. The effect of rating scale format on response styles: The number of response categories and response category labels. *Int J Res Mark*. 2010;27(3):236-247. doi:10.1016/j.ijresmar.2010.02.004

50.   Lange T, Kopkow C, Lützner J, et al. Comparison of different rating scales for the use in Delphi studies: Different scales lead to different consensus and show different test-retest reliability. *BMC Med Res Methodol*. 2020;20(1):1-11. doi:10.1186/s12874-020-0912-8

51.   Hsu CC, Sandford BA. The Delphi technique: Making sense of consensus. *Pract Assessment, Res Eval*. 2007;12(10):1-8.

52.   Rowe G, Wright G, Bolger F. Delphi: A reevaluation of research and theory. *Technol Forecast Soc Change*. 1991;39(3):235-251. doi:10.1016/0040-1625(91)90039-I

53.   Hanafin S. *Review of Literature on the Delphi Technique*. Dublin: National Children's Office; 2004.

54.   Okoli C, Pawlowski SD. The Delphi method as a research tool: An example, design considerations and applications. *Inf Manag*. 2004;42(1):15-29. doi:10.1016/j.im.2003.11.002

55.   Meijering J V., Tobi H. The effect of controlled opinion feedback on Delphi features: Mixed messages from a real-world Delphi experiment. *Technol Forecast Soc Change*. 2016;103:166-173. doi:10.1016/j.techfore.2015.11.008

56.   Scheibe M, Skutsch M, Schofer J. Experiments in Delphi Methodology. In: Linstone H, Turoff

M, eds. *The Delphi Method: Techniques and Applications.* Massachusetts, Reading: Addison Wesley; 1975:262-287.

57. Woudenberg F. An evaluation of Delphi. *Technol Forecast Soc Change*. 1991;40(2):131-150. doi:10.1016/0040-1625(91)90002-W

58. Miller L. Determining what could/should be: The Delphi technique and its application. In: *Proceedings of the 2006 Annual Meeting of the Mid-Western Educational Research Association*. Columbus, Ohio; 2006.

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

60. 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. doi:10.1111/j.1365-2648.2006.03716.x

61. 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). doi:10.1371/journal.pone.0020476

62. Fletcher AJ, Marchildon GP. Using the delphi method for qualitative, participatory action research in health leadership. *Int J Qual Methods*. 2014;13(1):1-18. doi:10.1177/160940691401300101

63. Dalkey N, Helmer O. *The Use of Experts for the Estimation of Bombing Requirements: A Project Delphi Experiment*. Santa Monica: Rand Corporation; 1951. https://www.jstor.org/stable/2627117?seq=1.

64. Dayé C. How to train your oracle: The Delphi method and its turbulent youth in operations research and the policy sciences. *Soc Stud Sci*. 2018;48(6):846-868. doi:10.1177/0306312718798497

65. Magnuson LA. *A Delphi Study to Understand Relational Bonds in Supervision and Their Effect on Rehabilitation Counselor Disclosure in the Public Rehabilitation Program.* PhD Dissertation, University of Iowa; 2012.

66. Spiller HA, Hale JR, De Boer JZ. The delphic oracle: A multidisciplinary defense of the gaseous vent theory. *J Toxicol - Clin Toxicol*. 2002;40(2):189-196. doi:10.1081/CLT-120004410

67. de Boer JA. The Oracle at Delphi: the pythia and the pneuma, intoxicating gas finds, and hypotheses. In: Wexler P, ed. *History of Toxicology and Environmental Health: Toxicology in Antiquity, Volume 1*. Amsterdam: Academic Press; 2014:83-91.

68. Marchais-Roubelat A, Roubelat F. The Delphi method as a ritual: Inquiring the Delphic Oracle. *Technol Forecast Soc Change*. 2011;78(9):1491-1499. doi:10.1016/j.techfore.2011.04.012

69. Turoff M, Hiltz S. Computer based Delphi processes. In: Adler M, Ziglio E, eds. *Gazing into the Oracle: The Delphi Technique and Its Application to Social Policy and Public Health*. London: Kingsley; 1996:56-86.

70. Sackman H. *Delphi Assessment: Expert Opinion, Forecasting and Group Process*. Vol R-1283-PR. California: Rand Corporation; 1974.