

Quality indicators for Delphi studies

Jon Landeta¹  | Aitziber Lertxundi²

¹Faculty of Economics and Business, The University of the Basque Country (UPV/EHU), Bilbao, Spain

²Faculty of Economics and Business, The University of the Basque Country (UPV/EHU), Donostia-San Sebastián, Spain

Correspondence

Jon Landeta, Faculty of Economics and Business, The University of the Basque Country (UPV/EHU), Lehendakari Aguirre 83, 48015 Bilbao, Spain.

Email: jon.landeta@ehu.es

Funding information

None

Abstract

The Delphi method is a technique of social research that seeks to obtain a reliable group opinion from experts. It was first created for military purposes in the mid-1950s. Since then, its use in the scientific field has continued to spread to different disciplines and aims. Despite this expansion, however, not set of indicators of the quality of Delphi studies has yet to be developed that might provide the reader—whether an expert in the technique or not—with some framework of reference whereby to gauge what credibility should be afforded to the results of the study. In this paper, following a thorough review of the literature on the criteria used to assess Delphi studies and the items of evaluation recommended for inclusion in Delphi reports, we determine what characteristics a quality evaluation indicator for this technique should have and propose a battery of indicators based on these characteristics, which should for preference be included in the final report of a Delphi study. The proposed indicators focus on three areas that are particularly relevant to the quality of Delphi research: the quality of the panel of participating experts, the way in which relevant information is obtained from the experts, and the quality of the interaction generated among the experts.

KEY WORDS

Delphi method, evaluation, indicators, methodology, quality, reporting

1 | INTRODUCTION

The Delphi method is a technique of social research that seeks to obtain a reliable collective opinion from a group of experts. It is a method of structuring a group communication process that is effective in allowing a group of individuals jointly to address a complex problem (Linstone & Turoff, 1975; Adler & Ziglio, 1996) using a methodology whose classic version is characterized by (a) an iterative process (experts are asked at least twice about the same issues); (b) participant anonymity (the experts do not know who has said what, thus avoiding unwanted influences among them); (c) controlled feedback (there is no direct interaction between the experts; instead, the leaders of the study furnish all participants with

the group opinion—in the form of a mean or median—and relevant individual contributions, before asking them again to give their opinions on the initial questions), and (d) a statistical group response (the opinions of all participating experts are contained in the final group response) (Rowe & Wright, 2001).

This technique was first conceived, developed, and applied in the early 1950s at the Rand Corporation, a private American organization which, at the height of the Cold War, sought to connect military planning to scientific research and development (Dalkey & Helmer, 1951). As initially devised, it was intended to provide forecasts of potential future events (initially of a military nature) by obtaining a consensus among the participating experts, when there were no objective past data on which to build projections or models,

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2023 The Authors. *Futures & Foresight Science* published by John Wiley & Sons Ltd.

or when it was felt that past behavior patterns could not be extrapolated to the future (Dalkey & Helmer, 1963).

Researchers at the Rand Corporation were aware of the limitations of subjective judgment as a basis for decision-making (Dalkey, 1968; Helmer & Rescher, 1959), but they were also conscious of the need to provide decision-makers working in conditions of uncertainty with reliable and reasoned information of an initially higher quality than resorting to mere experience and intuition (Helmer, 1967, 1983). They therefore wanted to develop a technique that would take advantage of the superiority of subjective group judgment over individual judgment, given the availability to the group of more information and knowledge and the positive effect the group could exert on the individual contributions of its members. At the same time, they wanted to minimize the undesirable effects of direct interaction between group members—effects such as the inhibition experienced by some individuals due to personal character traits or communication difficulties (Collins & Guetzkow, 1964); the influence some group members exercised over others based not on knowledge but on status or dominant personality; responses conditioned by a desire for acceptance by the group; distraction (Collins & Guetzkow, 1964); groupthink (Janis, 1982); and premature closure of the process (Lock, 1987).

Due to the technique's relatively simple methodology and the variety of goals that can be achieved with it, it has spread rapidly from the area of military forecasting to many different research disciplines, with a diverse range of purposes.

The original Delphi methodology has been adapted for different conditions and purposes. Among the most noteworthy modifications to the initial design—for their early development and their subsequent and current success—are the Policy Delphi (De Loë et al., 2016; Turoff, 1970) and the Real-Time Delphi (Gnatzy et al., 2011; Gordon, 2009; Gordon & Pease, 2006; Turoff, 1972). The Policy Delphi is primarily oriented toward dissent, or, at least, toward exploring divergent positions, i.e., bringing to light different and often opposing stances on a given problem, with their associated

arguments. Real-time Delphi is a computer-based, nonsequential modality, in which experts can make contributions and modify their answers continuously over a predetermined period of time, without adhering to closed rounds and without their contributions being filtered and coordinated by the researchers conducting the study. This significantly shortens the time needed to conduct it and its cost. Other methodological variants in the literature include the Mini Delphi (Helmer, 1967; Nelms & Porter, 1985) or the Hybrid Delphi (Landeta et al., 2011).

It has consequently been used in a very wide array of applications and has become a commonly used method in research and consultancy, especially in health and social sciences. This is evidenced by the increasing number of published articles and doctoral dissertations in which it is used (Flostrand et al., 2020; Gupta & Clarke, 1996; Landeta, 2006). The latest review published by Flostrand et al. (2020) contains data to mid-2017 showing the number of articles containing the words "Delphi" and "Method" in the title, keywords, or abstract, collected by years and disciplines in the Scopus database. The results clearly show a sustained upward trend in use of this technique, globally and by areas (see Table 1), with an even higher-than-average increase in total publications in this database. Thus, 2995 Delphi-related articles were published in Scopus-indexed journals in 2022, as compared to 1331 in 2017. This represents a 125% rise, greater than the relative increase (115%) in the total number of articles in the database during this period. There has therefore been an increase in both absolute and relative terms. Particularly noteworthy is the total number of articles published in health sciences (65% of the total), and the relative increase over the last 6 years of publications in engineering (438%) and computer sciences (256%).

However, despite its ever-greater and more consolidated use as a social research technique, Delphi is still relatively unknown, at least beyond the superficial definition of "group consultation of experts seeking a consensus, expressed in means or medians, and quartiles to show the degree of dispersion of responses." In other words, most

TABLE 1 Published articles related to the Delphi method.

2017	2018	2019	2020	2021	2022	1968–2022	Subject areas
799	968	1095	1371	1691	1821	14,254	Medicine, nursing, and health professions
170	201	235	305	391	468	3184	Social sciences
87	128	112	151	211	468	2944	Engineering
50	59	68	98	139	178	2069	Computer science
108	107	108	135	153	218	1951	Business, management and accounting
71	91	108	178	206	236	1528	Environmental sciences
42	54	43	82	88	90	794	Biochemistry, genetics and molecular biology
40	39	47	83	119	107	708	Energy
1331	1606	1726	2202	2645	2995	21,858	Total

Note: Increase in the number of annual Delphi articles collected in the Scopus database from 2017 to 2022: 125%. Increase in the total annual number of articles collected in the Scopus database from 2017 to 2022: 115%.

Source: Authors based on Scopus data.

readers or researchers consulting or using the results of a published Delphi application do not have (or need) the knowledge required to gauge whether the technique has been used rigorously and, therefore, whether the results they are consulting are as good as they believe. At the same time, no set of indicators or points of contrast showing and measuring the degree of quality of a Delphi study has been consolidated in the academic literature, of a kind similar to the statistical contrasts used to measure the goodness of fit of an econometric model or the reliability and validity of selected variables (Diamond et al., 2014; Niederberger & Spranger, 2020; Schmidt, 1997). Finally, there is no widely used and accepted model for reporting the results of Delphi research, with key points, including proof of the quality of the study (Belton et al., 2019; Hardy et al., 2004; Humphrey-Murto et al., 2017; Niederberger & Spranger, 2020; Powell, 2003; Spranger et al., 2022).

Aware of the lack of consolidated standards for rigorous reporting of the results of Delphi studies, and the practical absence of quality (or goodness) indicators in published articles, numerous authors have discussed the need to make Delphi studies more rigorous and credible—among them Banno et al. (2019, 2020), Foth et al. (2016), Hasson and Keeney (2011), Hasson et al. (2000), Humphrey-Murto and de Wit (2019), Iglesias et al. (2016), Jünger et al. (2017), Niederberger and Spranger (2020), Spranger et al. (2022), Toma and Piciooreanu (2016).

In view of this shortcoming, this paper seeks to offer a reasoned proposal for a synthetic set of indicators that could be included in any published study and which would contribute to measuring the quality of a Delphi study and help (even nonspecialist) readers to gauge the reliability of its results.

Given the variety of objectives, areas of application, and designs covered by the Delphi methodology, as well as the different scientific paradigms under which a Delphi application can be evaluated, it is practically impossible to make a proposal for indicators that would be suitable for all Delphis. Our proposal will therefore focus mainly on the more classical design, although most of the indicators will be capable of being used in any work that can be qualified as Delphi.

2 | LITERATURE REVIEW

Throughout its long history, the Delphi method has been applied in different areas (Flostrand et al., 2020; Gupta & Clarke, 1996; Landeta, 2006), for different purposes (Linstone & Turoff, 1975; Niederberger & Spranger, 2020) and with different designs (Hasson & Keeney, 2011; Landeta, 1999; Spranger et al., 2022). This ductility of technique has prevented any single valid way of conducting research (Rowe & Wright, 2001; Woudenberg, 1991) and, thus, of reporting and evaluating the results (Hasson & Keeney, 2011; Schmidt, 1997; Spranger et al., 2022) from prevailing over all others.

Concerns about the quality of Delphi studies, and the methodology itself, have been voiced since its inception. The greatest criticism was leveled against the technique from within its very birthplace at the Rand Corporation by Harold Sackman (1974). Sackman was an

exponent of the traditional scientific school, quantitative and purist in its methods (Rieger, 1986), in opposition to the new, more qualitative, and applied approach of the Delphi method. He attacked the technique in the round, from two perspectives. Firstly, he tested Delphi against the psychometric standards of the APA, and finding that it failed to meet many of them, he argued that the scientific nature of the technique was highly debatable. Secondly, he criticized its methodological principles and assumptions. Sackman thus concluded that it was neither a reliable technique nor was it validated by scientific procedure. Coates (1975) and Goldschmidt (1975) immediately countered by claiming that the proper criterion for assessing this technique was its social utility, not its scientific correctness, and that the standards of reliability and validity employed by Sackman were positivist criteria, valid for psychometric tests but not for this type of qualitative technique. They concluded that this technique was useful for alerting participants to the complexity of the issues involved, making them think and challenging their own assumptions, and that Sackman's criticisms were based more on misapplication of the technique than on actual methodological weaknesses.

These debates highlighted two fundamental elements for appraising the quality of Delphi studies:

- a) Validity and reliability are quality assessment criteria that can be taken as benchmarks, but they should not be applied with the same standards as in quantitative research.
- b) Rigour in applying the technique, adapting it to the user's context and correctly interpreting its results is fundamental for overcoming many of the limitations attributed to it.

In this regard, Trustworthiness (Lincoln & Guba, 1986; Krefting, 1991) is an approach that can be used to overcome, or make up for, the limitations of positivist criteria of reliability and validity in the measurement of quality in qualitative research, as in the case of the Delphi (Day & Bobeva, 2005). According to Guba and Lincoln (1982), within a post-positivist scientific paradigm, this trustworthiness can be configured through the criteria of credibility, dependability, confirmability, and transferability, in parallel with the positivist criteria used in the evaluation of quantitative research, but with more relaxed standards of rigour (Morrow, 2005).

- Credibility corresponds to the criterion of internal validity, and is defined as the quality of the results to be taken as credible, which according to Creswell and Miller (2000), can be determined using different lenses: the lens of the researcher (to see whether the constructs, categories, explanations and interpretations make sense (Patton, 1990), the lens of the participants (involves participants assessing whether the interpretations accurately represent them) (Belton et al., 2019), and the lens of individuals external to the study (unaffiliated reviewers who may help to establish validity and diverse readers for whom the account is written). Credibility can be achieved through negative case analysis, researcher reflexivity and participant checks, validation

and co-analysis (Morrow, 2005), as well as by providing information on researchers' previous training and experience, personal connections to the background theory, access to the study population and funding sources (Johnson et al., 2020).

- Dependability, on the other hand, emphasizes the demand for reliability (Morrow, 2005). Reliability, defined as consistency of measurement within a given study (Hair et al., 1998), that is, the extent to which a study conducted by different researchers, different experts or at different points in time arrives at similar results. It is a quality criterion that is difficult to apply to a specific Delphi study, at least in general terms, given the characteristics of this technique. By their very nature, Delphi studies seldom if ever produce the same information at different points in time, since it is mainly obtained through the subjective judgment of the participating experts, which may change over time. Moreover, the researcher intervenes to a significant degree in the process (Patton, 1990). Any increase in the reliability of this technique should therefore result from greater formalization and quantification of its performance and the rigorous use of defined methodologies. This reduces the variability resulting from the researcher's personal intervention and the different contributions of the same experts. For this purpose too, the use of Delphi pilots, to test whether all experts have the same understanding of the issues on which they are asked to give an opinion, could be assimilated to a parallel mode measurement, thus increasing the reliability of the application under analysis (Hasson & Keeney, 2011). Finally, one possible way of assessing the intra-response reliability of a Delphi application might be to ask the Delphi questions in two different ways (e.g., using reversed wording), as suggested by Belton et al. (2019), although we have not seen this practice used in any published study.
- Transferability is linked to external validity or generalizability. External validity, considered as the extent to which the results of the study can be generalized, is not decisive for a Delphi application based on a "convenience sample of experts," but to the extent that this sample is large, representative, and heterogeneous, it is possible to approach this type of validity (Bolger & Wright, 2011). The researcher conducting qualitative research is expected to provide sufficient information about him/herself, the context, processes, and participants and the researcher's relationships with them to enable the reader to assess the extent to which the research findings are transferable to other situations (Belton et al., 2019; Hasson & Keeney, 2011; Morrow, 2005).
- Finally, Confirmability is associated with the objectivity of the results, which implies that results should, as far as possible, represent the subject of study, rather than the beliefs or biases of the researcher (Gasson, 2004).

Trustworthiness, under other paradigms, is achieved by other criteria specific to these paradigms (Hideg, 2013; Morrow, 2005).

From an interpretive-constructivist paradigm, it is noted that the criterion of authenticity (Guba & Lincoln, 1989) is relevant. Fairness and authenticity (in their several variants) contribute to

trustworthiness. Thus, the diversity of interpretations among the participants, as well as being welcome in itself, matures and enriches the final result. At the same time, it stimulates a better understanding of other people's arguments. In this sense, the reflexivity of Delphi facilitators contributes to making them more aware that their own experiences and interpretations of the environment around them also affect the research being carried out (Morrow, 2005).

Likewise, within the critical-ideological paradigm, it is necessary to assess the extent to which the research has achieved its objectives of social and political change, if any, and the degree to which it has fostered a more critical discourse (Morrow, 2005).

In all cases, according to Morrow (2005), there are certain standard general criteria that ensure Trustworthiness, regardless of the scientific paradigm taken as a reference:

- *Social validity*: the purpose and usefulness of the study is clear and real, and the results are trustworthy and appear useful to the study's stakeholders and to the participating experts themselves (Landeta, 2006).
- *Adequacy of data*: the participants are genuine experts who are actively involved in designing the study and obtaining the results; they are sufficient in number and representative, and suitable strategies are used for eliciting and enriching their contributions. The development of the study is described in sufficient detail and rigour, including numerical, statistical and qualitative data.
- *Adequacy of interpretation*: the researchers conducting the study are experts in the technique and in the object of study who follow an appropriate methodology to adequately analyze and interpret the information they receive from the experts and from other sources used.
- *Adequate management of subjectivity*: every attempt should be made to avoid or mitigate, as far as possible, unwanted bias among the researchers conducting the study in their handling of qualitative information and in the design of the study or, at least, to be aware of and share with others any implicit assumptions and judgments of the researchers that might condition their research.

Consistent with this objective of improving the quality of implementation and communication of Delphi studies, and despite the difficulties listed at the beginning of this section, various researchers have been aware of the need to propose a reporting model that would include key aspects of Delphi research and would enable readers to judge the rigour with which the study was conducted and the decisions that were made. From a literature review conducted in the Scopus database, using the terms "Delphi," "method," and "reporting" in keywords, abstract and title, we have extracted the following studies, which offer checklists to facilitate complete and rigorous reporting: Boulkedid et al. (2011), Diamond et al. (2014), Hasson et al. (2000), Humphrey-Murto et al. (2017), Iglesias et al. (2016), Jünger et al. (2017), Sinha et al. (2011), Toma and Piciooreanu (2016), Niedemberg and Spranger (2020), Spranger et al. (2022).

These works propose guidelines for correct reporting, generally based on a review of a large number of published Delphi applications

and/or other review articles. However, few allude to the incorporation of a set of key indicators, identified as such, in the report presenting the results, which might serve to assess the quality of the results of the study being presented; even fewer discuss and propose any specific indicators for inclusion.

Thus, Hasson et al. (2000) are concerned with the reliability and validity of the technique, or its credibility, but do not propose concrete indicators for measuring it. Diamond et al. (2014), in their systematic review of 98 studies aimed at assessing consensus, used four quality indicators to classify the studies analyzed: were the participant-selection criteria given?; was the number of rounds to be carried out specified? was the criterion for rejecting items in each round clear?; was any other criterion specified for ending the study, apart from the number of rounds to be conducted? However, these indicators were used to assess and rank the papers but were not proposed for inclusion in the reporting guide. Only Spranger et al. (2022), in the most recently published proposal for a Delphi study reporting guide, propose—as the 13 of 15 points in their guide—the inclusion of quality criteria for qualitative or quantitative research, which should be applied depending on the orientation and variant of the Delphi conducted. However, they do not propose any specific criteria.

However, a few published works do use certain quality indicators for the Delphi application presented. Thus, Landeta (2006) uses as indicators the quality and stability of the panel of experts (percentage of acceptance of the invitation to participate and dropout rate), the time elapsed between rounds, the quality and intensity of the experts' participation (duration of the data collection interviews and number of qualitative contributions per question and expert), the modification of the initial response, derived from iteration and feedback (percentage of experts who modified their responses in the second round) and the greater degree of consensus and convergence of opinions in the second round, measured by the average standard deviation of the responses. Landeta et al. (2008) add to the previous indicators, mainly related to the reliability of Delphi application, an indicator of its validity, comparing the results achieved using this technique (estimates of input/output tables for Catalonia) with other results obtained using other techniques (in this case surveys and statistical inference carried out by the Spanish National Institute of Statistics). In a later publication, Landeta and Barrutia (2011) incorporate as quality indicators the degree of acceptance of the study and its results, both by the panel of experts involved and the users of the study results and by the people who commissioned it, based on specific surveys carried out after completion and implementation of the study. These indicators provided external information closely linked to the assessment of the application in terms of its capacity to adapt to the social and political context of its development and the usefulness as perceived by recipients and users of the study. Finally, Etxeandia-Pradera et al. (2022) use some of the indicators mentioned above: increasing the degree of consensus, degree of change of opinion from the first to the second round, expert participation (dropout rate of experts and average number of qualitative contributions per expert and question) and expert panel satisfaction (by means of a satisfaction survey included in the second round).

3 | PROPOSAL OF INDICATORS

The purpose of this contribution is to propose a synthetic and reasoned list of indicators of the quality of a Delphi exercise, which must be related to the quality criteria listed in the previous section.

The aim is not to evaluate the technique itself. Rather, it is to provide third parties reading the report of a Delphi exercise with instruments that will reflect the quality of the study and, by extension, its results. The purpose is to separate, as far as possible, the evaluation of the goodness of a Delphi application from the subjective evaluations made by those conducting the study, and to steer it toward "standardized" and objective indicators that an external observer can judge and appraise, without necessarily being an expert in the Delphi technique. The proposal is that this set of indicators should be included in a final section of the Delphi study report. This section need not include all the elements that establish the rigour of the Delphi application (these would be part of the full report); rather its purpose is to identify a few relevant indicators that provide information on the expected quality of the results of the Delphi application being presented.

We believe that in order for this list of indicators to be useful and its use to be generalizable, it must fulfill certain characteristics:

1. It must be short and synthetic. A limited number of indicators should be included in the final Delphi report, preferably after the results section.
2. The indicators should be simple for the researchers to assess. They should not require further or parallel research, or a high degree of technical knowledge. This would make it relatively easy to incorporate them into the study and prevent any noninclusion from being attributed to a lack of time or resources.
3. The indicators should be easy for readers to understand. By observing the results of an indicator, an inexperienced reader should be able to determine whether the work is rigorous and the results credible, without requiring special training. Where necessary, they should be accompanied by some optimal, widely accepted values for comparison and reference.
4. The indicators should be relevant. They should provide information on elements that are truly transcendental to the quality of the study and must therefore be related to the scientific quality criteria generally accepted in qualitative research.
5. The indicators should be as objective, quantitative, and external as possible. The aim is to avoid, as far as possible, the bias of the researcher conducting the study, even if he/she is the one who must prepare them.
6. The indicators should seek to satisfy the scientific quality criteria generally accepted by the scientific community for qualitative studies: validity, reliability, and trustworthiness.

In drawing up this list, we can also go back to the origins of the method in the Rand Corporation, and draw on the reflections of Olaf Helmer, one of its creators. Helmer (1965) wrote about the three areas in which action needs to be taken to improve the scientific use

of the subjective appraisals made by individuals (experts, in our terminology):

- (a) Improving the selection of the most suitable individuals, that is, having people who can really contribute to the objective of the study, because of their experience, knowledge, involvement, and motivation.
- (b) Improving the effective performance of the expert, that is, ensuring that the expert contributes what he/she can correctly.
- (c) Improving the development of methodologies that make it possible to achieve valuable results, based on the use of subjective information. This is the case of Delphi itself, as an integrator of individual contributions and enhancer of new contributions, so that the final result is superior to any of the individual contributions.

The first group of indicators should therefore refer to the value of the group of experts who have participated in the study, with an expert being deemed to be any person who can make useful contributions to the object of the study (Pill, 1971). Aspects to consider include:

1. **Number:** there is no predetermined number. Experiments carried out in the embryonic phase of the Delphi indicated that from seven experts upwards, there was an increasingly reduced improvement in the precision of the results (Dalkey, 1969). However, seven homogeneous experts are needed, that is, for each area of expertise required (field of knowledge, professional, role, etc.). Therefore, a reference could be 5–10 experts, per type of expert (Delbecq et al., 1975), albeit we are aware that this is a small number for statistical comparisons, or 15–30, for homogeneous panels (Belton et al., 2019; Clayton, 1997).
2. **Quality of the experts:** concise and objective information indicating the degree to which they fulfill what is required (average years of experience, average number of publications, professional or academic degree, average self-assessment by question or topic, absence of conflict of interest, etc.) (Mauksch et al., 2020; Van Zuuren et al., 2022). Such information could also include the results, of dissent, desirability bias, outlier, bipolarity, or stakeholder group analyzes among experts, if performed (Beiderbeck et al., 2021).
3. **Representativeness of the participating experts:** number of fields of knowledge required and number of experts per field (Beiderbeck et al., 2021; Mauksch et al., 2020).
4. **Motivation:** percentage of dropouts between rounds, percentage of experts providing additional qualitative information, and so on.

The second group of indicators focuses on how the study has managed to extract the maximum knowledge and experience from the expert, with the minimum of bias and influence from the researcher. The purpose is not to repeat what should already have been explained in the section on research development, but of listing objective evidence showing that the experts have effectively

contributed their knowledge and experience in the direction required by the study. We propose three main focuses:

1. Clear configuration of a common context. In other words, enumeration of concrete items that show that all the experts understood the meaning of the research and the issues on which they were being asked to give answers: study presentation session, initial interview with each expert, study presentation text, additional information provided by the researchers for each question (where applicable, objective data provided to focus the answers), diversity of sources of evidence to understand and interpret the context, and so on.
2. Clear questions (understood in the same way by all the experts), unbiased and free from the influence of the researchers: whether initial open questions were asked for the purpose of drawing up the subsequent Delphi questions, whether pilot tests or Delphis were carried out to check understanding, whether there were channels of interaction with the experts to ensure proper interpretation of the questions, and so on.
3. Compensation and conditions previously agreed with the experts: remuneration, commitment to deliver preliminary results, commitment to deliver publications, levels of anonymity, number of rounds or effort required, and so on.
4. Quality and relevance of information-gathering mechanisms: interviews, website, or software used. Particularly in the case of Real-Time Delphi, the success of its implementation depends to a large extent on the capabilities of the selected software, in terms of its features, data output, user friendliness, and administration (Aengenheyster et al., 2017; Beiderbeck et al., 2021).
5. Indicators of the researchers' proficiency in obtaining, processing, improving, and interpreting the information provided by the experts (expertise in the technique, subject matter, own assumptions, teamwork, and so on (Morrow, 2005)).

The third group of indicators focuses on showing the degree to which the technique has achieved its intended methodological objectives, i.e., to provide interaction between experts, as free as possible from unwanted interpersonal biases and influences, in such a way that with the new information compiled they have been able to change their initial beliefs and make contributions of greater value, resulting in a superior group result (Barrios et al., 2021; Bolger & Wright, 2011; Meijering & Tobi, 2016; Rowe et al., 1991; Yaniv & Milyavsky, 2007). These indicators might relate to:

1. **Information exchanged:** number or percentage of experts who provided rationales (qualitative input), number of new inputs (not included in the questions) collected, number of new rationales (feedback) provided to the experts per round (Drumm et al., 2022), and so on.
2. **Opportunities for change of opinion:** number of rounds. The logical minimum would be two rounds plus—where appropriate—a prior preliminary round of open questions, so that there is at least one opportunity to incorporate the feedback received from the

- first round in each expert's individual reflection in the second round. This indicator would obviously not be valid for Real-Time Delphis, although it could be replaced by others that measure this opportunity, for example, the time that the window for experts to provide rationales and modify previous responses remains open.
3. Stability of external conditions. A change of context between each round may cause changes of opinion to occur not because of the interaction and the Delphi methodology, but because of variations in environmental variables external to the technique and the panel of experts. Possible indicators might include: time taken to conduct the study from the sending of the first questionnaire to the closing of the last round, time elapsed between each round, total time from the beginning to the presentation of the results. Three to 4 weeks per round appears to be an acceptable period (Beiderbeck et al., 2021; Belton et al., 2019).
 4. Internalization of the information exchanged. If the participants in a Delphi study do not change their minds, it is probably a sign that the study has been poorly conducted or that the experts did not take it seriously. In either case, it suggests that the study should not have been conducted. A simple survey would have been quicker and cheaper. Indicators that the technique has influenced the experts, shifting their opinion toward a presumably more reasoned and higher quality opinion may be: percentage of experts who have changed their opinion per question; variation in each round in the measures of central tendency and dispersion used; variation in the degree of consensus (preferably by homogeneous subgroups of experts).
 5. External evaluation of the key elements of the study by external stakeholders, users or, more easily, by the experts participating in the study. It is important to introduce an element of evaluation and control that is external to the researcher conducting the study. This indicator is fully aligned with the criteria of trustworthiness (Creswell & Miller, 2000) and social utility (Coates, 1975), yet, surprisingly, it is very rare to find it in published Delphi applications. Indicators of this type might include:
 - a. Results of a final short survey of the panel experts, included in the last round (Etxeandia-Pradera et al., 2022), or conducted subsequently (Landeta et al., 2011; Turnbull et al., 2018), measuring, for example, *the degree to which the method has been effective in obtaining input from the expert panel members, the degree to which the feedback received has helped them to improve their confidence in the answers given, their confidence in the technique or methodology used, the quality and clarity of the presentation of the questions, their satisfaction with participation in the study, and the usefulness of the results achieved*. Another possibility would be to provide the output from a final focus group with the experts (Gisbert-Trejo et al., 2023).
 - b. Results of a survey of project stakeholders: persons or representatives from the institution that commissioned or financed the research, or persons who use or apply the results of the study (Landeta & Barrutia, 2011). This might include *their assessment of the degree to which the intended objectives of*

the research have been achieved, the degree of satisfaction of initial expectations, the usefulness of the results obtained, the intention to commission a Delphi again if similar circumstances and needs arise again.

- c. Results of a survey or consultation with experts in the Delphi methodology external to the research. This indicator, which is common in the dynamics of publication of scientific articles through anonymous reviewers, is not known to have been referenced in the report of any published study. Its use might undoubtedly increase the confidence of the reader of the study report, provided, obviously, that his or her assessment is good. These external experts should give their judgment on the key aspects of the study (Spranger et al., 2022): *appropriateness of the choice of technique for the intended objectives, quality, and suitability of the participating experts, questionnaire wording, research design, rigor in the development of data collection, quality of the data obtained, data analysis, presentation of results, interpretation, and discussion of the results, ethical assessment of the study and its development.*

The first two groups of external indicators are aimed, mainly, at assessing the extent to which the Delphi study has been able to respond to the user context, that is, to the conditions and needs of the participants, of the users of the results and of the agents that promoted it. In short, it seeks to evaluate its actual practical usefulness. The third group of indicators is oriented more toward the objective, technical, and scientific external evaluation of the study presented.

The results of these external evaluations on each of the above-mentioned topics, preferably measured on a quantitative scale of seven points, for example, should be analyzed by type of external evaluator, calculating their medians and ranges, or means and standard deviations. These would be the data that would be presented as quality indicators, simple to calculate by the authors of the study and easy to interpret by the readers of the study.

The battery of indicators finally selected (see Table 2) should be included in the final report, after the results, and before the conclusions, limitations, and future lines of research.

All indicators should be estimated and included by the researcher who has conducted the Delphi application, with the exception of the indicators that include external evaluations (fifth section of the third group of indicators), which would be fed by the information provided by these external evaluators, who have not conducted the Delphi study, although the survey administrators would be responsible for requesting this information, analyzing it and reporting it synthetically and objectively.

4 | CONCLUSIONS

If the Delphi method is to continue to gain acceptance among the scientific and professional community, it requires methodological instruments that will endow it with even greater rigour and credibility.

The proposal to include quality indicators in reports on the results of Delphi studies is intended to meet this purpose. Such

TABLE 2 Quality indicators for Delphi studies.

- Indicators of the value of the group experts
 - Appropriate number (indicative minimum of 5–10 per area of expertise)
 - Quality (expertise, experience, independence, etc.)
 - Representativeness
 - Motivation (dropouts, rationales provided, etc.)
- Indicators of adequate and effective expert contribution
 - Configuration of a common context
 - Clear and unbiased questions
 - Explicit benefits to experts for their contribution.
 - Quality and adequacy of the mechanisms for obtaining and managing the experts' knowledge (interviews, software, etc.).
 - Researchers' competence
- Indicators of the performance of the Delphi on experts' contributions
 - Quantity and quality of information exchanged
 - Opportunities to change opinion (number of rounds)
 - Stability of external conditions (time between each round, etc.)
 - Change in the experts' opinion
 - Assessment of the rigour, trustworthiness, and usefulness by non-conductors of the study (surveys of the panel of participant experts, users of the results of the study, funders of the research, experts in Delphi methodology, etc.).

Source: Authors.

indicators should take as their reference point the quality criteria of validity and reliability used to assess quantitative psychometric studies. Nonetheless, there should be an awareness that the same degree of rigor cannot be demanded, due to the very characteristics of the Delphi technique (which is qualitative in nature and based on the subjective opinion of a convenience sample of experts). Furthermore, external readers should be able easily to gauge the degree of trustworthiness the study warrants and its perceived usefulness.

The report should describe the process of origin, development, results, and evaluation of the Delphi study in sufficient detail to enable the reader to assess the rigor with which it has been carried out, its transferability to other contexts, situations or times, and the quality of the research. For this last task, it is essential to include a limited set of quality indicators that are easy to appraise and simple to understand, measure genuinely relevant aspects of the quality of the research and are as objective and free from the bias of the researcher conducting the study as possible.

To our knowledge, our proposal for indicators under these conditions is the first to be formulated in a structured, reasoned, and detailed fashion. It focuses on three areas that are particularly relevant to the quality of a Delphi study: the quality of the panel of participating experts, the way in which valuable information is obtained from the experts, and the quality of the interaction generated between the experts.

Use of the indicators should contribute to improving understanding of the quality of each Delphi application, both among those conducting the study and—above all—by readers without links to the research. At the same time, if researchers are aware on commencing a Delphi study that they are going to have to present quality

indicators at the end, they will probably be motivated to work harder and more rigorously to ensure that these indicators offer a more positive picture of their work. Thus, the very existence of consolidated indicators might, by itself, help to improve the average quality of the Delphi research being carried out. Finally, the incorporation of reasoned, objective, consolidated, and quality measuring elements, relatively free from researcher bias, marks another important step in moving this technique, based on subjective information, from the position of the "oracle of the gods" toward the scientific method of acquisition, transmission, and contrast of knowledge, which should guide scientific progress.

5 | LIMITATIONS AND FUTURE LINES OF RESEARCH

We have presented a personal methodological proposal, based on our theoretical and applied knowledge of the technique. The article attempts to introduce a reasoned proposal for the inclusion of indicators of quality of the design and execution of a generic Delphi application, not specifically linked to a specific field of knowledge or to a specific paradigm of interpretation. A subsequent approach would require an in-depth study based on different scientific paradigms under which a phenomenon or the results of the execution of a given technique can be assessed, to determine a set of indicators adjusted to a specific field of knowledge, such as, in our case, that of futures (Hideg, 2013).

Moreover, the selection of the control areas and indicators has not been contrasted and, each expert researcher in this technique might possibly construct his or her own. Many of the proposed indicators have been used satisfactorily by the authors of this work in previous studies, most of them published. However, so far, no assessment has been made, from the user's perspective, of the relevance, usefulness, and benefits of each of the suggested indicators.

It therefore requires the endorsement of the scientific community and practitioners, both in its theoretical aspects (which indicators are most relevant) and in its application (which indicators can most feasibly be used). It is intended as a discussion that is open to the knowledge and experience of academics and practitioners.

However, it would be useful to reach an acceptable level of consensus on which and how many indicators should be used, or at least on a minimum set of indicators that could be common to any study, to advance toward consolidating their use in academia and business. Should we try a dose of our own medicine? Would a Delphi study of this subject be an apt contribution to the advancement of the rigor and scientific acceptance of this technique?

ACKNOWLEDGMENTS

This work was improved thanks to the review and valuable suggestions for improvement made by George Wright (University of Strathclyde, Glasgow), José Guimón (Universidad Autónoma de Madrid), and the anonymous reviewers of F&F. Open access funding

by the University of the Basque Country (UPV/EHU). Research Group IT1735-22 Innovation Management, Responsible People and Organizations, of the Basque University System 2022–2025.

DATA AVAILABILITY STATEMENT

There are no data associated with this work, except those extracted directly from Scopus database, which have not been processed or modified.

ORCID

Jon Landeta  <https://orcid.org/0000-0002-3283-1576>

REFERENCES

- Adler, M., & Ziglio, E. (1996). *Gazing into the oracle: The Delphi method and its application to social policy and public health*. Jessica Kingsley Publishers.
- Aengenheyster, S., Cuhls, K., Gerhold, L., Heiskanen-Schüttler, M., Huck, J., & Muszynska, M. (2017). Real-time Delphi in practice—A comparative analysis of existing software-based tools. *Technological Forecasting and Social Change*, 118, 15–27. <https://doi.org/10.1016/j.techfore.2017.01.023>
- Banno, M., Tsujimoto, Y., & Kataoka, Y. (2019). Reporting quality of the Delphi technique in reporting guidelines: A protocol for a systematic analysis of the EQUATOR Network Library. *BMJ Open*, 9(4), e024942. <https://doi.org/10.1136/bmjopen-2018-024942>
- Banno, M., Tsujimoto, Y., & Kataoka, Y. (2020). The majority of reporting guidelines are not developed with the Delphi method: A systematic review of reporting guidelines. *Journal of Clinical Epidemiology*, 124, 50–57. <https://doi.org/10.1016/j.jclinepi.2020.04.010>
- Barrios, M., Guilera, G., Nuño, L., & Gómez-Benito, J. (2021). Consensus in the delphi method: What makes a decision change. *Technological Forecasting and Social Change*, 163, 120484. <https://doi.org/10.1016/j.techfore.2020.120484>
- Beiderbeck, D., Frevel, N., von der Gracht, H. A., Schmidt, S. L., & Schweitzer, V. M. (2021). Preparing, conducting, and analyzing Delphi surveys: Cross-disciplinary practices, new directions, and advancements. *MethodsX*, 8, 101401. <https://doi.org/10.1016/j.mex.2021.101401>
- Belton, I., MacDonald, A., Wright, G., & Hamlin, I. (2019). Improving the practical application of the Delphi method in group-based judgment: A six-step prescription for a well-founded and defensible process. *Technological Forecasting and Social Change*, 147, 72–82. <https://doi.org/10.1016/j.techfore.2019.07.002>
- Bolger, F., & Wright, G. (2011). Improving the Delphi process: Lessons from social psychological research. *Technological Forecasting and Social Change*, 78(9), 1500–1513. <https://doi.org/10.1016/j.techfore.2011.07.007>
- 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), e20476. <https://doi.org/10.1371/journal.pone.0020476>
- Clayton, M. J. (1997). Delphi: A technique to harness expert opinion for critical decision-making tasks in education. *Educational Psychology*, 17(4), 373–386. <https://doi.org/10.1080/0144341970170401>
- Coates, J. F. (1975). In defense of Delphi. *Technological Forecasting and Social Change*, 7, 193–194. [https://doi.org/10.1016/0040-1625\(75\)90058-X](https://doi.org/10.1016/0040-1625(75)90058-X)
- Collins, B. E. Y., & Guetzkow, H. (1964). *A social psychology of group processes for decision-making*. John Wiley & Sons.
- Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, 39(3), 124–130. https://doi.org/10.1207/s15430421tip3903_2
- Dalkey, N. (1968). *Experiments in group prediction*. The RAND Corporation.
- Dalkey, N. (1969). Analyses from a group opinion study. *Futures*, 1(6), 541–551. [https://doi.org/10.1016/S0016-3287\(69\)80044-3](https://doi.org/10.1016/S0016-3287(69)80044-3)
- Dalkey, N., & Helmer, O. (1951). *The use of experts for the estimation of bombing requirements*. A project Delphi experiment. The RAND Corporation.
- 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>
- Day, J., & Bobeva, M. (2005). A generic toolkit for the successful management of Delphi studies. *Electronic Journal of Business Research Methods*, 3(2), 103–116.
- Delbecq, A. L., Van de Ven, A. H., & Gustafson, D. H. (1975). *Group techniques for program planning: A guide to nominal group and Delphi processes*. Scott, Foresman. <https://doi.org/10.1177/1059601176001002>
- 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>
- Drumm, S., Bradley, C., & Moriarty, F. (2022). 'More of an art than a science'? The development, design and mechanics of the Delphi technique. *Research in Social and Administrative Pharmacy*, 18(1), 2230–2236. <https://doi.org/10.1016/j.sapharm.2021.06.027>
- Etxeandia-Pradera, J. I., Landeta, J., Gonzalez-Such, J., & Aguilar, E. J. (2022). How to improve training in descriptive psychopathology for psychiatry residents: A Delphi study. *Psychopathology*, 56, 206–219. <https://doi.org/10.1159/000525691>
- Flostrand, A., Pitt, L., & Bridson, S. (2020). The Delphi technique in forecasting—A 42-year bibliographic analysis (1975–2017). *Technological Forecasting and Social Change*, 150, 119773. <https://doi.org/10.1016/j.techfore.2019.119773>
- Foth, T., Efstratiou, N., Vanderspank-Wright, B., Ufholz, L. A., Dütthorn, N., Zimansky, M., & Humphrey-Murto, S. (2016). The use of Delphi and nominal group technique in nursing education: A review. *International Journal of Nursing Studies*, 60, 112–120. <https://doi.org/10.1016/j.ijnurstu.2016.04.015>
- Gasson, S. (2004). Rigor in grounded theory research: An interpretive perspective on generating theory from qualitative field studies. In M. Whitman & A. Woszcynski (Eds.), *The handbook of information systems research* (pp. 79–102). IGI Global.
- Gisbert-Trejo, N., Fernández-Ferrín, P., Albizu, E., & Landeta, J. (2023). Mapping mentee outcomes in mentoring programs for managers. *Industrial and Commercial Training*, 55, 284–294. <https://doi.org/10.1108/ICT-07-2022-0047>
- Gnatzy, T., Warth, J., von der Gracht, H., & Darkow, I.-L. (2011). Validating an innovative real-time Delphi approach—A methodological comparison between real-time and conventional Delphi studies. *Technological Forecasting and Social Change*, 78(9), 1681–1694. <https://doi.org/10.1016/j.techfore.2011.04.006>
- Goldschmidt, P. G. (1975). Scientific inquiry or political critique? *Technological Forecasting and Social Change*, 7(2), 195–213. [https://doi.org/10.1016/0040-1625\(75\)90059-1](https://doi.org/10.1016/0040-1625(75)90059-1)
- 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>
- Gordon, T. J. (2009). *Real-time Delphi: Futures research methodology V3.0. The Millennium Project*. <https://millennium-project.org/wp-content/uploads/2022/01/05-Real-Time-Delphi.pdf>
- Guba, E. G., & Lincoln, Y. S. (1982). Epistemological and methodological bases of naturalistic inquiry. *ECTJ*, 30(4), 233–252. <https://doi.org/10.1007/BF02765185>

- Lincoln, Y. S., & Guba, E. G. (1986). But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Program Evaluation*, 1986(30), 73–84. <https://doi.org/10.1002/ev.1427>
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Sage Publications.
- Gupta, U. G., & Clarke, R. E. (1996). Theory and applications of the Delphi technique: A bibliography (1975–1994). *Technological Forecasting and Social Change*, 53(2), 185–211. [https://doi.org/10.1016/S0040-1625\(96\)00094-7](https://doi.org/10.1016/S0040-1625(96)00094-7)
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis*. Englewood Cliff.
- Hardy, D. J., O'Brien, A. P., Gaskin, C. J., O'Brien, A. J., Morrison-Ngatai, E., Skews, G., Ryan, T., & McNulty, N. (2004). Practical application of the Delphi technique in a bicultural mental health nursing study in New Zealand. *Journal of Advanced Nursing*, 46(1), 95–109. <https://doi.org/10.1111/j.1365-2648.2003.02969.x>
- Hasson, F., & Keeney, S. (2011). Enhancing rigour in the Delphi technique research. *Technological Forecasting and Social Change*, 78(9), 1695–1704. <https://doi.org/10.1016/j.techfore.2011.04.005>
- 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>
- Helmer, O. (1965). *Social technology*. The RAND Corporation.
- Helmer, O. (1983). *Looking forward: A guide to futures research*. Sage.
- Helmer, O. (1967). *The future of science*. The RAND Corporation.
- Helmer, O., & Rescher, N. (1959). On the epistemology of the inexact sciences. *Management Science*, 6(1), 25–52.
- Hideg, É. (2013). Integral futures based on the paradigm approach. *Futures*, 45, S6–S15. <https://doi.org/10.1016/j.futures.2012.11.007>
- Humphrey-Murto, S., Varpio, L., Gonsalves, C., & Wood, T. J. (2017). Using consensus group methods such as Delphi and nominal group in medical education research. *Medical Teacher*, 39(1), 14–19. <https://doi.org/10.1080/0142159X.2017.1245856>
- Humphrey-Murto, S., & de Wit, M. (2019). The Delphi method—More research please. *Journal of Clinical Epidemiology*, 106, 136–139. <https://doi.org/10.1016/j.jclinepi.2018.10.011>
- Iglesias, C. P., Thompson, A., Rogowski, W. H., & Payne, K. (2016). Reporting guidelines for the use of expert judgement in model-based economic evaluations. *PharmacoEconomics*, 34, 1161–1172. <https://doi.org/10.1007/s40273-016-0425-9>
- Janis, I. L. (1982). *Groupthink*. Houghton Mifflin.
- Johnson, J. L., Adkins, D., & Chauvin, S. (2020). A review of the quality indicators of rigor in qualitative research. *American Journal of Pharmaceutical Education*, 84(1), 7120. <https://doi.org/10.5688/ajpe7120>
- 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/10.1177/02692163176906>
- Krefting, L. (1991). Rigor in qualitative research: The assessment of trustworthiness. *The American Journal of Occupational Therapy*, 45(3), 214–222. <https://doi.org/10.5014/ajot.45.3.214>
- Landeta, J. (1999). *El método Delphi: Una técnica de previsión para la incertidumbre*. Ariel.
- Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, 73(5), 467–482. <https://doi.org/10.1016/j.techfore.2005.09.002>
- Landeta, J., & Barrutia, J. (2011). People consultation to construct the future: A Delphi application. *International Journal of Forecasting*, 27(1), 134–151. <https://doi.org/10.1016/j.ijforecast.2010.04.001>
- Landeta, J., Barrutia, J., & Lertxundi, A. (2011). Hybrid Delphi: A methodology to facilitate contribution from experts in professional contexts. *Technological Forecasting and Social Change*, 78(9), 1629–1641. <https://doi.org/10.1016/j.techfore.2011.03.009>
- Landeta, J., Matey, J., Ruiz, V., & Galter, J. (2008). Results of a Delphi survey in drawing up the input-output tables for Catalonia. *Technological Forecasting and Social Change*, 75(1), 32–56. <https://doi.org/10.1016/j.techfore.2007.01.005>
- Linstone, H. A. & Turoff, M., (Eds.). (1975). *The Delphi method*. Addison-Wesley.
- Lock, A. (1987). Integrating group judgment in subjective forecast. In G. Wright & Ayton (Eds.), *Judgmental forecasting* (pp. 109–127). John Wiley & Sons.
- De Loë, R. C., Melnychuk, N., Murray, D., & Plummer, R. (2016). Advancing the state of policy delphi practice: A systematic review evaluating methodological evolution, innovation, and opportunities. *Technological Forecasting and Social Change*, 104, 78–88. <https://doi.org/10.1016/j.techfore.2015.12.009>
- Mauksch, S., von der Gracht, H. A., & Gordon, T. J. (2020). Who is an expert for foresight? A review of identification methods. *Technological Forecasting and Social Change*, 154, 119982. <https://doi.org/10.1016/j.techfore.2020.119982>
- Meijering, J. V., & Tobi, H. (2016). The effect of controlled opinion feedback on Delphi features: Mixed messages from a real-world Delphi experiment. *Technological Forecasting and Social Change*, 103, 166–173. <https://doi.org/10.1016/j.techfore.2015.11.008>
- Morrow, S. L. (2005). Quality and trustworthiness in qualitative research in counseling psychology. *Journal of Counseling Psychology*, 52(2), 250–260. <https://doi.org/10.1037/0022-0167.52.2.250>
- Nelms, K. R., & Porter, A. L. (1985). EFTE: An interactive Delphi method. *Technological Forecasting and Social Change*, 28(1), 43–61. [https://doi.org/10.1016/0040-1625\(85\)90072-1](https://doi.org/10.1016/0040-1625(85)90072-1)
- Niederberger, M., & Spranger, J. (2020). Delphi technique in health sciences: A map. *Frontiers in Public Health*, 8, 457. <https://doi.org/10.3389/fpubh.2020.00457>
- Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications Inc.
- Pill, J. (1971). The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Economic Planning Sciences*, 5(1), 57–71. [https://doi.org/10.1016/0038-0121\(71\)90041-3](https://doi.org/10.1016/0038-0121(71)90041-3)
- Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing*, 41(4), 376–382. <https://doi.org/10.1046/j.1365-2648.2003.02537.x>
- Rieger, W. G. (1986). Directions in Delphi developments: Dissertations and their quality. *Technological Forecasting and Social Change*, 29(2), 195–204. [https://doi.org/10.1016/0040-1625\(86\)90063-6](https://doi.org/10.1016/0040-1625(86)90063-6)
- 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). Springer.
- Rowe, G., Wright, G., & Bolger, F. (1991). Delphi: A reevaluation of research and theory. *Technological Forecasting and Social Change*, 39(3), 235–251. [https://doi.org/10.1016/0040-1625\(91\)90039-1](https://doi.org/10.1016/0040-1625(91)90039-1)
- Sackman, H. (1974). *Delphi assessment: Expert opinion, forecasting, and group process*. The RAND Corporation.
- Schmidt, R. C. (1997). Managing Delphi surveys using nonparametric statistical techniques. *Decision Sciences*, 28(3), 763–774. <https://doi.org/10.1111/j.1540-5915.1997.tb01330.x>
- Sinha, I. P., Smyth, R. L., & Williamson, P. R. (2011). Using the Delphi technique to determine which outcomes to measure in clinical trials: Recommendations for the future based on a systematic review of existing studies. *PLoS Medicine*, 8(1), e1000393. <https://doi.org/10.1371/journal.pmed.1000393>
- Spranger, J., Homberg, A., Sonnberger, M., & Niederberger, M. (2022). Reporting guidelines for Delphi techniques in health sciences: A methodological review. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 172, 1–11. <https://doi.org/10.1016/j.zefq.2022.04.025>
- Toma, C., & Picioreanu, I. (2016). The Delphi technique: Methodological considerations and the need for reporting guidelines in medical

- journals. *International Journal of Public Health Research*, 4(6), 47–59.
- Turnbull, A. E., Dinglas, V. D., Friedman, L. A., Chessare, C. M., Sepúlveda, K. A., Bingham, III, C. O., & Needham, D. M. (2018). A survey of Delphi panelists after core outcome set development revealed positive feedback and methods to facilitate panel member participation. *Journal of Clinical Epidemiology*, 102, 99–106. <https://doi.org/10.1016/j.jclinepi.2018.06.007>
- Turoff, M. (1970). The design of a policy Delphi. *Technological Forecasting and Social Change*, 2, 149–171. [https://doi.org/10.1016/0040-1625\(70\)90161-7](https://doi.org/10.1016/0040-1625(70)90161-7)
- Turoff, M. (1971). Delphi conferencing: Computer-based conferencing with anonymity. *Technological Forecasting and Social Change*, 3, 159–204. [https://doi.org/10.1016/S0040-1625\(71\)80012-4](https://doi.org/10.1016/S0040-1625(71)80012-4)
- 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](https://doi.org/10.1016/0040-1625(91)90002-W)
- Yaniv, I., & Milyavsky, M. (2007). Using advice from multiple sources to revise and improve judgments. *Organizational Behavior and Human Decision Processes*, 103(1), 104–120. <https://doi.org/10.1016/j.obhdp.2006.05.006>
- Van Zuuren, E. J., Logullo, P., Price, A., Fedorowicz, Z., Hughes, E. L., & Gattrell, W. T. (2022). Existing guidance on reporting of consensus methodology: A systematic review to inform ACCORD guideline development. *BMJ Open*, 12(9), e065154. <https://doi.org/10.1136/bmjopen-2022-065154>

How to cite this article: Landeta, J., & Lertxundi, A. (2024). Quality indicators for Delphi studies. *Futures & Foresight Science*, 6, e172. <https://doi.org/10.1002/ffo2.172>