Delphi Technique in Health Sciences: A Map

Marlen Niederberger* and Julia Spranger

Department of Research Methods in Health Promotion and Prevention, University of Education Schwaebisch Gmuend, Schwäbisch Gmünd, Germany

Objectives: In health sciences, the Delphi technique is primarily used by researchers when the available knowledge is incomplete or subject to uncertainty and other methods that provide higher levels of evidence cannot be used. The aim is to collect expert-based judgments and often to use them to identify consensus. In this map, we provide an overview of the fields of application for Delphi techniques in health sciences in this map and discuss the processes used and the quality of the findings. We use systematic reviews of Delphi techniques for the map, summarize their findings and examine them from a methodological perspective.

Methods: Twelve systematic reviews of Delphi techniques from different sectors of the health sciences were identified and systematically analyzed.

Results: The 12 systematic reviews show, that Delphi studies are typically carried out in two to three rounds with a deliberately selected panel of experts. A large number of modifications to the Delphi technique have now been developed. Significant weaknesses exist in the quality of the reporting.

Conclusion: Based on the results, there is a need for clarification with regard to the methodological approaches of Delphi techniques, also with respect to any modification. Criteria for evaluating the quality of their execution and reporting also appear to be necessary. However, it should be noted that we cannot make any statements about the quality of execution of the Delphi studies but rather our results are exclusively based on the reported findings of the systematic reviews.

Keywords: Delphi technique, method, health sciences, consensus, systematic review, map, methodological discussion, expert survey

INTRODUCTION

Delphi techniques are used internationally to investigate a wide variety of issues. The aim is to develop an expert-based judgment about an epistemic question. This is based on the assumption that a group of experts and the multitude of associated perspectives will produce a more valid result than a judgment given by an individual expert, even if this expert is the best in his or her field.

The relevance and objectives of Delphi techniques differ between the various disciplines. While Delphi techniques are primarily used in the context of the technical and natural sciences to analyze future developments (1), they are often used in health sciences to find consensus (2). According to the recommendations of the US Agency for Health Care Policy and Research (AHCPR), Delphi techniques are considered to provide the lowest level of evidence for making causal inferences and are thus subordinate to meta-analyses, intervention studies and correlation studies (3).

Nevertheless, Delphi techniques are also highly relevant in health science studies. Based on the findings of Delphi techniques, guidelines or white papers are drafted that act as an important basis for carrying out and evaluating studies or publications (4, 5). Another aspect is that the experts in Delphi studies can draw on various sources of information to make their judgments. On the one hand, they can call on their personal expertise and, on the other hand, they can call on knowledge from other types of studies, e.g., randomized controlled trials or metanalysis (6). This expertise appears to be especially relevant when an experimental design cannot be carried out due to, for example, practical research or ethical reasons. Jorm (2) puts it this way: "The quality of the evidence they produce depends on the inputs available to the experts (e.g. systematic reviews, experiments, qualitative studies, personal experience) and on the methods used to ascertain consensus". Accordingly, there is thus a need from a methodological and epistemological standpoint to investigate Delphi techniques and their epistemic and methodological assumptions in more detail. The following article makes a contribution to this by analyzing systematic reviews of the use of Delphi techniques in the health sciences.

Delphi techniques are structured group communication processes in which complex issues where knowledge is uncertain and incomplete are evaluated by experts using in an iterative process (7, 8). The defining feature is that the aggregated group answers from previous questionnaires are supplied with each new questionnaire, and the experts being questioned are able to reconsider their judgments on this basis, revising them where appropriate. Some authors define Delphi techniques more specifically and focus on reaching consensus between the experts (2, 9, 10). According to Dalkey and Helmer (11), it is a technique designed "to obtain the most reliable consensus of opinion of a group of experts […] by a series of intensive questionnaires interspersed with controlled feedback". However, narrowing the definition to just focus on the concept of consensus hardly seems tenable in view of the wide range of different applications for Delphi techniques. Based on the objectives of the Delphi techniques, Häder distinguished between three other methodological types of Delphi technique besides that of finding consensus: (1) for the aggregation of ideas, (2) for making future predictions, and (3) to determine experts' opinions (7). **Table 1** shows the different types of Delphi technique.

There are also critical arguments against the use of Delphi techniques. In intervention research in health sciences, surveys of experts are considered subordinate to evidence-based methods because they do not take account of any reliable findings on observed cause-effect relationships (12). In addition, Delphi techniques cannot be assigned to any specific paradigm. There are thus no commonly accepted quality criteria (13). From a normative perspective, it is possible to critically question the stability of the judgments, the composition of the expert groups, and the handling of divergent judgments. From a sociological perspective, these techniques raise questions about their validity, the dominance of possible thought collectives, and the reproduction of possible power structures. The focus of reaching consensus increases the risk of reproducing the "habitus mentalis" and possibly failing to take new impetus and

scientific findings sufficiently into account (14). In addition, it is not possible to make an a priori assumption that all groups are "wise" (2, 15). Previous analyses of Delphi techniques have thus indicated that expert judgments differ between different groups (16, 17). The influence of different and independent perspectives on the rationality and appropriateness of the judgments made by an expert group was also emphasized by Surowiecki in his book "Wisdom of crowds" in which he identifies five characteristics of wise groups:

1. Diversity of opinions
2. Independence of opinions
3. Decentralization and specialization of knowledge
4. Aggregation of private judgments into a collective decision
5. Trust and fairness within the group

Delphi techniques are used in health science studies in both medical/natural science and behavioral/social science disciplines (18–22). In the field of medical/natural science, they are used when large-scale observation studies or randomized and controlled clinical studies cannot be carried out due to economic, ethical, or pragmatic research reasons. Delphi techniques have proven useful in the explorative or theoretical phase of the research process because they generate knowledge that can increase the evidence for the desired effect of an intervention—and thus possible insights into its potential effectiveness.

In a behavioral/social science context, Delphi techniques are primarily used for the integration of knowledge. The studies are often prompted by contradictory expertise that generates behavioral uncertainty among consumers and could undermine trust in decision makers (23). The Delphi technique enables the identification of areas of consensus and characterization of areas of disagreement.

The following objectives are typical for Delphi techniques in health sciences:

- Identifying the current state of knowledge (23)
- Improving predictions of possible future circumstances (24, 25)
- Resolving controversial judgments (26)
- Identifying and formulating standards or guidelines for theoretical and methodological issues (2, 4, 27)
- Developing measurement tools and identifying indicators (28)
- Formulating recommendations for action and prioritizing measures (29)

In methodological literature on about Delphi techniques, five characteristics of classical Delphi techniques have been identified: (7, 23)

- Surveying experts who remain anonymous
- Using a standardized questionnaire that can be adapted for every new round of questions
- Determining group answers statistically using univariate analyses
- Anonymous feedback of the results to the participating experts with the opportunity for them to revise their judgments
- One or multiple repetitions of the questionnaire

TABLE 1 | Types of Delphi technique according to Häder (7).

	Type 1: Aggregation of ideas	Type 2: Most precise prediction of an uncertain issue	Type 3: Collecting expert opinions on a diffuse issue	Type 4: Consensus
Approach	Qualitative	Qualitative and quantitative	Primarily quantitative	Quantitative
Goal	Generating and aggregating different ideas and solutions for a problem	Achieving predictions that are as accurate as possible about an unclear issue or uncertain situation	Identifying and displaying the views of a group of experts on a diffuse issue	Establishing the highest possible degree of consensus among the participating experts
Sampling	Selection based on expertise Interdisciplinary composition Tends to have a low number of participants	Clear criteria for identifying experts Usually a very large number of experts (sometimes many thousand, transnational composition)	Full survey or deliberate selection of experts	Selection of experts based on a thematic framework Alongside scientific experts, the participants include interest groups or the interested public

Numerous different variants of Delphi techniques have been developed over the last few years. There are *Realtime Delphi's* in which expert judgments are fed back online in real time (30). In so-called *Delphi Markets*, the Delphi concept is combined with prediction markets and information markets, as well as with the findings of big data research, to improve its forecasting capabilities (31). In a *Policy Delphi*, the aim is to identify the level of dissensus, i.e., the range of the judgments (32). In an *Argumentative Delphi*, the focus is placed on the qualitative justification for the standardized judgments made by the experts (19). In a *Group Delphi technique*, the experts are invited to a joint workshop and can thus give contextual justifications for deviating judgments (23, 33).

The development of new variants has also been accompanied by epistemological and methodological changes to the traditional understanding of the Delphi method. The definition of the term expert has thus been broadened. The definition of an expert is either based on their individual's scientific/professional expertise or lifeworldly experience. Alongside members of certain professions, experts also include patients or users of an intervention (34, 35). The effects that the associated heterogeneous composition of the expert panel may have are quite unclear. However, previous analyses have shown that cognitive diversity in an expert group can support innovative and creative discussion processes and hence are just as important for forming a judgment as the individual abilities and expertise of the experts (36, 37). Hong and Page describe it in a nutshell with the phrase: "Diversity trumps ability" (37).

Ensuring the anonymity of the experts has always remained a constant feature during the evolution of the Delphi technique and the development of methodological variants; the names of the experts involved are only published in exceptional cases (21, 33). From an methodological perspective, some of the new Delphi studies are based on qualitative assumptions. Accordingly, the survey instruments do not only include standardized questionnaires, but also explorative instruments such as open-ended leading questions (38) or workshops (39).

A few methodical tests used to examine the basis of Delphi techniques and their evaluation have been conducted and have also led to contradictory recommendations in some cases (17, 40–44). So a survey of former participants in an international Delphi study in a clinical context demonstrated that up to five rounds of questionnaires was deemed acceptable by the experts (42). However, another study clearly indicated that the experts underestimated the work involved in a Delphi questionnaire, which is why two rounds were recommended (43). The influence of the feedback design with respect to the expert group is also unclear (44). While in one study no significant influence was identified (17), another study showed it does have an influence or its influence was considered unclear (42).

MATERIALS AND METHODS

The results of the systematic reviews of Delphi techniques in health sciences are summarized below. The presentation of the reviews is largely based on the PRISMA statement (45). As we did not complete our own systematic review but rather have created a map with a methodological focus, it was not possible to apply all of the topics and proposed analyses included therein. The following were not applicable: protocol, registration and additional analyses (e.g., sensitivity or subgroup analyses, summary measures, meta-regression).

The process for identifying systematic reviews of Delphi techniques in health sciences was carried out in 2019 using a search of the databases *PubMed [include Journal/Author Name Estimator (JANE)]* with the keywords "review" and "delphi" without any restriction placed on the year of publication. The abstracts for the articles identified were each read by one person and a decision about whether to include them in the analysis was made on the basis of inclusion and exclusion criteria. In addition, we searched the identified reviews for other possible studies and then investigated any articles that came into question and examined the abstracts of the articles.

The inclusion criteria for the reviews were that they were designed as systematic reviews, written in German or English, and were available as a full-text version. The contents of the articles included in the review also had to be based on Delphi techniques used in health sciences in general or in a subdiscipline. Articles not designed as a systematic review, whose contents did not involve the health sector or articles exclusively focused on a specific Delphi modification (e.g., Policy Delphi) were excluded.

We developed abductive categories in order to analyze the reviews (46, 47). These categories were based on constituent characteristics of Delphi techniques and guidelines on ensuring

the quality of the reporting of Delphi studies (cf. CREDES [Guidance on Conducting and REporting DElphi Studies]) (48). For each category, we focused on the publication practice and the findings reported in the reviews. The reviews were evaluated on the basis of a qualitative analysis of their contents with the aim of determining the contextual scope and also the frequency of the categories to some extent (**Table 2**).

The evaluation of the systematic reviews was carried out by two researchers who consulted with one another in the event of any uncertainty.

The category system presented above was the basis for the qualitative content analysis of the systematic reviews of Delphi techniques in health sciences. The results are presented below.

RESULTS

A total of 16 reviews from 1998 to 2019 were identified (**Supplementary Table 1**). Four were excluded from the analysis due to the stated criteria (AID13-AID16). Twelve reviews satisfied the inclusion criteria. In the articles, the authors investigated Delphi techniques used in the themes of health and well-being (ID5), in health care (ID2), palliative care (ID9), training in radiography (ID8), the care sector (ID4), health promotion (ID11), health reporting (ID1), the clinical sector (ID12), and medical education (ID6), as well as Delphi techniques used in the health sciences in general (ID3, ID7, ID10). The number of Delphi studies reviewed varied from 10 (ID2, ID7) to 257 (ID6). Overall, Delphi techniques from 1950 (ID10) through to 2016 (ID6, ID11) were included in the analysis. This means that this analysis is based on data accumulated from 883 Delphi techniques over a period of six decades. At the same time, this means that the following results cover a large period of time, even if more modern Delphi studies are more frequently represented. For example, six of the systematic review exclusively focused on Delphi studies carried out in the 2000s or even later (ID2, ID3, ID4, ID5, ID6, ID11).

The focus of the analysis in some reviews was explicitly placed on consensus Delphi techniques (ID3, ID4, ID5). In other reviews, the analysis covered all identified Delphi techniques irrespective of their objectives.

Category 1: Delphi Variants
An overview of the results of the analysis into the Delphi variants can be found in **Supplementary Table 2**.

Category 1.a: Reporting Quality
A specific definition of the underlying Delphi technique was found in 61% (ID11) and 88.2% (ID4) of the Delphi articles investigated.

Category 1.b: Delphi Variants
Nine of the reviews included an investigation of which Delphi variants had been used (ID1, ID2, ID3, ID4, ID6, ID7, ID9, ID11, ID12). Classical Delphi techniques were mostly dominant (ID4 classical 69.7%, ID11 76%). In other reviews, the authors mainly found modified techniques (e.g., ID1 62.8%). However, it should be noted that a differentiation between classical and modified was

only possible to a limited extent due to diverging or imprecise definitions. For example, the authors of one review included online questionnaires as a classical variant (ID11), while this was unclear in other reviews (e.g., ID1). In the articles investigated in the reviews, modifications included, for example, personal meetings of the experts (ID1), a combination of quantitative and qualitative data (ID5), or if different expert panels for each Delphi round were used (ID9). In some cases, modifications had been made to the Delphi techniques without describing them as such, and other studies did not include a specification of what adaptations had been made (ID9).

Category 2: Experts
A detailed presentation of the results for each review for the experts category can be found in **Supplementary Table 3**.

Category 2.a: Reporting Quality
Most of the Delphi studies analyzed in the reviews reported on the number of participating experts. The rates for the initial round were between 84% (ID6) and 100% (ID12). Four of the reviews investigated whether the number of experts was stated for each round (ID4, ID7, ID11, ID12). In one review based on 10 Delphi studies from health sciences (ID7), the authors discovered that the number of experts per round was stated in all articles. A review of 48 studies in a medical context indicated that the number of invited experts was stated less frequently with each round (ID6).

Seven of the 12 reviews investigated whether the backgrounds of the experts had been reported, what kind of expertise they possessed, and the criteria according to which they were selected (ID1, ID3, ID4, ID6, ID9, ID11, ID12). One review of Delphi techniques in a health context determined that the criteria for selecting the experts was reproduced in 65 of 100 articles (65%) (ID3) included in that particular review. In other reviews with a more specific focus, such as on health care, palliative medicine, or health promotion, the rates were higher at 69% (ID11), 70% (ID9) and 79% (ID1), respectively.

Based on the results of the reviews, the criteria by which the experts were selected and approached was not always clear. In one review of 100 studies from the care sector, the proportion of articles with unclear selection criteria was 11.2% (ID4), while the proportion was 93.3% in a review of 15 studies from the clinical sector (ID12).

Category 2.b: Number of Experts
Seven of the 12 review authors investigated whether the number of experts was stated in the analyzed articles (ID1, ID3, ID4, ID6, ID9, ID11, ID12). In this process, the authors of the reviews investigated the number of experts at different points in the Delphi process: at the beginning of the Delphi (ID6), in the last round (ID3) or at different points in time (ID11).

The number of experts included varied in the Delphi studies investigated in the reviews from three (ID1) to 731 experts (ID11). The average number of experts included was usually in the low to medium double-digit range (e.g., ID1: median = 17 invited experts; ID11: mean = 40 experts in the first Delphi round). Two reviews indicated the number of participants was

TABLE 2 | Overview of the categories.

Main category	Subcategory	Questions
1. Delphi variants variants of Delphi technique	1.a Reporting quality	• How often did the investigated articles of the reviews contain a report of the Delphi variants?
	1.b Variants	• Which Delphi variants were used in the studies examined in the review?
2. Experts	2.a Reporting quality	• How often did the investigated articles of the reviews include a report on aspects related to the definition and selection of experts?
	2.b Number of experts	• How many experts were included in the Delphi technique? • How did the number of experts change over the different rounds?
	2.c Selection of experts	• Which categories of experts were included in the Delphi technique? • How were the experts identified and approached?
	2.d Expert panel	• How was the expert panel put together?
3. Consensus (Delphi studies that were explicitly designed as consensus methods were included here)	3.a Reporting quality	• How often did the articles contain a report about consensus aspects?
	3.b Definition and measurement of consensus	• How was consensus defined and measured? • What role did the stability of the answers play?
	3.c Consensus reached	• For how many items in the Delphi studies was consensus reached?
4. Delphi process	4.a Reporting quality	• How often did the investigated articles of the reviews contain a report of the specific process used?
	4.b Number of rounds	• How many rounds were used in the Delphi technique?
	4.c Questionnaire and scale development	• How were the questionnaires and the specific items for a Delphi technique developed?
	4.d Response rate	• How high was the response rate from the experts both when initially approached and also for the individual rounds?
	4.e Feedback design	• How was the feedback designed?

higher than 100 experts in five of 100 articles (ID3) and two of 15 articles (ID12).

Category 2.c: Selection of Experts

The most commonly stated selection criteria for the experts in the investigated Delphi articles were organizational or institutional affiliation, recommendation by third parties, or the experience of the experts (measured in years) (ID1, ID4, ID9, ID11, ID12). Academic factors such as academic title or number of publications (ID9 22%), or geographical aspects (ID9 43.3%), also played a role in the composition of the expert panel. Identification of the experts was mostly based on multiple criteria (ID4 23%, ID11 28.6%).

Overall, the authors of the reviews indicated the experts were deliberately approached by the researchers (e.g., ID4, ID11) and their selection was not verified by a self-evaluation (ID11). Random selection of the experts remained an exception (ID4, ID11).

Category 2.d: Expert Panel

In seven reviews, there was a systematic investigation of the expert panels (ID1, ID4, ID6, ID7, ID9, ID11, ID12). A heterogeneous composition was identified in most cases. The Delphi studies included professionals from the health sector, scientists, managers, and representatives of specific organizations.

Patients were also included in some Delphi studies. The number of such Delphi studies was between 2% (ID4) and 27% (ID12). According to the information provided in the

investigated articles, the inclusion of those affected and involved increased the quality of the process (ID1, ID12).

Category 3: Consensus

In the various reviews, questions about the definition and presentation of consensus were investigated in detail (cf. **Supplementary Table 4**).

Category 3.a: Reporting Quality

Seven of the 12 reviews determined whether and when consensus was defined in the Delphi studies (ID1, ID3, ID4, ID6, ID9, ID11, ID12). The number of studies in which consensus was defined in the article was between 73.5% (ID3) and 83.3% (ID9) in the reviews.

Category 3.b: Definition and Measurement

The definition of consensus was mostly defined a priori. In a review of 100 Delphi studies (ID3), for example, 88.9% of the authors defined consensus in advance of development of the questionnaire. The proportion in other reviews was in the medium range (ID4 44.9%, ID6 43.2%, ID12 46.7%).

The results of the reviews demonstrate that different definitions and measurements for reaching consensus were used. In one review, the authors identified 11 different statistical definitions for consensus (ID3).

Consensus was usually measured in the Delphi studies using percent agreement, units of central tendency (especially the median), or a combination of percent agreement within a certain range and for a certain threshold (mostly the median) (ID1, ID3,

ID4, ID6, ID9, ID11). Likert type scales ranging from 3 to 10 points (ID4, ID1) were used, whereby five- or nine-point scales were the most common (ID9, ID11).

In particular, the definition using agreement that exceeded a certain percentage value was used in the Delphi studies investigated in the reviews (ID1 14.5%, ID3 34.7%, ID11 42.2%). The cut-off value, meaning the value from which consensus was assumed, varied between 20 and 100% agreement in one review (ID6). However, a threshold of 60% (ID4) or higher (ID3, ID9, ID6) was identified in most cases.

In on review, it was discovered that qualitative aspects tended to be used to define consensus in modified Delphi techniques (ID8).

According to the findings in the reviews, the stability of the judgments did not play a central role in the Delphi articles. In a review from the health care sector, one study was found that specified the stability of the judgments (ID1). The authors of the review on palliative care study identified two such studies (ID9).

Category 3.c: Consensus Reached
The question of whether consensus was reached in the Delphi studies was rarely a theme in the reviews. However, the authors of the reviews did indicate that consensus had been reached for most of the items on a questionnaire but not in all cases (ID3, ID11). In one review in health promotion, the authors discovered that on average consensus had been reached between the experts for more than 60% of the items (ID11). The level of consensus on items was influenced by how consensus was defined and the composition of the expert panel (ID11).

Category 4: Delphi Process
An overview of the individual results of the analysis of the Delphi process in each review is presented in **Supplementary Table 5**.

Category 4.a: Reporting Quality
The authors of seven reviews investigated whether the number of Delphi rounds was published (ID1, ID3, ID4, ID6, ID9, ID11, ID12). The number of Delphi rounds was stated in most of the Delphi studies (e.g., ID1 82.5%, ID4 91%, ID6 100%, ID9 49.3%, ID12 93.3%).

Six of the reviews included a report of the generation of the questionnaire (ID1, ID4, ID6, ID9, ID11, ID12). They demonstrated that up to 96.3% of the investigated articles reported on how the items for the questionnaire were developed (ID1). In contrast, this rate stood at 33.3% in the review of palliative care articles (ID9).

The authors of two reviews investigated the question of how the items were changed during the Delphi process based on the judgments submitted by the experts (ID3, ID12). In one of the reviews, the authors indicated that 59% of the analyzed articles had defined criteria for dropping items (ID3). In another review, the authors stated that all of the investigated Delphi studies included a report of "what was asked in each round" (ID12, p. 2).

The authors of the reviews reported about the feedback in most of the Delphi studies (ID11 67.9%, ID12 93.3%). The information provided about the response rate per Delphi round was less (ID1 and ID4 39%). According to the results of the

reviews, around half of the studies did not provide information about the feedback design between the Delphi rounds (ID1 40%, ID4 55.1%, ID6 37.7% ID12 40%).

According to the authors of the review on health promotion, the process—from formulating the issue being investigated through to the development of the questionnaire—was in general similar to a "black box," and the methodological quality of the survey instrument was almost impossible to evaluate using the published information (ID11, p. 318).

Category 4.b: Number of Rounds
The number of Delphi rounds varied relatively widely according to the findings presented in the reviews. The largest range of 0 to 14 rounds was identified by the authors of the review in a medical context (ID6). However, the authors did not state the specific research contexts for the extremes of 0 and 14 rounds. In the other reviews, the ranges were between 2 and 5 (ID11), 2 and 6 (ID12), 1 and at least 5 (ID3) or 1 and 5 rounds (ID9). The most common number of rounds in the Delphi process was two or three rounds (ID3, ID6, ID9, ID10, ID4, ID11, ID12).

In one review, the authors discovered that the range for the number of rounds in modified Delphi techniques was larger than for classical Delphi techniques (ID1). At the same time, the median number of rounds was lower than for classical Delphi techniques (ID1, ID4).

There was no further discussion on the reasons for the number of rounds. The authors of one review determined that the number of rounds was defined in advance in 18.3% of 257 Delphi studies (ID6). In the other studies, this was either not clearly explained or was decided post priori.

Category 4.c: Development of the Questionnaire
The items for a Delphi questionnaire were developed by the Delphi users based on literature on relevant subject matter in most of the studies investigated (ID1, ID4, ID6, ID11). The proportions ranged between 35.7% (ID11) and 70% (ID6). In some cases, the items were also identified from empirical analyses such as qualitative interviews or focus groups that were completed in advance or were taken from existing guidelines (ID1, ID4, ID11). The first (qualitative) round of questions in the Delphi process was also sometimes used to generate the items for a standardized questionnaire (ID4).

The specific development of the questionnaire during the Delphi process was rarely discussed in the reviews. A review of palliative care studies demonstrated that new items were developed (33.3%), items were modified (20%), and items were deleted (30%) during the Delphi process (ID9).

Category 4.d: Response Rate
The response rate was investigated in five of the reviews examined (ID1, ID4, ID8, ID11, ID12). In one review, a median for the response rate of 87% in the first round and 90% in the last round was determined for classical Delphis on the subject of health care (ID1). In the case of modified Delphi techniques, the median in the first round was a little higher at 92% and a little lower in the final round at 87% (ID1).

The authors of the review of health promotion studies also identified high response rates (ID11). Based on the number of invited experts in each case, the average response rate was 72% in the first round, 83% in the second wave, and 89% in the third wave. In a review on the subject of radiography, the authors identified a Delphi study with an increasing number of participants (ID8).

Category 4.e: Feedback Design

The authors of six reviews reported findings related to the feedback design (ID1, ID4, ID6, ID9, ID11, ID12). In most of the Delphi studies investigated, the researchers provided group feedback and less frequently individual feedback (ID1, ID4, ID11, ID12). As indicated in one review, the experts received no feedback at all in 9% of the Delphi studies investigated (ID4). This review showed that group feedback was provided less frequently for modified Delphi techniques than for classical Delphi techniques (ID4).

If data about feedback was published, the studies mostly contained a report of quantitative statistical feedback and less frequently a combination of quantitative and qualitative results or purely qualitative findings (ID1 quantitative 58.3%, quantitative and qualitative 39.6% and qualitative 2.1%, ID9 quantitative 36.7%, qualitative 26.7%, ID12 quantitative 53.3% and qualitative 26.7%).

DISCUSSION

By examining all of the results, it was possible to identify the following aspects of Delphi techniques in health sciences:

1. *There is no uniform definition for consensus.* Values in the various Delphi reviews varied, they generally showed that the proportion of definitions for consensus made a priori and the number where the definition of consensus was not or was unclearly reported were high. The appropriateness of the theoretical measurements and the possible consequences associated with using one or another definition for consensus were not discussed. There was little consideration of possible factors that may have influenced whether consensus was reached (18, 49).

2. The various *fields of application* demonstrated that although Delphi techniques are used, new variants such as Realtime Delphis are seldom found in the health sciences. Instead the authors from the reviews concluded (ID9, ID11), there appears to be a large number of less specific *modifications of Delphi techniques* for which it is barely possible or even impossible to understand the epistemic objectives and the research process using the publications.

3. The specific characteristics used to *identify types of experts* or the effect of taking account of evidence-based and lifeworldly expertise on the group communication process are not discussed. This appears to be important, especially when integrating patients into the studies, which is something that is generally being increasingly promoted and implemented in research. Studies have shown that this adds value because it enables insights that cannot be gained using other research designs (50). There was also no discussion of any validation of the expert possessing the attributed expertise. Some reflection on the different types of knowledge and the associated linking of assumed expertise to the issue being investigated would appear to be especially relevant for the significance of the results of a Delphi process.

4. Although the number of experts included in the studies varied, it was mostly in the low double-digit range. This number raises questions about the validity of the findings. The idea of collective intelligence [based on the "wisdom of the many" (15)], as used primarily in Delphi studies for making predictions, does not apply for such small numbers of experts. Instead, it raises the question of whether all relevant perspectives and scientific disciplines have been appropriately taken into account. Moreover, the effect that very small numbers may have on the risk of accumulating certain thought collectives to the detriment of peripheral concepts is unclear. The low number of experts is perhaps also an issue for reliability (51, 52). Previous analyses have demonstrated that the reliability of the Delphi technique can be highly diverse and also dependent on the number of participants.

5. The items for the Delphi questionnaires are usually taken from literature relevant to the subject matter, or collected during interviews or focus groups carried out in advance. However, there is little information published about the *process for developing and monitoring the questionnaires.* It is thus very difficult to evaluate the methodological quality of the survey instruments.

1. The *number of rounds* is interpreted here as a methodological rule or is defined based on pragmatic research arguments. It is only defined as an epistemological variable in exceptional cases. That means that many Delphi studies stopped the survey process for a certain projection when a predefined level of agreement, i.e., consensus, was achieved (AID14). This is connected to the fact that the stability of the individual or group judgments was rarely discussed in the Delphi studies included in these 12 review articles. This also has an effect on the critical reflection of the interrater and intrarater reliability, which could also not be examined in the reviews due to the lack of information in the primary articles [ID1 (53)].

2. The results of Delphi techniques are often presented on the basis of *consensus judgments.* Yet depending on how consensus was defined, up to 40% of the experts do not agree with the consensus. The identity of these experts and the judgments they have made remains unclear. There is a risk that relevant and unusual judgments will be neglected. In addition, there is no reflection on the possible reasons for dissensus.

3. The authors of some of the reviews identified *irregularities in the design and statistical analysis* of Delphi techniques. For example, one-off surveys of experts or preliminary studies are sometimes described as Delphi techniques (ID1, ID3, ID6, ID9). It is questionable whether these types of studies can be described as Delphi techniques. Furthermore, errors in the statistical analyses were discovered that were often associated with the measurement level being disregarded (AID14).

The findings in the reviews we analyzed indicated that there is no uniform process for carrying out and reporting Delphi techniques. In this context, recommendations such as those made by Hasson and Keeney (54) or by Jünger et al. (48) with CREDES are not taken into account.

LIMITATIONS

This analysis essentially compares apples with oranges because the reviews of the Delphi techniques focus on very diverse themes and questions. In addition, our analysis exclusively considered reviews of publications and we did not read the original literature. Therefore, we have only analyzed in this article what was reported in the reviews. There is thus a clear danger that we have replicated the limitations of the systematic reviews. However, the systematic collation of the reviews has allowed us to overcome gaps in the content of the individual reviews and ensures that this map provides a comprehensive picture of the application of Delphi techniques in health sciences.

We were also only able to analyze reviews written in German and English. Although the results provided us with insights into the research practices used for Delphi techniques, we do not claim these insights to be complete or representative in any way.

CONCLUSION

Following a critical examination of publication practice for Delphi techniques, Humphrey-Murto and de Wit (2018) reached the following conclusion: "More research please" (55). Our results also indicate deficits both in carrying out and also reporting Delphi techniques. In conclusion, we would like to highlight the lack of an epistemological and methodological basis for Delphi techniques (54). In terms of the main categories examined in this article, we believe that there is a need for further research and discussion, especially of a methodological nature, in the following areas:

- **Delphi variants:** There are a series of Delphi variants distinguished in the methodological discussion that seldom appear to be applied in health sciences. The use of these variants could generate contextual and methodological value. For example, a Group Delphi technique would enable the collection of contextual justifications for dissensus and a Realtime Delphi would make it possible to analyze the response latency of experts.
- **Experts:** Cognitive diversity in the composition of the expert panel is important for the robustness and validity of the findings. In preparation for a Delphi process, a fundamental

system analysis is thus required in order to identify all relevant groups of actors, scientific disciplines, and perspectives and to invite appropriate representatives or, if possible, all experts to participate in the Delphi process at an early stage. Diversity can have a decisive influence on the quality of the data and on whether the judgments are accepted and considered feasible later on, especially if the number of experts is rather low.

- **Consensus:** Identifying consensus amongst experts appears to be the central motivation for the application of Delphi techniques in health sciences. However, there is no general definition for what consensus actually is. In addition, there seems to be no discussion about which experts are in consensus and which are not. Possible distortions that may, for example, favor certain groups of experts, thus remain concealed. The Delphi techniques also do not usually allow any statements to be made about the stability of the judgments. This appears to be particularly virulent if the results lead to the publication of guidelines, definitions, or white papers, which often act as the basis for health research, medical practice, and diagnostics for many years.
- **Delphi process:** A Delphi process is a complex and challenging process that is now carried out using numerous different variations. Nevertheless, it is important to precisely describe, justify, and methodologically reflect on any modifications. This would increase the transparency of the findings. It is also necessary to discuss the critical and rationalistic criteria for the validity and reliability of the studies and the more constructivist characteristics of credibility, transparency, and transferability.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

MN: search literature, concept of the article, analyze and interpretation, and write the article. JS: support data analysis and formal aspects. All authors contributed to the article and approved the submitted version.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2020.00457/full#supplementary-material

REFERENCES

1. Cuhls K, Kayser V, Grandt S, Hamm U, Reisch L, Daniel H, et al. *Global Visions for the Bioeconomy - An International Delphi-Study*. Berlin: Bioökonomierat (2015).
2. Jorm AF. Using the Delphi expert consensus method in mental health research. *Aust N Z J Psychiatry*. (2015) 49:887–97. doi: 10.1177/0004867415600891

3. US Department of Health and Human Services, Public Health Services, Agency for Health Care Policy and Research. *Acute Pain Management: Operative or Medical Procedures and Trauma*. Rockville, MD (1992).
4. Negrini S, Armijo-Olivo S, Patrini M, Frontera WR, Heinemann AW, Machalicek W, et al. The randomized controlled trials rehabilitation checklist: methodology of development of a reporting guideline specific to rehabilitation. *Am J Phys Med Rehabil*. (2020) 99:210–5. doi: 10.1097/PHM.0000000000001370

5. Centeno C, Sitte T, De Lima L, Alsirafy S, Bruera E, Callaway M, et al. White paper for global palliative care advocacy: recommendations from a PAL-LIFE expert advisory group of the pontifical academy for life, Vatican City. *J Palliat Med.* (2018) 21:1389–97. doi: 10.1089/jpm.2018.0248

6. Morgan AJ, Jorm AF. Self-help strategies that are helpful for sub-threshold depression: a Delphi consensus study. *J Affect Disord.* (2009) 115:196–200. doi: 10.1016/j.jad.2008.08.004

7. Häder M. *Delphi-Befragungen: Ein Arbeitsbuch.* Wiesbaden: Springer VS (2014). p. 244.

8. Linstone HA, Turoff M, Helmer O, editors. *The Delphi Method: Techniques and applications.* Reading, MA: Addison-Wesley (2002). p. 620.

9. Ab Latif R, Mohamed R, Dahlan A, Mat Nor MZ. Using Delphi technique: making sense of consensus in concept mapping structure and multiple choice questions (MCQ). *EIMJ.* (2016) 8:89–98. doi: 10.5959/eimj.v8i3.421

10. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assess Res Eval.* (2007) 12:1–8. doi: 10.7275/pdz9-th90

11. Dalkey N, Helmer O. An Experimental application of the DELPHI method to the use of experts. *Manage Sci.* (1963) 9:458–67. doi: 10.1287/mnsc.9.3.458

12. Bödeker W, Bundesverband BKK. Evidenzbasierung in gesundheitsförderung und prävention: der wunsch nach legitimation und das problem der nachweisstrenge. motive und obstakel der suche nach evidenz. *Präv Zeitschr Gesundheitsförderung* (2007) 3:1–7.

13. Guzys D, Dickson-Swift V, Kenny A, Threlkeld G. Gadamerian philosophical hermeneutics as a useful methodological framework for the Delphi technique. *Int J Qual Stud Health Well Being.* (2015) 10:26291. doi: 10.3402/qhw.v10.26291

14. Scheele DS. "Reality construction as a product of Delphi interaction," In: Linstone HA, Turoff M, Helmer O, editors. *The Delphi Method: Techniques and Applications.* Reading, MA: Addison-Wesley (2002). p. 35–67.

15. Surowiecki J. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations.* New York, NY: Doubleday (2004). p. 296.

16. Campbell SM. How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? *Qual Saf Health Care.* (2004) 13:428–34. doi: 10.1136/qshc.2003.007815

17. MacLennan S, Kirkham J, Lam TB, Williamson PR. A randomized trial comparing three Delphi feedback strategies found no evidence of a difference in a setting with high initial agreement. *J Clin Epidemiol.* (2018) 93:1–8. doi: 10.1016/j.jclinepi.2017.09.024

18. Hart LM, Jorm AF, Kanowski LG, Kelly CM, Langlands RL. Mental health first aid for Indigenous Australians: using Delphi consensus studies to develop guidelines for culturally appropriate responses to mental health problems. *BMC Psychiatry.* (2009) 9:47. doi: 10.1186/1471-244X-9-47

19. Keeney S, Hasson F, McKenna H. *The Delphi Technique in Nursing and Health Research.* Oxford: Wiley-Blackwell (2011).

20. De Meyrick J. *The Delphi Method and Health Research.* Sydney, NSW: Macquarie University Department of Business (2001). p. 13.

21. Schmitt J, Petzold T, Nellessen-Martens G, Pfaff H. Priorisierung und konsentierung von begutachtungs-, förder- und evaluationskriterien für projekte aus dem innovationsfonds: eine multiperspektivische Delphi-studie. *Das Gesundheitswesen.* (2015) 77:570–9. doi: 10.1055/s-0035-1555898

22. De Villiers MR, De Villiers PJ, Kent AP. The Delphi technique in health sciences education research. *Med Teach.* (2005) 27:639–43. doi: 10.1080/13611260500069947

23. Niederberger M, Renn O. *Das Gruppendelphi-Verfahren: Vom Konzept bis zur Anwendung.* Wiesbaden: Springer VS (2018). p. 207.

24. Cuhls K, Blind K, Grupp H, Bradke H, Dreher C, Harmsen D-M, et al. *DELPHI, 98 Umfrage. Studie zur Globalen Entwicklung Von Wissenschaft und Technik.* Karlsruhe: Fraunhofer-Institut für Systemtechnik und Innovationsforschung (ISI) (1998).

25. Kanama D, Kondo A, Yokoo Y. Development of technology foresight: integration of technology roadmapping and the Delphi method. *IJTIP.* (2008) 4:184. doi: 10.1504/IJTIP.2008.018316

26. Zwick MM, Schröter R. Wirksame prävention? Ergebnisse eines expertendelphi. In: Zwick MM, Deuschle J, Renn O, editors. *Übergewicht und Adipositas bei Kindern und Jugendlichen.* Wiesbaden: VS Verl. für Sozialwiss (2011). p. 239–59.

27. Jünger S, Payne S, Brearley S, Ploenes V, Radbruch L. Consensus building in palliative care: a Europe-wide delphi study on common understandings and conceptual differences. *J Pain Symptom Manage.* (2012) 44:192–205. doi: 10.1016/j.jpainsymman.2011.09.009

28. Han H, Ahn DH, Song J, Hwang TY, Roh S. Development of Mental Health Indicators in Korea. *Psychiatry Investig.* (2012) 9:311–8. doi: 10.4306/pi.2012.9.4.311

29. Van Hasselt FM, Oud MJ, Loonen AJ. Practical recommendations for improvement of the physical health care of patients with severe mental illness. *Acta Psychiatr Scand.* (2015) 131:387–96. doi: 10.1111/acps.12372

30. Aengenheyster S, Cuhls K, Gerhold L, Heiskanen-Schüttler M, Huck J, Muszynska M. Real-time Delphi in practice — a comparative analysis of existing software-based tools. *Technol Forecasting Soc Change.* (2017) 118:15–27. doi: 10.1016/j.techfore.2017.01.023

31. Servan-Schreiber E. Prediction Markets. In: Landemore H, Elster J, editors. *Collective Wisdom: Principles and Mechanisms.* Cambridge: Cambridge University Press (2012). p. 21–37.

32. Turoff M. The Design of a Policy Delphi. *Technol Forecast Soc Change.* (1970) 2:149–71. doi: 10.1016/0040-1625(70)90161-7

33. Ives J, Dunn M, Molewijk B, Schildmann J, Bærøe K, Frith L, et al. Standards of practice in empirical bioethics research: towards a consensus. *BMC Med Ethics.* (2018) 19:68. doi: 10.1186/s12910-018-0304-3

34. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG, et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. *Ann Rheum Dis.* (2013) 72:1125–35. doi: 10.1136/annrheumdis-2012-202745

35. Guzman J, Tompa E, Koehoorn M, de Boer H, Macdonald S, Alamgir H. Economic evaluation of occupational health and safety programmes in health care. *Occup Med.* (2015) 65:590–7. doi: 10.1093/occmed/kqv114

36. Page SE. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies.* Oxford: Princeton University Press. (2008). p. 455.

37. Hong L, Page SE. Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proc Natl Acad Sci USA.* (2004) 101:16385–9. doi: 10.1073/pnas.0403723101

38. Kelly M, Wills J, Jester R, Speller V. Should nurses be role models for healthy lifestyles? Results from a modified Delphi study. *J Adv Nurs.* (2017) 73:665–78. doi: 10.1111/jan.13173

39. Teyhen DS, Aldag M, Edinborough E, Ghannadian JD, Haught A, Kinn J, et al. Leveraging technology: creating and sustaining changes for health. *Telemed J E Health.* (2014) 20:835–49. doi: 10.1089/tmj.2013.0328

40. Holey EA, Feeley JL, Dixon J, Whittaker VJ. An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC Med Res Methodol.* (2007) 7:52. doi: 10.1186/1471-2288-7-52

41. Birko S, Dove ES, Özdemir V. Evaluation of nine consensus indices in Delphi foresight research and their dependency on Delphi survey characteristics: a simulation study and debate on delphi design and interpretation. *PLoS ONE.* (2015) 10:e0135162. doi: 10.1371/journal.pone.0135162

42. Turnbull AE, Dinglas VD, Friedman LA, Chessare CM, Sepúlveda KA, Bingham CO, et al. A survey of Delphi panelists after core outcome set development revealed positive feedback and methods to facilitate panel member participation. *J Clin Epidemiol.* (2018) 102:99–106. doi: 10.1016/j.jclinepi.2018.06.007

43. Hall DA, Smith H, Heffernan E, Fackrell K. Recruiting and retaining participants in e-Delphi surveys for core outcome set development: evaluating the COMiT'ID study. *PLoS ONE.* (2018) 13:e0201378. doi: 10.1371/journal.pone.0201378

44. Brookes ST, Macefield RC, Williamson PR, McNair AG, Potter S, Blencowe NS, et al. Three nested randomized controlled trials of peer-only or multiple stakeholder group feedback within Delphi surveys during core outcome and information set development. *Trials.* (2016) 17:409. doi: 10.1186/s13063-016-1479-x

45. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* (2009) 6:e1000097. doi: 10.1371/journal.pmed.1000097

46. Eriksson K, Lindström UA. Abduction—a way to deeper understanding of the world of caring. *Scand J Caring Sci.* (1997) 11:195–8. doi: 10.1111/j.1471-6712.1997.tb00455.x

47. Peirce CS. *Writings of Charles S. Peirce: A Chronological Edition*. Vol. 5. Bloomington, IN: Indiana University Press (1982).

48. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and REporting DElphi Studies (CREDES) in palliative care: recommendations based on a methodological systematic review. *Palliat Med.* (2017) 31:684–706. doi: 10.1177/0269216317690685

49. Vidgen HA, Gallegos D. *What is Food Literacy and Does it Influence What We Eat: A Study of Australian Food Experts*. Brisbane, QLD: Queensland University of Technology (2011).

50. Serrano-Aguilar P, Trujillo-Martin MM, Ramos-Goni JM, Mahtani-Chugani V, Perestelo-Perez L, La Posada-de Paz M. Patient involvement in health research: a contribution to a systematic review on the effectiveness of treatments for degenerative ataxias. *Soc Sci Med.* (2009) 69:920–5. doi: 10.1016/j.socscimed.2009.07.005

51. Tomasik T. Reliability and validity of the Delphi method in guideline development for family physicians. *Qual Prim Care.* (2010) 18:317–26.

52. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, et al. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess.* (1998) 2:i–iv, 1–88. doi: 10.3310/hta2030

53. de Loë RC, Melnychuk N, Murray D, Plummer R. Advancing the state of policy Delphi practice: a systematic review evaluating methodological evolution, innovation, and opportunities. *Technol Forecast Soc Change.* (2016) 104:78–88. doi: 10.1016/j.techfore.2015.12.009

54. Hasson F, Keeney S. Enhancing rigour in the Delphi technique research. *Technol Forecast Soc Change.* (2011) 78:1695–704. doi: 10.1016/j.techfore.2011.04.005

55. Humphrey-Murto S, De Wit M. The Delphi method-more research please. *J Clin Epidemiol.* (2018) 106:136–9. doi: 10.1016/j.jclinepi.2018.10.011