Research article

Conducting and reporting the Delphi method in traditional Chinese medicine syndrome diagnosis research: A cross-sectional analysis

Xinyi Shi [a], Xudong He [b], Qiang Liu [c], Luda Feng [d], Yixuan Li [a], Xuebin Zhang [a], Zixin Cheng [a], Chi Zhang [a,*], Ying Gao [a,e,**]

[a] *Dongzhimen Hospital, Beijing University of Chinese Medicine, Beijing, China*
[b] *School of Nursing, Beijing University of Chinese Medicine, Beijing, China*
[c] *Center for Evidence-based Medicine, World Federation of Chinese Medicine Societies, Beijing, China*
[d] *Dongfang Hospital, Beijing University of Chinese Medicine, Beijing, China*
[e] *Institute for Brain Disorders, Beijing University of Chinese Medicine, Beijing, China*

ARTICLE INFO

ABSTRACT

Background: The Delphi method has been extensively used to reach a consensus in traditional Chinese medicine (TCM) syndrome diagnosis research when subjective judgment is not uniform and objective evidence is lacking. The conduct and reporting of the Delphi method in TCM syndrome diagnosis research have never been critiqued. Our study aims to explore the consistency of using this technique and assess the reporting quality.
Methods: A cross-sectional study was employed to scope articles reporting the conduct of the Delphi method in TCM syndrome diagnosis research. We searched the PubMed, Web of Science, CNKI, VIP, Wanfang and SinoMed databases with the restriction of Chinese and English language from their inception to March 25, 2023. A standardized extraction form was designed to collect demographics and methodological processes reflecting the rigor and transparency in TCM syndrome diagnosis research.
Results: A total of 1832 studies were screened, and 50 were included. The median number of panels was 30 (IQR 20–34.5) and only 12 (24.0 %) studies were with a heterogeneous sample of panels. Two rounds was most common (37/50; 74.0 %), followed by three (7/50; 14.0 %), and only 13 (26.0 %) studies determined the number of rounds a priori. The reporting quality varied, with 18.0 % (9/50) reporting anonymity, 30.0 % (15/50) describing the controlled feedback, 20.0 % (10/50) reporting the procedure duration (7.14 ± 3.29 months) and 26.0 % (13/50) predefining the consensus.
Conclusion: The Delphi method is inconsistently conducted and nontransparently reported in TCM syndrome diagnosis research. Standardized criteria are urgently needed for best practices in future research.

---

* Corresponding author. Dongzhimen Hospital, Beijing University of Chinese Medicine, 5 Haiyuncang Street, Dongcheng District, Beijing, 100070, China.

** Corresponding author. Institute for Brain Disorders, Beijing University of Chinese Medicine, Dongzhimen Hospital, Beijing University of Chinese Medicine, 5 Haiyuncang Street, Dongcheng District, Beijing, 100070, China.

E-mail addresses: saga618@126.com (C. Zhang), gaoying973@163.com (Y. Gao).

1. Introduction

The Delphi technique has been used since the 1950s as a generally acknowledged tool to develop professional guidance. The technique was originally developed for military forecasting and has extensively evolved into various disciplines from public policy to health services [1,2]. This method involves an iterative process in a group facilitation technique through a series of structured questionnaires when empirical evidence is contradictory or insufficient [3,4].

Traditional Chinese medicine (TCM) has been verified in clinical practice for over 3000 years, and precisely combines integration and personalization in accordance with the trend in the life sciences in the 21st century [5]. TCM is a vital part of the healthcare-focused medical system and the core of which includes the holistic view and the dialectical view [6]. The dialectical view refers to treatment of a disease based on accurate syndrome diagnosis (also known as syndrome differentiation). TCM syndrome differentiation, the most significant principle in the clinical practice of TCM, leads to new scientific findings for medical sciences, clinical trials, and modern new drug discoveries in TCM [7]. However, TCM syndrome differentiation is greatly affected by the subjective judgment of different physicians, and the syndrome terms have only been unified for less than 10 % [8], seriously hindering the standardization of TCM. Hence, standardizing the diagnostic procedure for syndromes is urgent and represents a key challenge.

The Delphi method has been applied extensively to reach a consensus in health and medical science. In 2004, researchers used the method to explore the common criteria of phlegm congealing syndrome due to liver depression and qi stagnation syndrome due to liver depression in depression, which is the first report available in this field [9]. The application of the Delphi method in TCM syndrome research has gained widespread acceptance, with an average of 17 relevant studies published annually from 2010 to 2020 [10]. It is believed that the Delphi method could establish relatively objective diagnostic criteria by breaking down the boundaries of qualitative or quantitative analysis. The Delphi method is a structured group communication process that integrates individual qualitative opinions for a consensus and addresses group collective opinions through statistically aggregating opinions [11], which can effectively answer the questions mentioned above and draw more accurate conclusions in this field. The method for TCM syndrome diagnosis criteria provides the possibility for syndrome differentiation to transition from a complicated nonlinear form to a linear result. However, since its emergence and widespread application, the Delphi method has been criticized for its inconsistent use without existing consensuses or guidelines overall, not only in TCM. At present, it is urgent to standardize the application and reporting specification of the Delphi technique for TCM syndrome diagnosis research.

In summary, the consistency of the use and reporting quality of the Delphi method in TCM syndrome diagnosis research remain unclear. This cross-sectional analysis seeks to explore and critique the use of the Delphi method for TCM syndrome diagnosis research and assess the reporting of the method.

2. Materials and methods

2.1. Data source

A cross-sectional analysis was undertaken to systematically examine the conduct and reporting of the Delphi method in TCM syndrome differentiation research. We comprehensively searched PubMed, Web of Science, China National Knowledge Infrastructure (CNKI), China Science and Technology Journal Database (VIP), Wanfang Database, and SinoMed from inception to March 25, 2023. Dissertations were also searched. The search strategies are presented in Supplemental Materials.

2.2. Study selection

We clarified the following inclusion criteria: (1) papers available in full-text format about the Delphi in TCM syndrome diagnosis research and (2) papers written in Chinese or English. We excluded studies reported only in abstract forms, conference papers, methodological studies or reviews, comments and duplicate publications. Additionally, expert consensuses or clinical practice guidelines for therapy and studies using the 'hybrid' consensus method (such as the Delphi method combined with brainstorming) were also excluded to ensure the focus on the Delphi method solely for TCM syndrome differentiation research.

2.3. Manually extracted data

On the basis of the prespecified selection criteria, two authors (XYS and ZXC) reviewed the titles and abstracts of the retrieved articles after excluding duplicates. The articles that failed to fulfill the eligibility criteria were removed, and any ambiguous abstracts were considered for full-text review. Two reviewers (XYS and XDH) independently screened the full texts of the remaining articles. Any disagreements in the full texts were resolved by discussion or by involving other professional researchers (YG and CZ) to arbitrate.

Reviewers clarified the definitions of each item and refined the type of data through the relevant literature and a first round of data extraction with ten different articles. The entire team discussed and reached a consensus on any potential problems or uncertainties. Reviewers, in pairs (XDH, YXL, ZXC, and XBZ), independently extracted the data using a preformulated data collection form. If there were more than 20 % interrater disagreements in the data extraction process, including the consistency and completeness of the extracted data, we would refine the definitions, make a consensus through an iterative process, and restart the data extraction process. Any inconsistencies were settled by agreement with other professional researchers (YG, CZ, and QL). We used a qualitative and quantitative data extraction form mainly based on relatively mature studies and reviews [4,12,13], because of the lack of reporting

criteria for the Delphi method to date. The form comprised five parts: (1) Article demographics, (2) Selection of the Delphi panels, (3) Quality assessment of the Delphi process, (4) The reporting of the Delphi rounds, and (5) Consensus and termination of the Delphi process.

2.4. Data analysis

Categorical variables are presented as counts (n) and percentages (%). Quantitative data were reported as the mean ± standard deviation (SD) for normally distributed data or median and interquartile range for nonnormally distributed data. The Shapiro-Wilk test was used to inspect the normality of the quantitative data. All data were collected and recorded in Microsoft Office Excel (Version 365). Statistical analyses were performed using IBM SPSS version 26 (IBM, Armonk, NY, USA).

A total of 37 reporting items were counted for the reporting quality analysis. The items were included mainly based on previous representative studies [4,13,14]. According to the primary principles including anonymity, the multiple rounds, the controlled feedback between rounds, and statistical group response, we selected four critical factors containing anonymity, iteration, controlled feedback, and data analysis. Meanwhile, we selected another five critical factors containing a priori criteria for panels, the heterogeneity of panels, the literature review, the Delphi procedure duration, and the a priori definition of consensus, mainly based on our systematic cognition of the Delphi process for TCM syndrome diagnosis research [15]. Of the included studies, we found the median and average of the number of reporting items were 24, and we defined a relatively high-quality reporting study as a study reported items more than 24. We used chi-squared tests and binary multivariable logistic regression to analyze the associations and sensitivity of the critical factors and the relatively high-quality reporting. A priori criteria for panels, the literature review, anonymity, iteration, controlled feedback, the Delphi procedure duration, data analysis and the a priori definition of consensus were dichotomized as reported versus not reported. The heterogeneity of panels was dichotomized as heterogeneity versus homogeneity or not reported. The nine factors we selected were based on the number of each factor after classification. The result was described using the 95 % Wald confidence intervals (CI).

3. Results

Our search yielded 1832 records, and 1099 records were excluded due to duplications. After the titles or abstracts were screened, 951 records were excluded due to the unconformity of the selection criteria. A total of 148 records were subjected to full-text screening. Of these, 50 studies published between 2007 and 2023 were included in the final dataset (Fig. 1).

3.1. Trial characteristics

For the geographical location of panels, 2 (4.0 %) studies had an international scope, 39 (78.0 %) studies had a national scope conducted by a within-country team, and 3 (6.0 %) studies had an explicitly local scope conducted by a local research team. The geographical scope could not be determined for 6 (12.0 %) studies. In 2 studies, the term 'modified Delphi' was mentioned and the modification details were simply specified. The general characteristics are listed in Table 1, and the studies included covered a wide range of diseases and syndromes (Table S1).
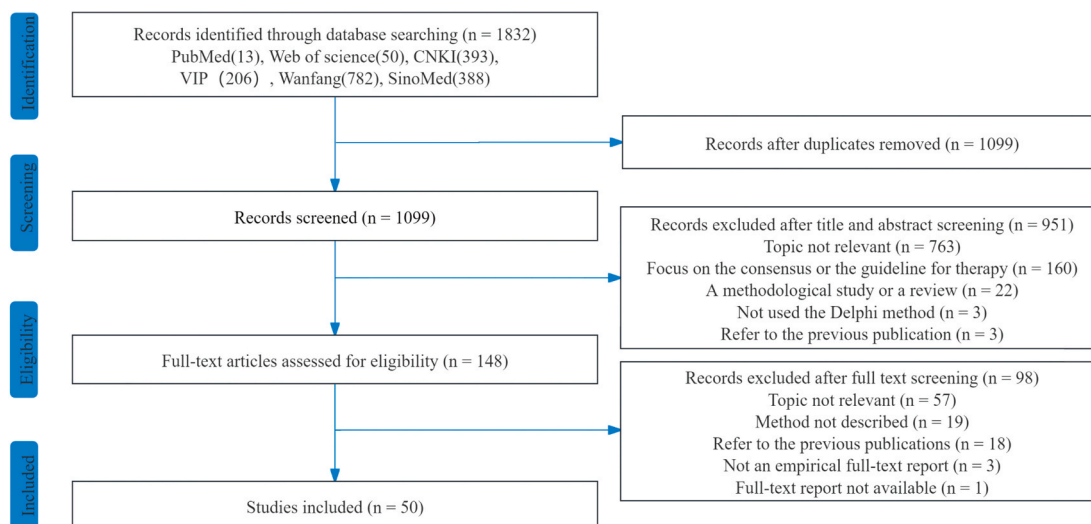

Fig. 1. Flow diagram of study inclusion.

Table 1
Descriptive statistics of 50 studies using the Delphi method in TCM syndrome diagnosis research.

Descriptive Statistics	No. Of articles	% of 50
Type of the Delphi		
Delphi	48	96.00 %
Modified Delphi	2	4.00 %
Geographical scope		
Local	3	6.00 %
National	39	78.00 %
International	2	4.00 %
Not reported	6	12.00 %
Types of panels		
Physician only	34	68.00 %
Physician + Methodologist	1	2.00 %
Physician + Diagnostician	2	4.00 %
Physician + Researcher	4	8.00 %
Physician + Researcher + Diagnostician	5	10.00 %
Not reported	4	8.00 %
Purpose for first questionnaire		
Generation	10	20.00 %
Ranking	6	12.00 %
Ranking + generation	33	66.00 %
Not reported	1	2.00 %
Method used to send questionnaires		
Mail	5	10.00 %
E-mail	3	6.00 %
WeChat	2	4.00 %
Face-to-face	2	4.00 %
On-scene	2	4.00 %
Multiple modes	15	30.00 %
Not reported	21	42.00 %
Rating scale utilized		
5-point Likert scale	13	26.00 %
9-point Likert scale	1	2.00 %
5-point and 9-point Likert scale	1	2.00 %
Not reported	35	70.00 %
No. Of rounds conducted		
1	5	10.00 %
2	37	74.00 %
3	7	14.00 %
5	1	2.00 %

3.2. The selection of panels

Of the 50 studies, the median number of panels was 30 (IQR 20–34.5). The types of panels overwhelmingly comprised physicians alone (34/50; 68.0 %) and only 12 (24.0 %) studies were with a heterogeneous sample of panels; the remaining 4 (8.0 %) studies did not describe the composition. Regarding the proportion of each type of panel, 46 (92.0 %) studies precisely stated the data, even though most comprised physicians alone. Notably, 16 (32.0 %) studies provided background information for panels, and only 3 (6.0 %) studies provided a clear description of the method for panels before the first round. The a priori criteria used to choose potential panels were stated in most of the studies (45/50; 90.0 %), including years of panel experience (40/50; 80.0 %). Furthermore, only 1 (2.0 %) study reported conflicts of interest, emphasizing no competing interests from the funders.

3.3. The reporting of the Delphi rounds

We noted that 13 (26.0 %) studies determined the number of rounds at the beginning of the Delphi process. The number of rounds varied from one to five in each study, with two (37/50; 74.0 %) or three (7/50; 14.0 %) rounds in most studies. Meanwhile, we were concerned that 5 (10.0 %) studies only described one round. Only 10 (20.0 %) studies reported the procedure duration with an average of 7.14 ± 3.29 months. More information about each round is listed in Table 2 and Table S2.

3.4. Consensus and termination

Of the 50 studies, only 13 (26.0 %) adequately defined the a priori consensus, and almost all were based on the degree of agreement. Furthermore, whether the consensus was reached was reported in 25 (50.0 %) studies. Similarly, the reason the Delphi method was terminated was reported in only 14 (28.0 %) studies.

Table 2
Reporting quality in 50 studies using the Delphi method in TCM syndrome diagnosis research.

Reporting item	No. Of articles	Reporting rate
Article demographics		
Type of the Delphi	50	100.00 %
Geographical scope	44	88.00 %
Major topic	50	100.00 %
The selection of the Delphi panel		
No. of panels	50	100.00 %
Types of panels	46	92.00 %
Proportion of each type of panels	36	72.00 %
Background information for panels before round 1	16	32.00 %
Clear description for panels before round 1	3	6.00 %
A priori criteria for panels	45	90.00 %
Conflicts of interest	1	2.00 %
The development of the Delphi process		
Literature review conducted	46	92.00 %
How items were generated for the first questionnaire	50	100.00 %
Purpose for the first questionnaire	49	98.00 %
Method used to send questionnaires	29	58.00 %
Rating scale utilized	15	30.00 %
Feedback to panels	15	30.00 %
Private decisions collected (anonymity)	9	18.00 %
No. Of rounds predefined	13	26.00 %
The reporting of the Delphi rounds		
No. Of rounds conducted	50	100.00 %
No. Of participants indicated	49	98.00 %
No. Of respondents for round 1 indicated	49	98.00 %
No. Of respondents for round 2 indicated	43	95.56 %[a]
No. Of completed for round 1 indicated	49	98.00 %
No. Of completed for round 2 indicated	43	95.56 %[a]
Delphi procedure duration overall	10	20.00 %
Delphi procedure duration for round 1 indicated	8	16.00 %
Delphi procedure duration for round 2 indicated	8	17.78 %[a]
Question formulation for round 1 indicated	48	96.00 %
Question formulation for round 2 indicated	43	95.56 %[a]
Were non-responders invited to subsequent rounds	1	5.00 %[e]
Criteria for dropping items at each round	37	84.09 %[b]
Data analysis clearly justified and reported	49	98.00 %
Formal feedback of results	50	100.00 %
Transparent reporting of results	49	98.00 %
Consensus and termination of the Delphi process		
Consensus defined or not	13	26.00 %
Consensus reached or not	25	50.00 %
Termination of the Delphi	14	28.00 %

[a] 5 articles are not applicable for this item.
[b] 30 articles are not applicable for this item.

3.5. Quality assessment

The foundational procedures of the Delphi process were evaluated; Table 2 presents data regarding whether the included studies were scientifically conducted and transparently reported. Below, we describe particularly noteworthy findings observed in Tables 1 and 2.

Most studies (46/50; 92.0 %) included a literature review. We noted that the purpose for the first questionnaire had been reported in most of the studies (49/50; 98.0 %), principally including generation (10/50; 20.0 %), ranking (6/50; 12.0 %) and a combination of generation and ranking (33/50; 66.0 %). It is worth mentioning that 13 (26.0 %) studies indicated that they utilized the 5-point Likert scale. The method used to send questionnaires was stated in 29 (58.0 %) studies, with diverse combinations of mail, email, telephone, face-to-face, on-scene and Wechat. In addition, only 15 studies explicitly reported controlled feedback to panels. The approaches to collecting the individual decisions of panels, which could be referred to verify whether the anonymity did, were described in 9 (18.0 %) studies. Nearly all studies (49/50; 98.0 %) transparently reported the statistical results. Moreover, only 5 (10.0 %) studies used the flow chart to describe the Delphi process, and 4 were in English.

Furthermore, we attempted to explore whether the year of publication, the geographical scope or the scale (number of panels) had an impact on the reporting quality. However, no clear trend of the reporting quality was observed related to the three factors (Fig. 2).

3.6. Factors that might affect the reporting quality

Nine analytic factors were mentioned in the data analysis. The chi-squared analysis revealed that when the number of reporting
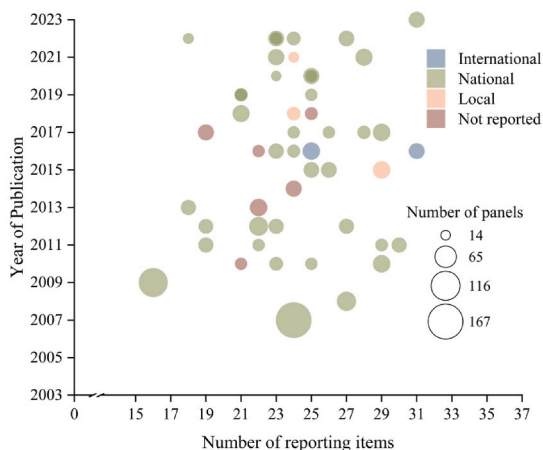
Fig. 2. The bubble plot of the effect of the year of publication, the geographical scope, and the scale (number of panels) on the reporting quality. No clear trend in the reporting quality was observed related to the year of publication, the geographical scope, or the number of panels. The color of the bubble corresponds to the geographical location of panels, with blue denoting an international scope, green denoting a national scope conducted by a within-country team, orange denoting a local scope conducted by a local research team, and brown denoting an unreported scope. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

items was delimited by 24, there was statistically significant difference between the two groups in the rate of relatively high-quality reports of the controlled feedback (p = 0.02, OR 4.36 95 %CI 1.20 to 15.84) (Table 3). Through binary multivariable logistic regression analysis of the results, we did not find a statistical difference between the two groups. Overall, the nine factors included had no statistically significant impact on the reporting quality. The specific results are shown in Supplementary Materials (Table S2). These data suggest that the relatively poor use and reporting of the Delphi method are the common problems in TCM syndrome diagnosis research.

Table 3
Characteristics associated with relative-high quality reporting (items **>** 24).

	Reporting items > 24 (N = 21)	Reporting items ≤24 (N = 29)	Odds ratio (95 %CI)	P value
A priori criteria for panels				0.57
Yes	20 (95.2 %)	25 (86.2 %)	3.20 (0.33–30.94)	
Not reported	1 (4.8 %)	4 (13.8 %)	Reference	
Types of panels				
Homogeneity	18 (85.7 %)	16 (55.2 %)	a	
Heterogeneity	3 (14.3 %)	9 (31.0 %)		
Not reported	0 (0.0 %)	4 (13.8 %)		
Literature review				0.85
Yes	20 (95.2 %)	26 (89.7 %)	2.31 (0.22–23.89)	
Not reported	1 (4.8 %)	3 (10.3 %)	Reference	
Anonymity				1.00
Yes	4 (19.0 %)	5 (17.2 %)	1.13 (0.26–4.84)	
Not reported	17 (81.0 %)	24 (82.8 %)	Reference	
Iteration				0.07
Yes	21 (100.0 %)	24 (82.8 %)	a	
Not reported	0 (0.0 %)	5 (17.2 %)		
Controlled feedback				0.02
Yes	10 (47.6 %)	5 (17.2 %)	4.36 (1.20–15.84)	
Not reported	11 (52.4 %)	24 (82.8 %)	Reference	
Delphi procedure duration				< 0.01
Yes	9 (42.9 %)	1 (3.4 %)	21.00 (2.39–184.63)	
Not reported	12 (57.1 %)	28 (96.6 %)	Reference	
Data analysis				1.00
Yes	21 (100.0 %)	28 (96.6 %)	a	
Not reported	0 (0.0 %)	1 (3.4 %)		
A priori definition of consensus				0.10
Yes	8 (38.1 %)	5 (17.2 %)	2.95 (0.80–10.90)	
Not reported	13 (61.9 %)	24 (82.8 %)	Reference	

CI, confidence interval.
[a] The Odds ratio is not applicable as the frequency count of "Not reported" in "Reporting items **>** 24" is zero.

4. Discussion

A total of 50 studies were conducted using the Delphi or the 'modified Delphi' method in TCM syndrome diagnosis research. Our study is unique in that it appraises the application and reporting related to the Delphi method for TCM syndrome diagnosis and represents the first study to discover the generally inconsistent use and poor quality of reporting in this field. Our study demonstrates that the studies included did not transparently report the procedures that were vital to the interpretation of the results; therefore, we could not infer whether the studies prescriptively used the Delphi technique and whether the results were scientific and reliable.

4.1. The rigor of the design and application of the Delphi method

The 'classic Delphi' was adapted to the 'modified Delphi' [3] with diverse modifications, such as in-person meetings, teleconferences or online meetings. Notably, the 'modified Delphi' is used with no concrete definitions or guidelines, increasing methodological confusion [4]. Only 2 studies reported the use of the 'modified Delphi', a proportion far lower than that in other medical fields [12,16]. Five additional studies involved in-person meetings during the process, which was a variation; thus, these studies should be considered as using the 'modified Delphi'. Furthermore, we wondered whether other Delphi studies used modifications, such as an in-person meeting. Thirteen studies reported on-scene or face-to-face approaches as methods used to distribute questionnaires. In truth, it is difficult to judge such a process to ensure the strict observance of anonymity in studies with generally poor reporting quality.

We consider that researchers in this field might lack adequate cognition of the Delphi and 'modified Delphi' processes. It is suggested that future studies for TCM syndrome diagnosis research strictly distinguish between the classical Delphi and the 'modified Delphi' to avoid technical critique caused by methodological confusion and to increase the rigor of the research. Additionally, the 'modified Delphi' is considered to be more suitable for the procedure of syndrome diagnosis to facilitate the discussion of outstanding items, with an additional face-to-face consensus meeting and an anonymous voting.

4.2. The selection of panels: crucial Delphi procedure

The Delphi method is regarded as an approach that brings individual opinions into a group consensus [17]. Hence, the selection of panels may be crucial for high-quality responses [18]. The Delphi method uses a nonrandom sampling technique for selecting panels [19], considering the requirements of the experience, capacity, willingness, and time to participate [20]. No consensuses or guidelines exist for the selection of panels, and the criteria for potential panels vary considerably across research topics. Our study found that 80.0 % studies used years of clinical/practical experience to gauge expertise, and the similar criteria had been chosen in other medical research [21]. Notably, Baker argued that expertise cannot be evaluated solely through years of experience [22]. We also insist that background information and a clear description of the method should be provided for panels to ensure the willingness and proper time for the whole process to maintain a higher response rate. Furthermore, it is advisable to report whether the members of panels have conflicts of interest. If conflicts of interest exist, we suggest to entrust an independent researcher with the main coordination of the Delphi process [13].

Another consideration when designing a Delphi study is the size of the panels. The optimal panel size has not been found, and it is assumed that larger panel sizes will lead to more stable findings [23]. Nonetheless, an exaggerated number of panels might lead to potential data collection and management challenges [24]. The studies obviously lacked guidance because the number of panels differed dramatically, ranging from 12 to 167. A recent narrative review suggested that the ideal number of panels could be between 8 and 23 with consideration of time and money [4], and we suggest that more research is needed to guide this issue.

In addition to panel size, heterogeneity must also be considered because the composition of the panel can affect the results and the data available for analysis [2]. It is suggested to have a heterogeneous sample of panels for better performance [25], despite the possible difficulties in collecting and analyzing data [20]. Our study found that only 24.0 % of the studies we screened included other types of interprofessional experts, while another group previously highlighted that interprofessional groups were infrequent in general [12]. We suggest that multidisciplinary studies can reach better performances from clinical, diagnostic, methodological, and other perspectives, and will contribute to higher quality responses and more convincing consensuses for the standardization of syndrome diagnosis research.

4.3. The primary principles underpin the Delphi method

Four primary principles underpin the methodology, including anonymity, the multiple rounds, the controlled feedback between rounds, and statistical group response [15].

Anonymity is a foundational principle for high-quality consensus, differentiating the Delphi method from other consensus methods. Panels could present and change individual opinions privately unbiased by the dominance of identities and pressures [26]. However, approximately 82.0 % studies omitted presenting anonymity in the articles. We wonder whether anonymity was designed scientifically in these studies, and it is necessary for future researchers to report anonymity to demonstrate the rigor of their study.

Iteration is the central feature of consensus methods [27], and at least two rounds are required for the entire procedure. We found that the classic Delphi has four rounds [17], whereas the choice of three rounds seems optimal for a better-balanced procedure duration, panel fatigue and meaningful findings [14]. We found that 37/50 (74.0 %) studies reported 2 rounds in total, and whether stability can be confirmed requires further investigation [28]. Concerningly, 5 studies reported only one survey round; an iterative process should involve at least two rounds. From a critical perspective, such a research design cannot be considered as the Delphi

method.

Iteration with controlled feedback requires panels to share collected perspectives and judge whether to rerank personal responses [20,29]. The group consensus is facilitated through a repeated process between rounds and controlled feedback. Notably, 70.0 % of the included studies did not report the controlled feedback process. It is advisable that such strengths of the Delphi method should be transparently reported.

We found a high reporting rate (98.0 %) in statistical results, probably because diagnostic criteria should be explicitly provided in this field. However, we found it difficult to evaluate the quality of the results due to the lack of scientific and unanimous application of the Delphi. On the other hand, researchers are suggested that they should interpret the statistical results cautiously, as the results cannot be automatically used to represent stability.

We found that there were misunderstandings of the rationale and the application of Delphi in the previous TCM syndrome diagnosis research. Indeed, the primary principles of the Delphi should be strictly followed for addressing relevant topics to set apart from the normal expert meeting.

4.4. A priori definition of the consensus is recommended for scientificity

Whether an a priori consensus definition was identified should be highlighted because only 26.0 % of the studies we reviewed reported this issue. Almost all of the studies mentioned above reached a consensus based on the degree of agreement. However, a general agreement degree does not refer to the Delphi method, as the degree depends on the number of panels, research objectives and other issues [17]. It is mentioned that the stability of the response could be a more reliable indicator of consensus [30]. Meanwhile, only 50.0 % of the included studies reported whether a consensus was reached. The Delphi method is regarded as an approach to force a consensus, which is criticized for the lack of opportunity for discussion and stating of views [31]. The result of a consensus is not equal to the right opinion and merely represents group identification of the definite selected panels. We also suggest that this critical procedure must be described.

The consensus could be defined using measures of not only the percent agreement but also the central tendency of responses and dispersion of responses. It should be stressed that the mean can not be calculated as the Likert scale data can be merely suitable for data on an interval scale [32]. At present, few recommendations are available on the optimal size of Likert scales in the relevant literature and future research should aim for setting agreement on the optimal size of Likert scales [14]. Given the significance of the inclusion of a midpoint in the Likert scale, researchers should cautiously and scientifically choose between the odd number of categories and the even number of categories. As the choice of an even number of categories can avoid the possibility of a neutral attitude, which might affect the reliability and validity [33]. We hope that panels can make a non-neutral choice as the definite positive and negative symptoms are meaningful for syndrome diagnosis and treatment.

We did not verify whether the results of the studies we reviewed were practically used in syndrome differentiation. Nevertheless, the final consensus might be conservative considering the loss of controversial opinions, and we suggest utilizing the results with caution due to the poor quality. This limits the development of TCM syndrome diagnosis to a great extent, and there is an urgent need for consistency and quality both in the conduct and in the reporting of studies using the Delphi method. And Table 4 shows the list of considerations and recommendations for the scientific design and detailed reporting for future studies built on the results of our study.

4.5. Strengths and limitations

To the best of our knowledge, this is the first study to explore and critique the conduct and reporting of the Delphi method in TCM syndrome diagnosis research. Meanwhile, we made recommendations for best practices in future research that are expected to improve results.

This study has some limitations. We only searched six databases for original articles, with Chinese and English language limitations. In addition, we did not include the relevant gray literature, which might contain relevant studies. Nevertheless, the six databases we searched are considered pertinent to obtain sufficient local studies for analysis. Additionally, the data extraction process was

Table 4
Considerations and recommendations for the scientific design and detailed reporting.

Procedures of the Delphi	Considerations and recommendations
The priori criteria for panels	The a priori criteria are needed for potential panels, considering the relatively comprehensive requirements of the experience, capacity, willingness, and time to participant.
The panel size and panel heterogeneity	Exaggerated number of panels might lead to the potential data collection and management trouble. The heterogeneous panels could reach more convincing consensus from multidisciplinary perspectives.
Conflicts of interest	If the members of panels have any (potential) conflict of interest, please state it and entrust an independent researcher with the main coordination of the Delphi process.
Literature review	It is suggested that the initial structured questionnaire should be conjunct with literature reviews.
The rating scale utilized	Researchers should cautiously take into account the size of Likert scales and whether to include a midpoint. An "open text boxes" could be combined to add in comments if needed.
Iteration with controlled feedback	We recommend three or four rounds to balance the levels of panel fatigue and meaningful findings. The iteration with controlled feedback facilitates the group consensus, which is highly advisable to be reported.
Statistical results	Statistical results are highly advisable to be interpreted cautiously, as the results cannot be automatically used to represent stability.

independently conducted by four reviewers in pairs with at least 80 % interrater agreement, to increase the robustness and credibility of the analysis. Furthermore, the Delphi technique is a quite valuable tool but with methodological uncertainty, contributing to the variability in the conducting and poor reporting quality. Hence, future research should be devoted to settling the international guidelines and then we hope the ongoing cross-sectional analysis will be based on more scientific studies for further analysis and recommendations. Despite the limitations mentioned above, we believe that our study provides convincing and reliable findings.

5. Conclusion

We have noted the popularity of the Delphi method with regard to TCM syndrome diagnosis research and discovered considerable variability in its implementation and poor reporting quality. A narrative literature search of the Delphi in health sciences research also indicated that there were universally ongoing debates and inconsistencies in many procedures of the Delphi [4]. Therefore, further research and improved criteria are needed to develop universal guidelines for issues of fuzziness in the Delphi method; ultimately standardizing TCM syndrome diagnosis research will improve scientific credibility.

Ethical statement

No ethical approval was required as this study did not involve human participants or laboratory animals.

Data availability

The data that support the findings of this study are available within the article and its supplementary materials. There is no additional data available.

CRediT authorship contribution statement

Xinyi Shi: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xudong He:** Resources, Investigation, Data curation. **Qiang Liu:** Writing – review & editing, Validation. **Luda Feng:** Writing – review & editing, Validation. **Yixuan Li:** Visualization, Validation, Resources, Investigation, Formal analysis. **Xuebin Zhang:** Investigation. **Zixin Cheng:** Investigation. **Chi Zhang:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Ying Gao:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e25162.

References

[1] N. Dalkey, D.O. Helmer, An experimental application of the DELPHI method to the use of experts, Manag. Sci. 9 (3) (1963) 458–467. https://doi.org/10.1287/mnsc.9.3.458.
[2] J. Jones, D. Hunter, Qualitative research: consensus methods for medical and health services research, BMJ 311 (7001) (1995) 376–380. https://doi.org/10.1136/bmj.311.7001.376.
[3] H.P. McKenna, The Delphi technique: a worthwhile research approach for nursing? J. Adv. Nurs. 19 (6) (1994) 1221–1225. https://doi.org/10.1111/j.1365-2648.1994.tb01207.x.
[4] Z. Shang, Use of Delphi in health sciences research: a narrative review, Medicine (Baltim.) 102 (7) (2023) e32829. https://doi.org/10.1097/MD.0000000000032829.

[5] B. Yuan, Towards a clinical efficacy evaluation system adapted for personalized medicine, Pharmgenomics Pers Med. 14 (2021) 487–496. https://doi.org/10.2147/PGPM.S304420.

[6] D.Z. Sun, S.D. Li, Y. Liu, Y. Zhang, R. Mei, M.H. Yang, Differences in the origin of philosophy between Chinese medicine and Western medicine: exploration of the holistic advantages of Chinese medicine, Chin. J. Integr. Med. 19 (9) (2013) 706–711. https://doi.org/10.1007/s11655-013-1435-5.

[7] M. Jiang, C. Lu, C. Zhang, J. Yang, Y. Tan, A. Lu, et al., Syndrome differentiation in modern research of traditional Chinese medicine, J. Ethnopharmacol. 140 (3) (2012) 634–642. https://doi.org/10.1016/j.jep.2012.01.033.

[8] Y. Wang, Suggestions on improving the methodological system of traditional Chinese medicine syndrome differentiation, J. Tradit. Chin. Med. 10 (2004) 729–731. https://doi.org/10.3321/j.issn:1001-1668.2004.10.003.

[9] W. Song, S. Hu, H. Zhang, Z. Chen, Analysis of second round expert questionnaire about criterions of syndromes of stagnation and phlegm retention of the hepatic-qi in depression, Chin. J. Tissue Eng. Res. 8 (3) (2004) 488–489. https://doi.org/10.3321/j.issn:1673-8225.2004.03.046.

[10] H. Fu, G. Dai, W. Gao, Z. Zhao, X. Liu, Y. Li, Systematic evaluation of the application of Delphi method in traditional Chinese medicine clinical syndrome, World Sci. Technol.-Modernization of Tradit. Chin. Med. 24 (10) (2022) 4082–4088. https://doi.org/10.11842/wst.20210909010.

[11] H. Wu, L. Xu, Y. Zheng, L. Shi, L. Zhai, F. Xu, Application of the Delphi method in the study of depressive disorder, Front. Psychiatr. 13 (2022) 925610. https://doi.org/10.3389/fpsyt.2022.925610.

[12] S. Humphrey-Murto, L. Varpio, T.J. Wood, C. Gonsalves, L.A. Ufholz, K. Mascioli, et al., The use of the Delphi and other consensus group methods in medical education research: a review, Acad. Med. 92 (10) (2017) 1491–1498. https://doi.org/10.1097/ACM.0000000000001812.

[13] S. Jünger, S.A. Payne, J. Brine, L. Radbruch, S.G. Brearley, Guidance on Conducting and REporting DElphi Studies (CREDES) in palliative care: recommendations based on a methodological systematic review, Palliat. Med. 31 (8) (2017) 684–706. https://doi.org/10.1177/0269216317690685.

[14] E.G. Trevelyan, N. Robinson, Delphi methodology in health research: how to do it? Eur. J. Integr. Med. 7 (4) (2015) 423–428. https://doi.org/10.1016/j.eujim.2015.07.002.

[15] S. Humphrey-Murto, T.J. Wood, C. Gonsalves, K. Mascioli, L. Varpio, The Delphi method, Acad. Med. 95 (1) (2020) 168. https://doi.org/10.1097/ACM.0000000000002887.

[16] T. Foth, N. Efstathiou, B. Vanderspank-Wright, L.A. Ufholz, N. Dütthorn, M. Zimansky, et al., The use of Delphi and Nominal Group Technique in nursing education: a review, Int. J. Nurs. Stud. 60 (2016) 112–120. https://doi.org/10.1016/j.ijnurstu.2016.04.015.

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

[18] C. Duffield, The Delphi technique: a comparison of results obtained using two expert panels, Int. J. Nurs. Stud. 30 (3) (1993) 227–237. https://doi.org/10.1016/0020-7489(93)90033-q.

[19] S. Keeney, F. Hasson, H.P. McKenna, A critical review of the Delphi technique as a research methodology for nursing, Int. J. Nurs. Stud. 38 (2) (2001) 195–200. https://doi.org/10.1016/s0020-7489(00)00044-4.

[20] G.J. Skulmoski, F.T. Hartman, J. Krahn, The Delphi method for graduate research, J. Inf. Technol. Educ. 6 (1) (2007) 1–21. https://www.learntechlib.org/p/111405/.

[21] S. McPherson, C. Reese, M.C. Wendler, Methodology update: Delphi studies, Nurs. Res. 67 (5) (2018) 404–410. https://doi.org/10.1097/NNR.0000000000000297.

[22] J. Baker, K. Lovell, N. Harris, How expert are the experts? An exploration of the concept of 'expert' within Delphi panel techniques, Nurse Res. 14 (1) (2006) 59–70. https://doi.org/10.7748/nr2006.10.14.1.59.c6010.

[23] A.F. Jorm, Using the Delphi expert consensus method in mental health research, Aust. N. Z. J. Psychiatr. 49 (10) (2015) 887–897. https://doi.org/10.1177/0004867415600891.

[24] C.C. Hsu, B.A. Sanford, The Delphi technique: making sense of consensus, Practical Assess. Res. Eval. 12 (2007) 10. https://doi.org/10.7275/pdz9-th90.

[25] C. Powell, The Delphi technique: myths and realities, J. Adv. Nurs. 41 (4) (2003) 376–382. https://doi.org/10.1046/j.1365-2648.2003.02537.x.

[26] C.M. Goodman, The Delphi technique: a critique, J. Adv. Nurs. 12 (6) (1987) 729–734. https://doi.org/10.1111/j.1365-2648.1987.tb01376.x.

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

[28] M.K. Murphy, N.A. Black, D.L. Lamping, C.M. McKee, C.F. Sanderson, J. Askham, et al., Consensus development methods, and their use in clinical guideline development, Health Technol. Assess. 2 (3) (1998) i–88.

[29] G. Rowe, G. Wright, The Delphi technique as a forecasting tool: issues and analysis, Int. J. Forecast. 15 (4) (1999) 353–375. https://doi.org/10.1016/S0169-2070(99)00018-7.

[30] J. Crisp, D. Pelletier, C. Duffield, A. Adams, S. Nagy, The Delphi method, Nurs. Res. 46 (2) (1997) 116–118. https://doi.org/10.1097/00006199-199703000-00010.

[31] M.A. Walker, M.J. Selfe, The Delphi method: a useful tool for the allied health researcher, Int. J. Ther. Rehabil. 3 (12) (1996) 677–681. https://doi.org/10.12968/bjtr.1996.3.12.14731.

[32] H.A. von der Gracht, Consensus measurement in Delphi studies: review and implications for future quality assurance, Technol. Forecast. Soc. Change 79 (8) (2012) 1525–1536. https://doi.org/10.1016/j.techfore.2012.04.013.

[33] K.K. Tsang, The use of midpoint on Likert Scale: the implications for educational research, Hong Kong Teach. Cent. J. 11 (2012) 121–130.