Interdisciplinary Research in Medical Sciences Specialty
ISSN: 3062-4401
2023, Volume 3, Issue 1, Page No: 71-81

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

A Cross-Sectional Analysis of the Application and Reporting of the Delphi Method in Traditional Chinese Medicine Syndrome Diagnosis

Anders Johansson[1], Erik Lundqvist[1], Johan Persson, Mikael Berg[1]*

[1]Department of Translational Science, Faculty of Medicine, Uppsala University, Uppsala, Sweden.

***E-mail** ✉ mikael.berg.med@gmail.com

Received: 19 November 2022; **Revised:** 23 February 2023; **Accepted:** 26 February 2023

ABSTRACT

In traditional Chinese medicine (TCM) syndrome diagnosis research, the Delphi technique is often used to build expert consensus when evidence is limited and opinions differ. Despite its frequent use, there has been no systematic evaluation of how this method is conducted or reported in this context. This study investigates the consistency of Delphi method application and evaluates the quality of reporting in TCM research. A cross-sectional analysis was performed to identify studies utilizing the Delphi approach in TCM syndrome diagnosis. Searches were conducted in PubMed, Web of Science, CNKI, VIP, Wanfang, and SinoMed for publications in English or Chinese up to March 25, 2023. A structured extraction form captured study characteristics and methodological details to assess rigor and transparency. From 1832 screened records, 50 studies were included. The median panel size was 30 (IQR 20–34.5), with only 24% having diverse panel composition. Two Delphi rounds were most frequent (74%), followed by three rounds (14%), and only a quarter of studies predetermined the number of rounds. Key reporting elements were inconsistently documented: anonymity was mentioned in 18%, controlled feedback in 30%, process duration (mean 7.14 ± 3.29 months) in 20%, and predefined consensus criteria in 26% of studies. The use of the Delphi method in TCM syndrome diagnosis research shows considerable variation and often lacks transparent reporting. Developing standardized guidelines is crucial to enhance methodological consistency and reporting quality in future studies.

Keywords: Cross-sectional analysis, Syndrome diagnosis research, Traditional Chinese medicine, The Delphi

How to Cite This Article: Johansson A, Lundqvist E, Persson J, Berg M. A Cross-Sectional Analysis of the Application and Reporting of the Delphi Method in Traditional Chinese Medicine Syndrome Diagnosis. Interdiscip Res Med Sci Spec. 2023;3(1):71-81. https://doi.org/10.51847/6rA7zEwAL1

Introduction

Since its development in the 1950s, the Delphi technique has been widely recognized as a structured method for formulating expert guidance. Originally designed for military forecasting, it has since expanded across numerous fields, including public policy and healthcare [1, 2]. The method employs an iterative, group-based process using structured questionnaires to achieve consensus when empirical evidence is limited or conflicting [3, 4].

Traditional Chinese medicine (TCM), with a clinical history spanning over 3,000 years, emphasizes both integration and individualized treatment, aligning with contemporary trends in life sciences [5]. Central to TCM is a holistic and dialectical approach, with the latter guiding treatment based on precise syndrome diagnosis, also known as syndrome differentiation [6]. Syndrome differentiation is pivotal in TCM clinical practice and has contributed to scientific advances, clinical trials, and the discovery of novel TCM-based therapies [7]. However, this process is heavily influenced by physician subjectivity, and less than 10% of syndrome terminology has been standardized [8], creating significant barriers to uniformity. Standardizing diagnostic procedures for TCM syndromes is therefore a pressing challenge.

The Delphi method has increasingly been employed in medical and health research to establish consensus. Its first documented application in TCM syndrome research occurred in 2004, focusing on phlegm congealing syndrome and qi stagnation syndrome associated with liver depression in depression [9]. From 2010 to 2020, approximately

17 studies per year applied the Delphi method in TCM syndrome research [10]. By systematically integrating expert opinion, the Delphi technique offers a way to derive objective diagnostic criteria from subjective qualitative judgments. It aggregates individual perspectives through statistical methods, transforming complex, nonlinear syndrome assessments into more structured outcomes [11]. Despite its widespread use, the Delphi method has faced criticism for inconsistent implementation and the absence of standardized guidance, not only in TCM but across disciplines. Establishing clear application and reporting standards for the Delphi method in TCM syndrome research remains an urgent need.

In summary, the consistency of Delphi method application and the quality of its reporting in TCM syndrome research are unclear. This study aims to systematically examine how the Delphi method is applied and reported in TCM syndrome diagnosis research through a cross-sectional analysis.

Materials and Methods

Data sources

We conducted a cross-sectional analysis to systematically review the use and reporting of the Delphi method in TCM syndrome differentiation studies. Comprehensive searches were performed in PubMed, Web of Science, CNKI, VIP, Wanfang, and SinoMed from their inception to March 25, 2023. Dissertations were also included.

Study selection

Inclusion criteria were: (1) full-text studies applying the Delphi method in TCM syndrome diagnosis, and (2) publications in English or Chinese. Exclusion criteria included abstracts only, conference papers, methodological reviews, commentaries, duplicate publications, clinical practice guidelines, and studies using hybrid consensus methods (e.g., Delphi combined with brainstorming), ensuring focus on the Delphi method alone.

Data extraction

Two authors (XYS and ZXC) initially screened titles and abstracts after removing duplicates. Articles not meeting eligibility criteria were excluded, and ambiguous abstracts underwent full-text review. Full texts were independently evaluated by two reviewers (XYS and XDH), with disagreements resolved through discussion or arbitration by additional researchers (YG and CZ).

Definitions and data types were refined using literature review and a pilot extraction of ten articles. The team reached consensus on ambiguities. Data extraction was performed in pairs (XDH, YXL, ZXC, XBZ) using a standardized form. If more than 20% discrepancies occurred, definitions were clarified, consensus achieved, and extraction repeated. Remaining inconsistencies were resolved with input from YG, CZ, and QL.

The extraction form, adapted from previous studies [4, 12, 13] due to the lack of formal reporting criteria for the Delphi method, included five sections: (1) Article demographics, (2) Delphi panel selection, (3) Quality assessment of the Delphi process, (4) Reporting of Delphi rounds, and (5) Consensus and termination criteria.

Data analysis

Categorical variables were summarized using counts (n) and percentages (%), while continuous variables were reported as mean ± standard deviation (SD) for normally distributed data or as median with interquartile range (IQR) for data not following a normal distribution. The Shapiro-Wilk test was applied to assess the normality of quantitative variables. All data were organized and stored in Microsoft Excel (Version 365), and statistical analyses were conducted using IBM SPSS version 26 (IBM, Armonk, NY, USA).

For evaluating reporting quality, 37 specific items were extracted, primarily informed by previous representative studies [4, 13, 14]. Based on core Delphi principles—including anonymity, iterative rounds, controlled feedback, and statistical aggregation of responses—four key factors were selected: anonymity, iteration, controlled feedback, and data analysis. Additionally, five more factors were considered critical in the context of TCM syndrome diagnosis research: pre-specified panel criteria, panel heterogeneity, literature review, duration of the Delphi process, and predefined consensus criteria [15]. Across all included studies, the median and mean number of reported items was 24; studies reporting more than 24 items were classified as having relatively high reporting quality. Associations between the selected critical factors and high-quality reporting were analyzed using chi-square tests and multivariable binary logistic regression. Dichotomization was applied: a priori panel criteria, literature review, anonymity, iteration, controlled feedback, procedure duration, data analysis, and predefined

consensus were coded as reported versus not reported, while panel heterogeneity was classified as heterogeneous versus homogeneous or unreported. Results were presented with 95% Wald confidence intervals (CI).

Results and Discussion

The literature search initially identified 1,832 records, of which 1,099 duplicates were removed. Screening of titles and abstracts led to the exclusion of 951 records that did not meet inclusion criteria. Full texts of 148 articles were assessed, and 50 studies published between 2007 and 2023 were ultimately included in the analysis **(Figure 1).**

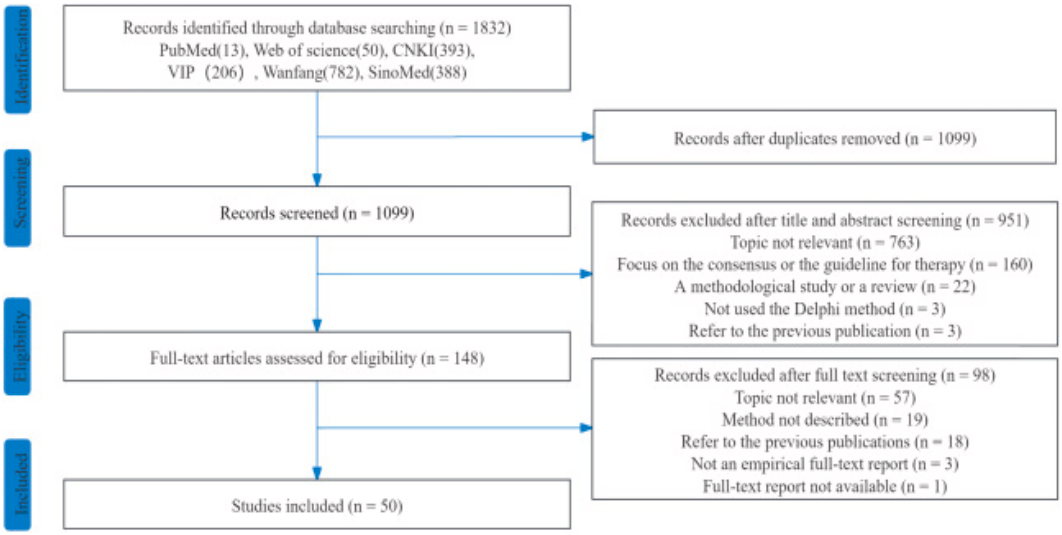

Figure 1. Flowchart of Study Selection

Study characteristics

Regarding the geographic scope of the Delphi panels, 2 studies (4.0%) included international participants, 39 studies (78.0%) were conducted at a national level within a single country, and 3 studies (6.0%) were explicitly local, led by regional research teams. For 6 studies (12.0%), the geographic coverage could not be determined. Two studies referred to a "modified Delphi" approach, although details of the modifications were minimally described. Key study characteristics are summarized in **Table 1**, and the included studies addressed a diverse array of diseases and syndromes.

Table 1. Summary Characteristics of 50 Studies Employing the Delphi Method in TCM Syndrome Diagnosis Research

Characteristic	Number of Studies	Percentage (%)
Delphi Type		
Standard Delphi	48	96.00
Modified Delphi	2	4.00
Geographic Scope of Panels		
Local	3	6.00
National	39	78.00
International	2	4.00
Not reported	6	12.00
Panel Composition		
Physicians only	34	68.00
Physicians + Methodologists	1	2.00
Physicians + Diagnosticians	2	4.00
Physicians + Researchers	4	8.00
Physicians + Researchers + Diagnosticians	5	10.00
Not reported	4	8.00
Purpose of First Questionnaire		
Item Generation	10	20.00

Ranking	6	12.00
Item Generation + Ranking	33	66.00
Not reported	1	2.00
Mode of Questionnaire Distribution		
Mail	5	10.00
Email	3	6.00
WeChat	2	4.00
Face-to-face	2	4.00
On-site	2	4.00
Multiple methods	15	30.00
Not reported	21	42.00
Rating Scale Used		
5-point Likert	13	26.00
9-point Likert	1	2.00
Both 5- and 9-point Likert	1	2.00
Not reported	35	70.00
Number of Rounds Conducted		
1 round	5	10.00
2 rounds	37	74.00
3 rounds	7	14.00
5 rounds	1	2.00

Panel selection

Across the 50 included studies, the median panel size was 30 (IQR 20–34.5). Most panels were composed solely of physicians (34/50; 68.0%), while only 12 studies (24.0%) included a heterogeneous mix of panelists, and 4 studies (8.0%) did not report the panel composition. Despite the predominance of physicians, 46 studies (92.0%) provided precise data on panel types. Background information about panel members was presented in 16 studies (32.0%), but only 3 studies (6.0%) clearly described the method used to select panels prior to the first round. Most studies (45/50; 90.0%) specified a priori criteria for panel inclusion, most commonly based on years of professional experience (40/50; 80.0%). Only a single study (2.0%) reported conflicts of interest, indicating no competing interests from funders.

Reporting of Delphi rounds

Thirteen studies (26.0%) pre-specified the number of Delphi rounds at the outset. The total number of rounds ranged from one to five, with two rounds being the most common (37/50; 74.0%), followed by three rounds (7/50; 14.0%). Notably, 5 studies (10.0%) reported conducting only a single round. Procedure duration was documented in just 10 studies (20.0%), with an average length of 7.14 ± 3.29 months. Additional details for each round are summarized in **Table 2.**

Table 2. Reporting Quality of the Delphi Method in 50 Studies on Traditional Chinese Medicine (TCM) Syndrome Diagnosis

Reporting Item	Number of Articles Reporting the Item	Reporting Rate
Article Characteristics		
Type of Delphi method used	50	100.00%
Geographical scope of the study	44	88.00%
Main research topic	50	100.00%
Composition and Selection of Expert Panel		
Number of expert panels	50	100.00%
Types of experts included in the panel	46	92.00%
Proportion of each expert type	36	72.00%
Provision of background information before Round 1	16	32.00%
Clear description of pre-Round 1 information	3	6.00%
Pre-defined inclusion/exclusion criteria for experts	45	90.00%
Disclosure of conflicts of interest	1	2.00%

Design and Development of the Delphi Process		
Performance of a literature review	46	92.00%
Description of how items were generated for Round 1 questionnaire	50	100.00%
Statement of the purpose of the initial questionnaire	49	98.00%
Method of questionnaire distribution	29	58.00%
Specification of the rating scale used	15	30.00%
Provision of feedback to experts between rounds	15	30.00%
Maintenance of respondent anonymity	9	18.00%
Pre-specification of the number of rounds	13	26.00%
Reporting of Delphi Rounds		
Total number of rounds performed	50	100.00%
Total number of participants reported	49	98.00%
Number of respondents in Round 1 reported	49	98.00%
Number of respondents in Round 2 reported	43	95.60% (a)
Number of fully completed questionnaires in Round 1	49	98.00%
Number of fully completed questionnaires in Round 2	43	95.6% (a)
Total duration of the entire Delphi process	10	20.00%
Duration of Round 1 reported	8	16.00%
Duration of Round 2 reported	8	17.8% (a)
Wording of questions in Round 1 provided	48	96.00%
Wording of questions in Round 2 provided	43	95.6% (a)
Whether non-respondents were re-invited	1	5.0% (e)
Criteria for removing items at each round	37	84.1% (b)
Clear justification and description of statistical analysis	49	98.00%
Provision of formal feedback on group results	50	100.00%
Transparent presentation of final results	49	98.00%
Consensus Definition and Process Termination		
Explicit definition of consensus	13	26.00%
Statement of whether consensus was achieved	25	50.00%
Description of termination criteria for the Delphi	14	28.00%

(a) 5 articles did not have a second round (not applicable).

(b) 30 articles did not involve item elimination (not applicable).

Consensus and termination

Among the 50 studies analyzed, only 13 (26.0%) clearly specified the a priori criteria for consensus, with most relying on the level of agreement among participants. Information on whether consensus was ultimately achieved was provided in 25 studies (50.0%). Similarly, just 14 studies (28.0%) reported the rationale for terminating the Delphi process.

Quality assessment

The core procedures of the Delphi methodology were assessed, and **Table 2** summarizes whether the included studies were executed rigorously and reported transparently. Key observations from **Tables 1 and 2** are highlighted below.

A majority of studies (46/50; 92.0%) incorporated a literature review. The objective of the initial questionnaire was reported in nearly all studies (49/50; 98.0%), primarily for generation (10/50; 20.0%), ranking (6/50; 12.0%), or a combination of generation and ranking (33/50; 66.0%). Notably, 13 studies (26.0%) employed a 5-point Likert scale. The mode of questionnaire delivery was detailed in 29 studies (58.0%), using various combinations of mail, email, telephone, in-person, on-site, and WeChat. Only 15 studies explicitly mentioned providing controlled feedback to the panels. Methods for collecting individual panel responses, which help confirm anonymity, were described in 9 studies (18.0%). Nearly all studies (49/50; 98.0%) clearly reported statistical outcomes. However, only 5 studies (10.0%) used flowcharts to illustrate the Delphi procedure, with 4 of these published in English.

Additionally, an exploration of the influence of publication year, geographical coverage, and panel size on reporting quality revealed no evident patterns **(Figure 2).**

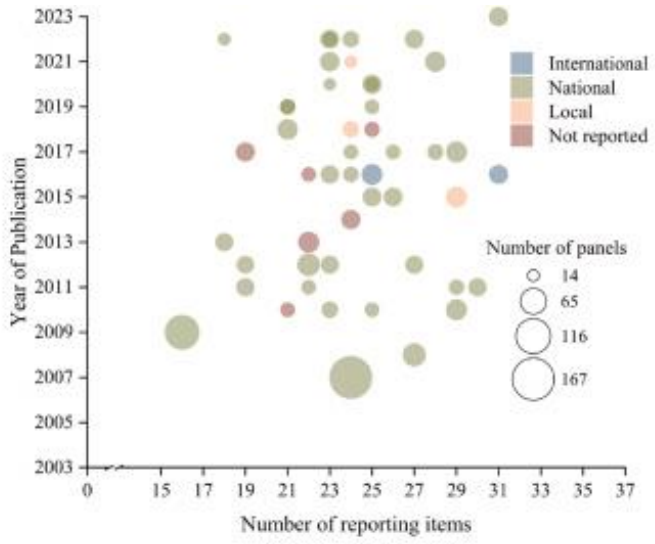

Figure 2. A bubble plot illustrating how the year of publication, geographic coverage, and panel size relate to the quality of reporting.

Factors influencing reporting quality

The analysis considered nine potential factors. Chi-square testing showed that when the threshold for reporting items was set at 24, studies providing controlled feedback were significantly more likely to be of relatively high reporting quality (p = 0.02; OR 4.36; 95% CI 1.20–15.84) **(Table 3)**. Nevertheless, results from binary multivariable logistic regression did not reveal any statistically significant differences between the groups. In general, none of the nine variables demonstrated a meaningful impact on reporting quality. These observations indicate that inadequate implementation and reporting of the Delphi method remain widespread challenges in research on TCM syndrome diagnosis.

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 %)		
Heterogeneity	3 (14.3 %)	9 (31.0 %)	a	
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 %)		

Not reported	12 (57.1 %)	28 (96.6 %)	21.00 (2.39–184.63) 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.

aThe Odds ratio is not applicable as the frequency count of "Not reported" in "Reporting items＞24" is zero.

This review examined 50 studies employing the Delphi or "modified Delphi" method within TCM syndrome diagnosis research. Our work is novel in evaluating both the implementation and reporting quality of the Delphi method in this field, revealing generally inconsistent application and substandard reporting. The findings indicate that most studies lacked transparent documentation of procedures critical for interpreting results, making it difficult to determine whether the Delphi methodology was properly applied and whether the outcomes can be considered scientifically reliable.

Rigor in the design and application of the Delphi method

The traditional "classic Delphi" has frequently been adapted into a "modified Delphi" [3], incorporating variations such as in-person meetings, teleconferences, or online discussions. However, the modified version lacks standardized definitions or guidelines, often causing methodological ambiguity [4]. Only two studies explicitly identified the use of a modified Delphi, a notably lower proportion compared with other medical disciplines [12, 16]. An additional five studies included in-person meetings during the process, which could also be categorized as modified Delphi. Moreover, thirteen studies reported face-to-face or on-site distribution of questionnaires, but given the overall poor reporting quality, it is challenging to ascertain whether participant anonymity was adequately preserved.

These observations suggest that researchers in TCM syndrome diagnosis may have insufficient familiarity with the Delphi or modified Delphi methodologies. Future studies should clearly distinguish between the classic and modified Delphi approaches to reduce methodological confusion and enhance rigor. The modified Delphi may be particularly suitable for syndrome diagnosis, as it allows in-person consensus discussions alongside anonymous voting for unresolved items.

Panel selection: a key Delphi procedure

The Delphi technique aims to consolidate individual perspectives into collective consensus [17], making panel selection critical for ensuring high-quality outcomes [18]. Panels are typically chosen using non-random sampling based on expertise, experience, willingness, and availability [19, 20]. No universal guidelines exist for panel selection, and criteria vary across studies. In our review, 80.0% of studies assessed expertise primarily by years of clinical or practical experience, a method similarly applied in other medical research [21]. However, as noted by Baker, experience alone is not sufficient to determine expertise [22]. We emphasize the importance of reporting panelists' backgrounds and clearly describing selection methods, including availability and commitment, to maintain high response rates. Disclosure of conflicts of interest is also recommended; if conflicts exist, independent coordination of the Delphi process should be considered [13].

Panel size is another important consideration, though an optimal number has not been established. Larger panels may enhance result stability [23], but overly large panels can create logistical and management challenges [24]. In the studies reviewed, panel sizes varied widely from 12 to 167 participants, reflecting a lack of standardized guidance. A recent narrative review suggests that 8–23 panelists may be ideal when considering practical constraints such as time and cost [4], though further research is needed to refine these recommendations.

Panel composition heterogeneity is also crucial, as it can influence outcomes and the breadth of data collected [2]. A diverse panel is generally recommended for optimal performance [25], despite potential challenges in data collection and analysis [20]. In our sample, only 24.0% of studies included interprofessional experts, consistent with previous observations that multidisciplinary panels are uncommon [12]. Encouraging a multidisciplinary approach could enhance the quality of responses, strengthen consensus validity, and improve the standardization of syndrome diagnosis research.

Core principles of the Delphi method

The Delphi methodology is founded on four key principles: anonymity, multiple iterative rounds, controlled feedback between rounds, and statistical aggregation of group responses [15]. Anonymity is essential for achieving unbiased consensus, as it allows panel members to express and revise their opinions without being influenced by dominant personalities or social pressures [26]. However, around 82.0% of the studies reviewed did not mention how anonymity was maintained, raising concerns about whether this principle was implemented rigorously. Future research should explicitly report anonymity procedures to strengthen study credibility.

Iteration is a defining feature of the Delphi approach [27], with at least two rounds required to complete the process. While the classical Delphi typically involves four rounds [17], three rounds are often considered optimal to balance procedural length, panel fatigue, and meaningful outcomes [14]. In this review, 37 of 50 studies (74.0%) conducted two rounds, but the stability of responses was seldom assessed [28]. Alarmingly, five studies reported only a single round, which fails to meet the minimum requirement for a true Delphi procedure.

Controlled feedback allows participants to review the collective responses and adjust their own rankings, fostering convergence toward consensus [20, 29]. Yet, 70.0% of studies did not report the feedback mechanism, limiting transparency. Despite a high reporting rate for statistical outcomes (98.0%), interpretation is complicated by inconsistent application of Delphi methodology. Researchers are therefore advised to treat statistical results cautiously, as they do not automatically indicate stable or reliable consensus.

Overall, the review indicates widespread misinterpretation and misuse of Delphi principles in TCM syndrome diagnosis research. Adhering strictly to these core principles is necessary to distinguish Delphi studies from ordinary expert discussions.

Importance of a priori consensus definition

Defining consensus a priori is critical for scientific rigor, yet only 26.0% of reviewed studies reported this step. Most studies determined consensus merely by the degree of agreement, which is insufficient because agreement levels depend on panel size, objectives, and other contextual factors [17]. Response stability is considered a more reliable indicator of consensus [30]. Additionally, only 50.0% of studies reported whether consensus was actually reached. As the Delphi method is designed to drive group agreement, its outcome should not be mistaken for the "correct" answer; it reflects only the collective opinion of the selected panel [31]. Transparent reporting of this process is therefore essential.

Consensus can be quantified using multiple metrics, including percentage agreement, measures of central tendency, and response dispersion. However, calculating means for Likert scale data is inappropriate, as these data are ordinal rather than interval [32]. There is currently limited guidance on the optimal number of Likert scale points, and future research should aim to standardize this. Careful consideration is also needed regarding odd versus even numbers of scale points. An even number avoids a neutral midpoint, encouraging panelists to take a definitive position, which is particularly relevant when identifying positive and negative symptoms for syndrome diagnosis [33].

The practical application of the reviewed studies' results in syndrome differentiation was not assessed. Nevertheless, the final consensus may be conservative due to the exclusion of dissenting opinions. Given the generally poor quality of reporting, these results should be applied cautiously. The lack of methodological rigor and consistency hampers the advancement of TCM syndrome diagnosis, highlighting the urgent need for both standardized conduct and transparent reporting. **Table 4** presents recommendations and key considerations to guide the design and reporting of future Delphi studies in this field.

Table 4. Considerations and Recommendations for Scientific Design and Transparent Reporting of Delphi Studies

Delphi Procedure	Considerations and Recommendations
A priori criteria for panels	Clearly define criteria for potential panelists, taking into account their experience, expertise, willingness, and availability to participate.
Panel size and heterogeneity	Avoid excessively large panels to reduce logistical and data management challenges; include diverse, multidisciplinary panelists to achieve more robust and credible consensus.
Conflicts of interest	Disclose any actual or potential conflicts of interest among panel members and, if present, assign an independent researcher to coordinate the Delphi process.

Literature review	Ensure the initial structured questionnaire is informed by relevant literature to provide a solid evidence base.
Rating scale selection	Carefully choose the number of Likert scale points and consider whether to include a neutral midpoint; optional open-text fields may be added for additional comments.
Iteration with controlled feedback	Conduct three or four rounds to balance panel fatigue with the collection of meaningful results; report the controlled feedback process transparently to facilitate consensus.
Statistical results	Interpret statistical outcomes with caution, as they do not automatically indicate stability or consensus; clearly describe the methods used.

Strengths and limitations

To our knowledge, this is the first study to systematically evaluate both the implementation and reporting of the Delphi method in TCM syndrome diagnosis research, and to provide evidence-based recommendations aimed at enhancing the quality of future studies. However, the study has several limitations. Our literature search was restricted to six databases and limited to publications in English and Chinese, and we did not include gray literature, which may contain additional relevant studies. Despite this, the selected databases are widely recognized and considered sufficient to capture a representative sample of local research. Data extraction was conducted independently by four reviewers in paired assessments, achieving at least 80% interrater agreement, which strengthens the reliability of our findings. The inherent methodological uncertainty of the Delphi technique itself contributes to variability in its execution and reporting, highlighting the need for internationally standardized guidelines. Future studies should aim to establish such guidance, enabling more rigorous and consistent analyses in subsequent cross-sectional investigations. Despite these limitations, we believe our study provides meaningful and trustworthy insights.

Conclusion

The Delphi method is widely employed in TCM syndrome diagnosis research, yet its application varies considerably, and reporting quality is generally poor. Similar patterns of inconsistency and debate regarding Delphi procedures have been observed in broader health sciences research [4]. These findings underscore the urgent need for standardized guidelines to address methodological ambiguities, and for improved reporting criteria to enhance transparency. Establishing such standards will strengthen the scientific credibility of TCM syndrome diagnosis research and support more reliable and reproducible outcomes.

Acknowledgments: The authors would like to express the appreciation to Peipei Du who provided help regarding statistical analyses during the research.

Conflict of Interest: None

Financial Support: This work was supported by the 2023 Beijing University of Chinese Medicine Graduate Students Independent Research Project (Grant No. ZJKT2023087).

Ethics Statement: None

References

1. Dalkey N, Helmer DO. An experimental application of the Delphi method to the use of experts. Manag Sci. 1963;9(3):458-67. doi:10.1287/mnsc.9.3.458
2. Jones J, Hunter D. Qualitative research: consensus methods for medical and health services research. BMJ. 1995;311(7001):376-80. doi:10.1136/bmj.311.7001.376
3. McKenna HP. The Delphi technique: a worthwhile research approach for nursing? J Adv Nurs. 1994;19(6):1221-5. doi:10.1111/j.1365-2648.1994.tb01207.x
4. Shang Z. Use of Delphi in health sciences research: a narrative review. Medicine (Baltimore). 2023;102(7):e32829. doi:10.1097/MD.0000000000032829
5. Yuan B. Towards a clinical efficacy evaluation system adapted for personalized medicine. Pharmacogenomics Pers Med. 2021;14:487-96. doi:10.2147/PGPM.S304420

6. Sun DZ, Li SD, Liu Y, Zhang Y, Mei R, Yang MH. Differences in the origin of philosophy between Chinese medicine and Western medicine: exploration of the holistic advantages of Chinese medicine. Chin J Integr Med. 2013;19(9):706-11. doi:10.1007/s11655-013-1435-5
7. Jiang M, Lu C, Zhang C, Yang J, Tan Y, Lu A, et al. Syndrome differentiation in modern research of traditional Chinese medicine. J Ethnopharmacol. 2012;140(3):634-42. doi:10.1016/j.jep.2012.01.033
8. Wang Y. Suggestions on improving the methodological system of traditional Chinese medicine syndrome differentiation. J Tradit Chin Med. 2004;10:729-31. doi:10.3321/j.issn:1001-1668.2004.10.003
9. Song W, Hu S, Zhang H, Chen Z. 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. 2004;8(3):488-9. doi:10.3321/j.issn:1673-8225.2004.03.046
10. Fu H, Dai G, Gao W, Zhao Z, Liu X, Li Y. Systematic evaluation of the application of Delphi method in traditional Chinese medicine clinical syndrome. World Sci Technol (Modernization of Tradit Chin Med). 2022;24(10):4082-8. doi:10.11842/wst.20210909010
11. Wu H, Xu L, Zheng Y, Shi L, Zhai L, Xu F. Application of the Delphi method in the study of depressive disorder. Front Psychiatry. 2022;13:925610. doi:10.3389/fpsyt.2022.925610
12. Humphrey-Murto S, Varpio L, Wood TJ, Gonsalves C, Ufholz LA, Mascioli K, et al. The use of the Delphi and other consensus group methods in medical education research: a review. Acad Med. 2017;92(10):1491-8. doi:10.1097/ACM.0000000000001812
13. 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(8):684-706. doi:10.1177/0269216317690685
14. Trevelyan EG, Robinson N. Delphi methodology in health research: how to do it? Eur J Integr Med. 2015;7(4):423-8. doi:10.1016/j.eujim.2015.07.002
15. Humphrey-Murto S, Wood TJ, Gonsalves C, Mascioli K, Varpio L. The Delphi method. Acad Med. 2020;95(1):168. doi:10.1097/ACM.0000000000002887
16. Foth T, Efstathiou N, Vanderspank-Wright B, Ufholz LA, Dütthorn N, Zimansky M, et al. The use of Delphi and Nominal Group Technique in nursing education: a review. Int J Nurs Stud. 2016;60:112-20. doi:10.1016/j.ijnurstu.2016.04.015
17. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs. 2000;32(4):1008-15. doi:10.1046/j.1365-2648.2000.t01-1-01567.x
18. Duffield C. The Delphi technique: a comparison of results obtained using two expert panels. Int J Nurs Stud. 1993;30(3):227-37. doi:10.1016/0020-7489(93)90033-Q
19. Keeney S, Hasson F, McKenna HP. A critical review of the Delphi technique as a research methodology for nursing. Int J Nurs Stud. 2001;38(2):195-200. doi:10.1016/S0020-7489(00)00044-4
20. Skulmoski GJ, Hartman FT, Krahn J. The Delphi method for graduate research. J Inf Technol Educ. 2007;6(1):1-21.
21. McPherson S, Reese C, Wendler MC. Methodology update: Delphi studies. Nurs Res. 2018;67(5):404-10. doi:10.1097/NNR.0000000000000297
22. Baker J, Lovell K, Harris N. How expert are the experts? An exploration of the concept of "expert" within Delphi panel techniques. Nurse Res. 2006;14(1):59-70. doi:10.7748/nr2006.10.14.1.59.c6010
23. Jorm AF. Using the Delphi expert consensus method in mental health research. Aust N Z J Psychiatry. 2015;49(10):887-97. doi:10.1177/0004867415600891
24. Hsu CC, Sandford BA. The Delphi technique: making sense of consensus. Pract Assess Res Eval. 2007;12:10. doi:10.7275/pdz9-th90
25. Powell C. The Delphi technique: myths and realities. J Adv Nurs. 2003;41(4):376-82. doi:10.1046/j.1365-2648.2003.02537.x
26. Goodman CM. The Delphi technique: a critique. J Adv Nurs. 1987;12(6):729-34. doi:10.1111/j.1365-2648.1987.tb01376.x.
27. 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

28. 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(3):i-88.
29. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. Int J Forecast. 1999;15(4):353-75. doi:10.1016/S0169-2070(99)00018-7
30. Crisp J, Pelletier D, Duffield C, Adams A, Nagy S. The Delphi method. Nurs Res. 1997;46(2):116-8. doi:10.1097/00006199-199703000-00010
31. Walker MA, Selfe MJ. The Delphi method: a useful tool for the allied health researcher. Int J Ther Rehabil. 1996;3(12):677-81. doi:10.12968/bjtr.1996.3.12.14731
32. von der Gracht HA. Consensus measurement in Delphi studies: review and implications for future quality assurance. Technol Forecast Soc Change. 2012;79(8):1525-36. doi:10.1016/j.techfore.2012.04.013
33. Tsang KK. The use of midpoint on Likert scale: the implications for educational research. Hong Kong Teach Cent J. 2012;11:121-30.