Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

Development and validation of a quality appraisal tool for case reports in traditional Chinese medicine using the Delphi method

Duoting TAN[a, b], Hao LIANG[a, c], Yipin YU[a, c], Jin GUO[a, c], Liqin ZHONG[a, c], Zhixi HU[a, c*]

a. School of Traditional Chinese Medicine, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

b. Institute of Fundamental of Traditional Chinese Medicine, Hunan Academy of Chinese Medicine, Changsha, Hunan 410013, China

c. Hunan Provincial Key Laboratory of Traditional Chinese Medicine Diagnostics, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

ARTICLE INFO

Article history
Received 14 November 2024
Accepted 06 May 2025
Available online 25 June 2025

Keywords
Case report
Traditional Chinese medicine (TCM)
Delphi study
Quality appraisal tool
Quality rating
Reporting recommendations

ABSTRACT

Objective To develop a quality appraisal tool for case reports in traditional Chinese medicine (TCM) based on their characteristics.

Methods An extensive literature search was conducted in Chinese Biomedical Literature Database (CBM), China National Knowledge Infrastructure (CNKI), and China Science and Technology Journal Database (CSTJ), focusing on expert consensus statements and checklists for TCM case reports. Relevant items were extracted, and a Delphi method involving 34 experts was used in two rounds to rate each item on a 5-point Likert scale. Items were screened based on measures of central tendency and coordination (including total score, mean score, percentage of items rated as unimportant, and coefficient of variation). The weighted average method was used to determine item weights and construct the appraisal tool. Internal consistency was assessed using Cronbach's α coefficient. The finalized tool was pilot-tested by two reviewers independently appraising 20 case reports, with an additional four reviewers evaluating 5 of these cases to compare inter-rater consistency.

Results A total of 9 513 articles were retrieved, and 96 items from 25 articles were extracted. After two rounds of the Delphi method, 27 items across 10 domains were retained. The Cronbach's α coefficient was 0.72 in the first round (acceptable range), and 0.96 in the second round, indicating strong internal consistency. The tool was piloted by six reviewers, achieving a kappa value of 0.663 and a Kendall's coefficient of concordance of 0.845, demonstrating high consistency among reviewers.

Conclusion The developed TCM case report quality appraisal tool, consisting of 27 items in 10 domains, offers a scientific and reliable means of assessing the quality of TCM case reports. The tool showed high consistency and practical utility, and its application is expected to enhance the standardization, scientific rigor, and evidence quality of TCM case reports, facilitating the integration of traditional medical knowledge with modern evidence-based standards.

1 Introduction

Traditional Chinese medicine (TCM), with a history spanning over thousands of years, has its theoretical foundations deeply rooted in extensive clinical practice. A case report in TCM serves as a detailed narrative that records the diagnosis, therapeutic principles, prescriptions, and prognosis of specific practical cases [1]. It fully

*Corresponding author: Zhixi HU, E-mail: 003405@hnucm.edu.cn.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2025.05.004

Citation: TAN DT, LIANG H, YU YP, et al. Development and validation of a quality appraisal tool for case reports in traditional Chinese medicine using the Delphi method. Digital Chinese Medicine, 2025, 8(2): 137-146.

embodies TCM theory and practitioners' understanding of diseases, which leaves a deep impression on readers. Consequently, case reports are vital in the inheritance and exchange of clinical experiences within the TCM community [2], playing an irreplaceable role in the preservation of TCM theory [3, 4]. Numerous TCM classics, such as *Shanghan Jiushi Lun* (《伤寒九十论》, *Ninety Treatises on Cold Damage*) and *Mingyi Leian* (《名医类案》, *Classified Case Records of Famous Physicians*), are replete with such case reports.

However, in contemporary times, while the publication of case reports has seen a marked increase, there exists a notable disparity in their quality [3, 5]. We systematically reviewed articles on the quality assessment of TCM case reports and found that there were significant variations in content and format [6-14]. For instance, a study found that only 54% of case reports included examination findings and diagnosis of conventional medicine, with less than 30% detailing the rationale behind treatment modifications, and none provided information on informed consent [11].

Nevertheless, many scholars have conducted research on the reporting standards of case reports. Progress in this area has been made, with notable contributions such as GOU's development of a quality evaluation tool for case reports on the treatment of typhoid fever in *Shanghan Lun* using the Delphi method [15]. Additionally, ZHANG et al. [16] in 2016 proposed recommended reporting items for TCM case reports based on the Case Report Guidelines (CARE), and the Consensus-based Recommendations for Case Report in Chinese Medicine (CARC) was introduced in the same year [17]. These studies have significantly advanced the understanding and standardization of TCM case report writing and publication, but a comprehensive quality appraisal tool specifically designed for evaluating TCM case reports is still lacking [11]. The absence of such a tool makes it challenging to objectively assess and distinguish the quality of case reports. Therefore, this study aims to develop a quality appraisal tool for TCM case reports through a combination of literature review and the Delphi method.

2 Data and methods

2.1 Literature review

The literature review method was employed to establish the item pool of case reports. The following databases were searched from 1992 to 2022: Chinese Biomedical Literature Database (CBM), China National Knowledge Infrastructure (CNKI), and China Science and Technology Journal Database (CSTJ). The search items included the following terms: (i) "case report" OR "case study" OR "case studies" OR "case histories"; (ii) "recommendations" OR "standards" OR "quality" OR "elements" OR "design"; (iii) combination of (i) and (ii).

The inclusion criteria for the articles comprised the following aspects: (i) articles containing checklists, quality evaluation, or recommendations on case reports; (ii) the current status of TCM case reports; and (iii) any aspect of standardization of TCM case reports. The exclusion criteria included: (i) duplicate publications; (ii) literature not published in Science Citation Index (SCI), Social Science Citation Index (SSCI), Chinese Social Science Citation Index (CSSCI), national-level journals, core journals, provincial-level journals, or China National Serial Number (CN) publications; (iii) studies that were only experimental or involved data mining of case reports; or (iv) literature that only included abstracts.

Following the identification of duplicate articles, all redundant articles were removed while retaining a single representative document. Articles were read independently by two reviewers. Any disagreements were resolved through reaching a consensus in group discussions. Finally, an item pool was formed, and an expert questionnaire was developed, containing all the possible items a case report could have based on the content of the item pool. Experts were asked to score each item and provide comments during the Delphi process.

2.2 Delphi study

The Delphi technique, a widely acknowledged consensus development method in healthcare [18], was employed in this study. This method is known for its efficiency in collecting information from a group of knowledgeable people [19]. In this study, the Delphi technique was applied to assess the extent of agreement and resolve disagreement in case report items.

Before the Delphi study, a clinical expert group was formed. The selection of experts was based on three criteria: (i) having more than 10 years of clinical experience in TCM; (ii) being affiliated with the Department of TCM Internal Medicine, Department of Diagnosis of Traditional Chinese Medicine, or Department of Acupuncture; and (iii) having published case reports in academic journals. The demographic characteristics of the experts were collected, including age, gender, professional rank, major field, and years of working experience. Descriptive analysis was performed to analyze these characteristics.

The survey was conducted in two rounds from March to July 2022. Face-to-face interviews with some experts were held before the first round to discuss the content of the questionnaire. The online survey, hosted on the Wenjuanxing platform (https://www.wjx.cn/), took approximately 15 to 20 min to complete. The research team carefully reviewed the items extracted from the CARC

guideline and other articles [5, 17, 20-23]. The first round of survey collected demographic information and experts' opinions on the importance of each item. The reasons for judgments were also provided in the questionnaire. These items covered five domains including: the basic information of patients, clinical information, treatment procedure, outcome, and discussion. Experts were invited to suggest additional items that they deemed necessary for TCM case reports. With reference to the item inclusion and exclusion criteria established by the team during the earlier researches of scale development [24], items that met the following standards were included in the second round: total score ⩽ 71, mean ⩽ 4, unimportant percentage ⩾ 0.15, and coefficient of variation (CV) ⩾ 0.15. The rest were included in the final appraisal tool, as 70% of experts considered them as very important or important, which were considered to have reached a consensus in this study (unimportant percentage: the ratio of experts giving the item a low score (0, 1, or 2) to total experts; a higher ratio indicates the item is less important. Coefficient of variation: the degree of dispersion of the experts' evaluation of the importance of an item; the smaller the coefficient of variation, the smaller the degree of dispersion and the higher the degree of harmonization of experts' opinions). The importance of each item was rated on a 5-point Likert scale for both surveys, including 1 (not important at all), 2 (not very important), 3 (neutral), 4 (important), and 5 (very important). All the participants were given 2 – 3 weeks to respond to the questionnaire. In the second round, the items were excluded if they failed to meet the following standards: total score ⩾ 67, mean ⩾ 4, unimportant percentage ⩽ 0.15, and CV ⩽ 0.15.

2.3 Data analysis

Data were analyzed using SPSS 23.0. Descriptive statistics were applied to analyze the demographic information of the experts. The degree of concentration and coordination among the experts was assessed using total score, mean, unimportant percentage, and coefficient of variation. The higher the mean score of an item, the more frequently it received full scores, indicating its greater importance and the experts' concentration on it. The coordination degree was measured by the coefficient of variation, which reflected the variation among the experts. Based on previous literature [25], items selected by more than 70% of experts were considered to have reached a consensus in this study (70% of experts considered them as very important or important). The independent sample *t* test was used to compare whether the scores of the items were significantly different from the mean score. Cronbach's α and split-half reliability were employed to assess the internal consistency of the questionnaire. A score of 0.7 or higher was indicative of high internal consistency. Kendall's coefficient of concordance (W) was used to assess the reliability of the results. Weight coefficients (WC) for each item were calculated using the weighted average method: WC = Σ (value × weight)/sum of weight.

2.4 Piloting the appraisal scale in 20 case reports

The 10-criteria appraisal scale was pilot-tested by a group of six reviewers. This group included four evaluators specializing in TCM, consisting of one expert, one clinician, and two graduate students specializing in TCM diagnosis. To assess the general applicability of the scale, the remaining two evaluators were professionals outside the field of TCM. Two reviewers evaluated 20 case reports; another four reviewers assessed 5 of these cases (corresponding to No. 1, No. 4, No. 17, No. 11, and No. 20) (specific details of the medical cases are provided in Supplementary Table S1). They independently appraised the methodological quality of the case reports using the 10-criteria scale. The case reports were selected from medical journals, such as *Chinese Journal of Integrated Traditional and Western Medicine*, *Journal of Traditional Chinese Medicine*, and *World Journal of Traditional Chinese Medicine*. The consistency of the rating results was evaluated using the mean of items, sum of ranks, and the kappa value.

The scores for each item were recorded, and the case reports were classified into three levels: A (total score ⩾ 80), B (total score 61 – 79), and C (total score ⩽ 60). Data analysis included calculating total scores, mean ± standard deviation (SD), kappa values, and Kendall's W. This analysis aimed to test the consistency of the reviewers' evaluations of items and case reports, as well as the applicability of the appraisal tool.

3 Results

3.1 Literature research

A total of 9 513 articles were retrieved. After a meticulous review of titles, 9 322 articles were deemed irrelevant and excluded, leaving 191 articles for preliminary inclusion. Upon thorough examination of the full texts, 20 articles and 5 ancient medical books, including guidelines for reporting TCM case reports were identified [1, 3, 5, 8, 9, 11, 15, 16, 21, 23, 26-35] (Figure 1). Additionally, based on the preliminary literature review [21], five ancient medical case works addressing the standards and formats of case reports were included as well. These include *Hanshi Yitong* (《韩氏医通》, *Han's Clear View of Medicine*) by Mao HAN, *Maiyu* (《脉语》, *Essence of Pulse Diagnosis*) by Kunwu in the Ming Dynasty, the "Ten Questions in Diagnosis" in *JingYue QuanShu* (《景岳全书》, *Complete Works of Jingyue*), the "Discussion of Disease Format" by Chang

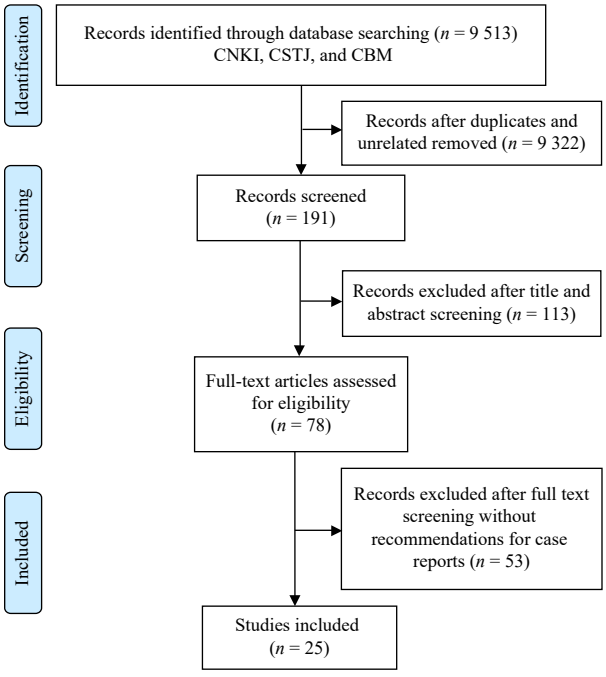

Figure 1 Literature retrieval flowchart

YU in *Yuyi Cao* (《寓意草》) during the Qing Dynasty, and "Eight Principles of Medical Case Report" designed by Lianchen HE. Three studies on the quality evaluation of TCM case reports cited the CARE Guidelines, CARC Guidelines, and the Joanna Briggs Institute (JBI) Case Report Quality Evaluation Tool [8, 9, 11]. Two researchers read the articles and extracted a total of 96 case report items. The results showed that the most frequently reported items included prescription composition (21), demographic information (20), chief complaint and present medical history (19), TCM diagnostic information (19), and symptom or tongue and pulse condition (18). Items deemed unsuitable for quality evaluation, such as "doctor's title and position" "author's qualifications" and "acknowledgments" were deleted. Based on the results of the literature study and several group discussions, 33 items about case reports were collected through literature review and group consultation.

3.2 Delphi study

The questionnaires were distributed to 17 experts who met the inclusion criteria. All participants completed the first round of the survey. In the second round, 17 questionnaires were sent to experts, and 16 experts completed the survey. Table 1 provides the demographic details of the experts.

Table 2 and 3 present the content of items in each round. In the first round, experts rated 33 items. Following this, 11 items and 7 new items (Q8 – Q14) suggested by the experts were included in the second round, while the remaining 22 items were directly included in the final list as the experts unanimously agreed on their importance. After the second round of the questionnaires, 5 items met the criteria and were included in the final list alongside the previously selected 22 items. Ultimately, a total of 27 items in 10 domains were retained in the final appraisal tool (Table 4).

The concentration and coordination indices of the experts were presented in Supplementary Table S2 and S3. The average scores of all items ranged from 3.47 to 4.94. In the first round, the highest scores were observed in the categories of "chief complaint and present illness" "symptoms and signs" "TCM four diagnosis results" and "TCM diagnosis results and rationale" (4.94), with a full score rate of 94.1%. This was followed by "therapeutic principle" and "herbal intervention" (4.88). The lowest scores were in "informed consent form" (3.47), "quality control standards of herbal ingredients" (3.59), and "manufacturing procedure of herbal ingredients" (3.76). The items with the highest agreement rate among experts were "chief complaint and present illness" "symptoms and signs" "TCM four diagnosis results" and "TCM diagnosis results and basis", followed by "TCM treatment principles" and "herbal intervention". The level of agreement among experts was low in these items: "informed consent form" (0.30), "quality standards of Chinese herbal medicine" (0.27), and "principles of acupuncture point selection" (0.25). Weight coefficients for each item were calculated using the weighted average method. To facilitate the application of the tool, the weight results were multiplied by 10 and rounded to the nearest whole number. We have simplified the scale into a 10-question appraisal scale for ease of use (Table 5).

Table 6 illustrates that the Cronbach's α for the first and second rounds were 0.72 and 0.96, respectively. The split-half reliability was 0.637 and 0.91 ($P < 0.05$), indicating a high degree of coordination and reliability.

Table 1 The demographics of experts

Round	Age (year)	Gender (*n*)		Ranking (*n*)				Major (*n*)			Working experiencing (*y*)		
		Male	Female	Professor	Associate professor	Lecturer	Assistant	TCM diagnosis	Acupuncture	TCM internal medicine	1 – 10	11 – 20	> 20
First	50.8 ± 8.76	14	3	14	1	1	1	5	1	11	2	3	12
Second	47.0 ± 9.05	16	1	10	3	3	1	7	4	6	3	6	8

Table 2 Items for expert consensus in the first round Delphi survey

No.	Item
Q1	The words “case report” or “medical record” appear in the title
Q2	Abstract
Q3	Keywords
Q4	The reason for reporting this case
Q5	Informed consent
Q6	Patient’s information (surname, gender, age, and date of consultation)
Q7	Main complaints and present condition
Q8	Symptoms and signs
Q9	Examination findings of four diagnosis (observation, auscultation, interrogation, and palpation)
Q10	Medical history and family inheritance history
Q11	The differentiation of TCM pattern and diagnosis rationale
Q12	Differential diagnosis in TCM
Q13	Examination findings and diagnosis of conventional medicine
Q14	Differential diagnosis in conventional medicine
Q15	Therapeutic principle
Q16	Herbal intervention (composition, dose of ingredients, and treatment courses)
Q17	Manufacturing procedure of herbal ingredients
Q18	Quality control standards of herbal ingredients
Q19	The name/dosing/treatment courses of Chinese proprietary medicine
Q20	Acupuncture intervention (name, location of points, and operating procedure)
Q21	Report the rationale of selected points in acupuncture
Q22	Needle type (diameter, length, or material)
Q23	Moxibustion intervention (name, location of the points, operation procedure, and technique)
Q24	Report the rationale of selected points for moxibustion
Q25	Report other external treatment procedures in detail
Q26	Use the widely recognized gold standard assessment criteria
Q27	Follow-up visit and consequent
Q28	Advice and precautions on diet, emotions and living
Q29	Examination findings (figures and tables) are relevant to this case
Q30	Narrative descriptions related to this case
Q31	Take-away messages in this case
Q32	The rationale of prescription and TCM theory in this case
Q33	Literatures are relevant to this case

3.3 Piloting the newly developed quality assessment tool

The scoring results were detailed in Supplementary Table S4 and S5. Two case reports were rated as Grade A, with average scores of 90 and 86.5, respectively, and were published in *Journal of Traditional Chinese Medicine* and *Journal of Jinan University (Natural Science & Medicine Edition)*. Three case reports were rated as C-level (the lowest level). Two of them were posted on WeChat official accounts, while another was published in *Journal of Shanxi Traditional Chinese Medicine*. The highest score rates were observed in item 3 (description of medical history), and item 6 (treatment intervention). Notably, none of the case reports included an informed consent form. The item with the highest standard deviation was item 9 (discussion) (5.34), indicating the largest score difference in this section. In contrast, the items with the smallest score difference were item 2 (informed consent) and item 5 (diagnosis of conventional medicine). The standard deviation values revealed that the No. 5 case achieved the highest level of agreement (3.54), while the No. 2 case report had the lowest level of agreement (6.89), followed by the No. 1 case report (6.62). This suggests that higher-quality case reports tend to have greater consistency in the evaluation scores.

The ratings by the reviewers were then compared. An interrater reliability analysis using the kappa statistic (SPSS 23.0) was conducted to assess the consistency of scoring between the two reviewers. The kappa value was 0.663, indicating a moderate level of agreement on consistency among the evaluation results (Table 7).

Six reviewers appraised five case reports using the 10-criteria appraisal scale. Two case reports were rated as “excellent” by all reviewers. The No. 1 case report had the

Table 3 Items for expert consensus in the second round Delphi survey

No.	Item
Q1	Differential diagnosis in conventional medicine
Q2	Report the rationale of selected points in acupuncture
Q3	Needle type (diameter, length, or material)
Q4	Report the rationale of selected points for moxibustion
Q5	Report other external treatment procedures in detail
Q6	Advice and precautions on diet, emotions and living
Q7	Literatures are relevant to this case
Q8	Social life history and constitution
Q9	Menstruation, vaginal discharge, and pregnancy (female only)
Q10	Standardized terminology
Q11	Follow the recommendation guidelines of TCM case reports
Q12	Elaboration on rationale of prescription
Q13	Changes in symptoms (pulse and tongue) during the treatment period
Q14	Using objective assessment criteria (Xray, CT, or other examination findings)
Q15	Informed consent
Q16	Manufacturing procedure of herbal ingredients
Q17	Quality control standards of herbal ingredients
Q18	Narrative descriptions related to this case

Table 4 Quality appraisal tool of TCM case report

Item	Domain	Content	Weight score
1	Complete title and abstract	The words “case report” or “medical record” appear in the title Abstract Keywords The reason to report this case	10
2	Informed consent	Informed consent	3
3	Complete medical history	Patient’s demographic information Main complaints and present condition Symptoms and signs Medical history and family inheritance history	15
4	Diagnosis of TCM and rationale	Examination findings of four diagnosis (observation, auscultation, interrogation, and palpation) The differentiation of TCM pattern and diagnosis rationale Differential diagnosis in TCM	12
5	Diagnosis of conventional medicine	Examination findings and diagnosis of conventional medicine	3
6	Complete treating process	Therapeutic principle Herbal intervention (composition, dose of ingredients, and treatment courses) Manufacturing procedure of herbal ingredients The name/dosing/treatment courses of Chinese proprietary medicine Acupuncture intervention (name, location of points, and operating procedure) Moxibustion intervention (name, location of the points, operation procedure, and technique)	15
7	Outcome assessment and description	Use the widely recognized gold standard assessment criteria Examination findings (figures and tables) are relevant to this case	12
8	Follow up and the others	Follow-up visit and consequent Advice and precautions on diet, emotions and living Narrative descriptions related to this case	10
9	Discussion	Take-away messages in this case The rationale of prescription and TCM theory in this case	15
10	Normative evaluation	Standardized terminology Follow the recommendation guidelines of TCM case reports	5

Table 5 Quality appraisal tool for TCM case report (revised version)

Item	Question	Yes	No	Not clear
1	Were there complete title and abstract?	2☆	0	1☆
2	Did the patient give informed consent?	1☆	0	/
3	Were there clear reporting of medical history and clinical findings of the patient?	3☆	0	2☆
4	Was the diagnosis clarified (including disease differentiation and pattern identification of TCM)?	2☆	0	1☆
5	Was there a diagnosis by conventional medicine?	1☆	0	0
6	Did the report have detailed descriptions of each intervention?	3☆	0	2☆
7	Were the outcomes clearly reported?	2☆	0	1☆
8	Were there any follow-up visits and precautions/advice for the patient?	2☆	0	1☆
9	Was there a discussion/comment?	3☆	0	2☆
10	Was the report written in accordance with CARC guidelines?	1☆	0	0

Yes, full compliance. No, non-compliance. Not clear, incomplete compliance. 0, non-compliance with the criteria. 1☆, general importance. 2☆, moderate importance. 3☆, critical importance. /, this option is not available.

Table 6 Reliability testing for two-round questionnaires

Round	Number	Cronbach's α	Split-half reliability
First round	33	0.72	0.637
Second round	14	0.96	0.910

Table 7 The kappa value of 20 case reports evaluated by two reviewers

Kappa value	Asymptotic standard error	*P* value
0.663	0.172	< 0.000 1

lowest mean score (58.67). The case report No. 4 had the highest mean score (88.30), followed by No. 5 (87.20). The Kendall's W of the evaluation results was 0.845 > 0.800 (Table 8), indicating a high level of consistency in the evaluation of the five case reports among the reviewers. The test results also showed significant differences in the scores of case reports with lower quality, while there was a higher consistency in scoring those with higher quality. The items with the largest score differences were domain 1 (complete title and abstract), domain 9 (discussion), and domain 6 (complete treating process).

Table 8 Kendall's W of six evaluators' rating results

Case No.	Kendall's W	Chi-square	Degree of freedom	*P* value
6	0.845	20.269	4	< 0.000 1

4 Discussion

The CARE guideline is a widely recognized and adopted framework for writing high-quality case reports. It provides a comprehensive and flexible approach that ensures that all relevant information is included, improving the transparency and credibility of case reports. CARC guidelines share the similar advantages and characteristics, focusing on clarity, transparency, and best practices in TCM case report writing. The guideline covers all aspects that should be included in a case report, emphasizing its comprehensiveness, while the quality assessment tool focuses more on the core content of case reports.

4.1 Development and validation of the TCM case report appraisal tool

This study followed the standard procedures of the Canada Institute of Health Economics (IHE) methodology papers to develop an appraisal tool for TCM case reports [25, 36]. Through a systematic review of 25 relevant literature sources, we identified 33 candidate items, which were subsequently refined using a Delphi study with high expert consensus (Cronbach's α > 0.85). The finalized 27-item scale prioritized clinical relevance, assigning the highest scores to domains including main complaints, diagnostic rationale, herbal interventions, and take-away messages—a pattern consistent with established frameworks such as the CARC Guidelines and the Treatise on Febrile Diseases Quality Tool [29]. Notably, while items like informed consent and manufacturing procedures received lower initial ratings, ethical considerations from expert feedback led to the retention of informed consent as a mandatory component [37, 38].

4.2 Analysis of research results

Outcome assessment is the core of medicine and the vitality of TCM. Both successful and unsuccessful cases have their own value, especially for cases where the initial efficacy is poor but improves after adjusting the prescriptions. Compared with the previous reporting standard [17], this scale categorized efficacy evaluation into diagnostic gold standard, patient-reported outcomes, and biochemical indicators, bridging the gap between TCM's narrative tradition and modern evidence-based requirements [39]. We expanded the applicability of the scale compared with previous appraisal studies [5, 16, 17, 21, 31, 35, 40],

and this tool can score each TCM case report and give a quality level.

The weighted average method was used to calculate the weight coefficients of each item, and the results showed no significant differences among them except for "informed consent" (2.8). Generally, only one to two treatment methods (herbal treatment or acupuncture) are used in a single case report. Therefore, the scores of the four items "herbal treatment" "Chinese patent medicine treatment" "acupuncture treatment" and "moxibustion treatment" were integrated into a score of 10. Quality appraisal tool should focus on evaluating the core content of case reports. Hence, the research team discussed and adjusted the scores from 15 to 10 for items that are not related to diagnosis and treatment procedures, such as "title and abstract".

The huge differences were mainly due to the following reasons. (i) The content of high-quality case reports is more complete compared with low-quality case reports, the latter having arbitrary titles and rarely containing keywords and abstracts, resulting in significant differences in the scores of titles and abstracts. (ii) Inconsistent understanding of the items was revealed. For example, "take-away message of case reports" can include medication experience, ancient literature, prescription basis, or the characteristics and difficulties of the case. Due to the broad content, different individuals may have different understandings, making it difficult to reach a consensus in scoring. (iii) Lack of certain background knowledge was noted. For evaluators who are not specialized in TCM, it is difficult to understand the concepts in the items.

4.3 Clinical and research implications

First, according to the CARC, standardized titles, abstracts, and keywords are inevitable trends in digital development. This part of the content provides convenience for the storage, classification, and retrieval of case reports, addressing a critical need in the era of big data. Second, by providing quantifiable metrics for case report quality, it enables systematic comparison across TCM modalities, which is a prerequisite for establishing evidence-based practice guidelines, and is beneficial for the objectification and standardization of TCM case reports. Third, high-quality case reports also have strong educational value. The emphasis on take-away messages and treatment rationales enables the transmission of clinical wisdom among TCM practitioners. Last, the scoring system identifies high-value case reports for meta-analyses, particularly those documenting treatment failures or dose adjustments—historically under-reported yet clinically invaluable data.

4.4 Limitations and future research

However, this research also has certain limitations. First, it was suggested that experts should be selected from a wide spectrum of vested interests [41]. The convenience sampling of experts from Hunan and Guangdong provinces (versus international representation) may introduce regional bias, though our panel size aligns with Delphi study recommendations [42, 43]. Second, initial feedback highlighted scoring complexity, prompting development of a streamlined 10-question version using JBI methodology [44]. Third, this study only measured the reliability of the scale and failed to evaluate the content validity of the scale. Some experts also pointed out that "truthfulness does not represent practicality". High-scoring case reports may not indicate high practical value. This needs to be further considered in future research.

5 Conclusion

This study addresses the necessity for a critical appraisal tool in the realm of TCM case reports. Given that high-quality experimental studies are limited, case reports may represent the best available evidence to inform clinical practice. Developed in alignment with the CARC guidelines, the primary innovation of this study resides in the novel application of the Delphi method to develop a quality assessment instrument for TCM case reports. Post-validation analyses demonstrated robust internal consistency and high practical applicability in clinical settings, establishing methodological advancements in evidence evaluation for TCM. Implementation is projected to enhance evidence traceability in real-world TCM practice while enabling structured data mining for syndrome differentiation research.

Fundings

General Program of National Natural Science Foundation of China (82274412), Natural Science Foundation of Hunan Province (2020JJ4062), and Science and Technology Innovation Program of Hunan Province (2020RC 2061).

Competing interests

Hao LIANG and Zhixi HU are editorial board members for *Digital Chinese Medicine* and were not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

References

[1] RILEY DS, BARBER MS, KIENLE GS, et al. CARE guidelines for case reports: explanation and elaboration document. Journal

of Clinical Epidemiology, 2017, 89: 218–235.
[2] HE LL, SHEN X, XIE HB. Critical reading of traditional Chinese medical records. Chinese Medicine and Philosophy, 2021, 42(14): 72–76.
[3] JIAO S, CUI JY, LI CM. The status, significance and existing problems of TCM case report papers. Journal of Chinese Medicine, 2016, 57(17): 1472–1476.
[4] YU L. Preliminary study on the medical value of medical records before Song Dynasty. Culture of Traditional Chinese Medicine, 2020, 15(5): 71–82.
[5] WANG J, WU LQ, MA XY, et al. Thoughts and suggestions on the standard of TCM medical record reporting. Chinese Journal of Evidence-Based Medicine, 2021, 21(11): 1347–1352.
[6] ZHANG LL, ZHANG RS, YAO NL, et al. Research status, quality problems and countermeasures of famous old Chinese medical records. World Science and Technology-Modernization of Chinese Medicine, 2020, 22(5): 1608–1613.
[7] CHEN Z, XUN YQ, JIA XD, et al. Evaluation on methodological quality of TCM case series in diagnosis and treatment of novel coronavirus pneumonia. Chinese Journal of Evidence-Based Cardiovascular Medicine, 2021, 13(8): 926–931.
[8] JIA XD, XUN YQ, CHEN Z, et al. Study on the report quality of TCM treatment of novel coronavirus pneumonia cases. Chinese Journal of Evidence-Based Cardiovascular Medicine, 2021, 13(11): 1304–1309.
[9] XIAO L, NIU LQ, FU SF, et al. Evaluation of case report quality of integrated Chinese and western medicine in treating COVID-19 based on CARC. Tianjin Traditional Chinese Medicine, 2021, 38(5): 548–555.
[10] WEI C, XU Y, DENG XG, et al. Assessment of reporting completeness in acupuncture studies on patients with functional constipation using the STRICTA guidelines. Complementary Therapies in Medicine, 2022, 70: 102849.
[11] TIAN H, SILVER H, YANG H, et al. Evaluation of the quality of TCM medical records in recent 5 years based on case report standard guidelines. World Science and Technology-Modernization of TCM, 2021, 23(9): 3408–3414.
[12] AN GH, TANG XT, CHEN YL, et al. Reporting characteristics of case reports of acupuncture therapy with CARE guidelines. Chinese Journal of Integrative Medicine, 2018, 24(1): 56–63.
[13] JIAO S, DENG Y, LI CM. Selection strategy of TCM case report. Journal of TCM, 2016, 57(18): 1560–1564.
[14] JIAO S, CHAI QY, LI CM. Description of diagnosis and treatment information in TCM case report papers. Journal of Chinese Medicine, 2016, 57(19): 1653–1657.
[15] GOU XC, SONG JS, GAO C, et al. Establishment of medical record quality scoring system for Treatise on Februarytis. Liaoning Journal of Traditional Chinese Medicine, 2012, 39(7): 1211–1213.
[16] ZHANG K, HAN M, LIANG N, et al. Norms for writing and reporting TCM medical records. Beijing Traditional Chinese Medicine, 2016, 35(1): 59–62.
[17] FU SF, CHENG CW, ZHANG L, et al. Consensus-based recommendations for case report in Chinese medicine (CARC). Chinese Journal of Integrative Medicine, 2016, 22(1): 73–79.
[18] MCPHERSON S, REESE C, CECILIA WENDLER M. Methodology update: Delphi studies. Nursing Research, 2018, 67(5): 404–410.
[19] HUMPHREY-MURTO S, VARPIO L, WOOD TJ, et al. The use of the Delphi and other consensus group methods in medical education research: a review. Academic Medicine, 2017, 92(10): 1491–1498.
[20] FU SF, KUN W, ZENG XX, et al. Urgent need to improve the quality of case report in traditional Chinese medicine: assessment on reporting quality of 3 417 cases. Chinese Journal of Integrative Medicine, 2016, 22(6): 473–480.
[21] GAO DP, WANG YH, ZHANG RS, et al. Review on the standardization of TCM medical records. China Information Journal of Traditional Chinese Medicine, 2018, 25(5): 131–135.
[22] WANG H, YANG QL, WANG ZX, et al. Suggestions on writing standard of TCM parallel medical records. Modern Chinese Medicine Clinic, 2019, 26(3): 6–10.
[23] YU Q, LI JH, LI ZY, et al. Research on ontology based TCM medical case knowledge service and sharing system construction. China Digital Medicine, 2017, 12(5): 103–105.
[24] SHOU XT. Research and product a diagnostic scale of integrated traditional Chinese and western medicine for different pathological factors of chronic heart failure. Changsha: Hunan University of Chinese Medicine, 2020.
[25] Development of a quality appraisal tool for case series studies using a modified Delphi technique. Available from: https://www.ihe.ca/advanced-search/development-of-a-quality-appraisal-tool-for-case-series-studies-using-a-modifieddelphi-technique.
[26] LI JH. The Study on the Quality Evaluation to TCM Treatment Literature and Model Application. Beijing: China Academy of Chinese Medical Sciences, 2009.
[27] YAO L, CHEN YL, WANG Q, et al. Interpretation of Reporting Guidelines for Case Reports. Chinese Journal of Evidence-Based Pediatrics, 2014, 9(3): 216–219.
[28] SUN GR. Thoughts and methods of medical record research and writing. Journal of Beijing University of Chinese Medicine, 2013, 20(5): 3–6.
[29] GAO C, SONG JS, XIONG J, et al. Research on the quality standard of medical record publication of Treatise on Febrile Diseases based on Delphi method. Chinese Journal of Traditional Chinese Medicine, 2012, 27(2): 296–298.
[30] HE HE. Research on the core value of TCM medical records. Changsha: Hunan University of Traditional Chinese Medicine, 2014.
[31] YANG H, FEI YT, LIU JP. Report method of TCM clinical medical records and expert experience—design of case report. Journal of Chinese Medicine, 2008(3): 215–217.
[32] YANG JY. On the writing and editing of medical records – on the academic thoughts of Professor Guangrong SUN, Master of Chinese medicine, on studying typical medical records of Chinese medicine. Chinese Association of Chinese Medicine Annual Conference 2013 – Chinese Medicine Continuing Education Forum, 2013: 7.
[33] YUAN T, PAN YX, GU WJ, et al. Evaluation on the quality of case reports published in Pediatrics Journal of Chinese Science Citation Database from 2010 to 2016. Chinese Journal of Evidence-Based Pediatrics, 2017, 12(1): 27–32.
[34] ZHOU YF, GU Y, HU Y, et al. JBI Center for evidence-based health care quality evaluation tool for different types of studies: case report and case series quality evaluation. Journal of Advanced Nursing, 2018, 33(4): 310–312.

[35] CHEN HL, LU JG, SHI ZM, et al. Discussion on the standardization of TCM medical records. China Information Journal of Traditional Chinese Medicine, 2007, 14(11): 6–7.
[36] MA LL, WANG YY, YANG ZH, et al. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? Military Medical Research, 2020, 7(1): 7.
[37] COCANOUR CS. Informed consent—it's more than a signature on a piece of paper. American Journal of Surgery, 2017, 214(6): 993–997.
[38] TAN DT, YU YP, LIANG H, et al. Construction of TCM medical record characteristics and quality evaluation system. Chinese Journal of Traditional Chinese Medicine, 2022, 37(9): 4924–4927.
[39] GUI T. From history to medical records: a study on the narrative origins of ancient Chinese medical records. Culture of Traditional Chinese Medicine, 2022, 17(3): 255–263.
[40] ZHANG WY, XIONG J, CHEN RX. Evaluation of case report quality of chinese acupuncture and moxibustion based on Chinese medical case report. Chinese Journal of Acupuncture and Moxibustion, 2018, 38(10): 1129–1132.
[41] NASA P, JAIN R, JUNEJA D. Delphi methodology in healthcare research: how to decide its appropriateness. World Journal of Methodology, 2021, 11(4): 116–129.
[42] ROMERO-COLLADO A. Essential elements to elaborate a study with the (e) Delphi method. Enfermeria Intensiva, 2021, 32(2): 100–104.
[43] NJUANGANG S, LIYANAGE C, AKINTOYE A. Application of the Delphi technique in healthcare maintenance. International Journal of Health Care Quality Assurance, 2017, 30(8): 737–754.
[44] MUNN Z, BARKER TH, MOOLA S, et al. Methodological quality of case series studies: an introduction to the JBI critical appraisal tool. JBI Evidence Synthesis, 2020, 18(10): 2127–2133.

基于德尔菲法中医病例报告质量评价工具的构建与验证

谭朵廷[a,b], 梁昊[a,c], 余怡嫔[a,c], 郭瑾[a,c], 钟俐芹[a,c], 胡志希[a,c*]

a. 湖南中医药大学中医学院, 湖南 长沙 410208, 中国
b. 湖南省中医药研究院中医基础研究所, 湖南 长沙 410013, 中国
c. 湖南中医药大学中医诊断学湖南省重点实验室, 湖南 长沙 410208, 中国

【摘要】目的 基于中医病例报告特点开发一种中医病例报告的质量评估工具。**方法** 研究首先系统检索了中国生物医学文献数据库（CBM）、中国知网（CNKI）以及中国科技期刊数据库（CSTJ）中有关中医病例报告专家共识及相关清单的文献，筛选并提取了相关条目。在此基础上，邀请 34 位专家采用德尔菲法进行了两轮咨询，专家根据条目重要性采用李克特五级量表评分，并依据集中度与协调度（包括总分、均分、不重要率和变异系数）筛选条目。采用加权平均法确定各条目权重，形成质量评价工具。工具的内部一致性通过克隆巴赫系数进行评价。最终工具由两位评审员独立评价 20 份病例报告，另有四位评审员对其中 5 份病例进行重复评价，以比较一致性。**结果** 文献检索共纳入 9 513 篇文献，最终提取 25 篇文献、96 个相关条目。德尔菲法两轮问卷分析后，最终保留了 10 个领域、共 27 个条目。第一轮克隆巴赫系数为 0.72，处于可接受范围；第二轮系数为 0.96，表明具有高度一致性。工具经六位评审员试用，kappa 值为 0.663，Kendall 协调系数为 0.845，说明评分结果高度一致。**结论** 本研究制定的中医病例报告质量评价工具涵盖 10 个领域、27 个条目，具有良好的科学性和实用性。工具的一致性验证结果良好，有望推动中医病例报告的规范化、标准化，提升学科证据质量，促进传统医学经验与现代循证医学的融合。

【关键词】病例报告；中医；德尔菲法；质量评估工具；质量评级；报告建议