International Journal of Nursing Sciences 12 (2025) 130–136

HOSTED BY

ELSEVIER

Contents lists available at ScienceDirect

International Journal of Nursing Sciences

journal homepage: http://www.elsevier.com/journals/international-journal-of-nursing-sciences/2352-0132

Review

Evaluation of the measurement properties of online health information quality assessment tools: A systematic review

Yating Li [a], Hui Ouyang [a], Gan Lin [a], Yichao Peng [b], Jinghui Yao [b], Yun Chen [c, *]

[a] *School of Nursing, Southern Medical University, Guangzhou, Guangdong, China*
[b] *Center for Orthopedic Surgery, The Third Affiliated Hospital of Southern Medical University, Guangzhou, Guangdong, China*
[c] *Department of Outpatient, The Third Affiliated Hospital of Southern Medical University, Guangzhou, Guangdong, China*

ARTICLE INFO

Article history:
Received 29 June 2024
Received in revised form
27 December 2024
Accepted 18 February 2025
Available online 21 February 2025

Keywords:
Consumer health information
Data accuracy
Internet
Quality indicators
Systematic review

ABSTRACT

Objectives: This study aimed to evaluate the measurement properties and methodological quality of instruments developed to evaluate the quality of online health information.
Methods: In this study, a systematic search was conducted across a range of databases, including the China National Knowledge Infrastructure (CNKI), Wanfang, China Science and Technology Journal (VIP), SinoMed, PubMed, Web of Science, CINAHL, Embase, the Cochrane Library, PsycINFO, and Scopus. The search period spanned from the inception of the databases to October 2023. Two researchers independently conducted the literature screening and data extraction. The methodological quality of the included studies was assessed using the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) Risk of Bias checklist. The measurement properties were evaluated using the COSMIN criteria. The modified Grading, Recommendations, Assessment, Development, and Evaluation (GRADE) system was used to determine the quality grade.
Results: A total of 18 studies were included, and the measurement properties of 17 scales were assessed. Fifteen scales had content validity, three had structural validity, six had internal consistency, two had test-retest reliability, nine had interrater reliability, one had measurement error, six instruments had criterion validity, and three scales had hypotheses testing for construct validity; however, the evaluation of their methodological quality and measurement properties revealed deficiencies. Of these 17 scales, 15 were assigned a Level B recommendation, and two received a Level C recommendation.
Conclusions: The Health Information Website Evaluation Tool (HIWET) can be temporarily used to evaluate the quality of health information on websites. The Patient Education Materials Assessment Tool (PEMAT) can temporarily assess the quality of video-based health information. However, the effectiveness of both tools needs to be further verified.

What is known?

- The Internet has become the primary source of health information for the public.
- The quality of health information available on the Internet is a matter of concern, as inaccurate health information can potentially cause harm to individuals.
- Current evidence on the measurement properties of instruments intended to evaluate the quality of online health information is scarce. Therefore, additional research is urgently needed to assess their validity.

What is new?

- Seventeen scales have been identified for assessing the quality of online health information. The Health Information Website Evaluation Tool (HIWET) can be temporarily used to evaluate the quality of health information on websites. The Patient Education Materials Assessment Tool (PEMAT) can temporarily assess the quality of video-based health information.
- All the identified scales necessitate additional research and validation to substantiate their psychometric properties.

* Corresponding author.
E-mail address: chenyun88@smu.edu.com (Y. Chen).
Peer review under responsibility of Chinese Nursing Association.

https://doi.org/10.1016/j.ijnss.2025.02.015

1. Introduction

Health information encompasses information closely related to physical and mental health, including general health status, specific disease information, mental health, and infectious disease prevention [1]. The Internet, characterized by its accessibility, ease of use, and affordability, has emerged as the primary source of health information [2]. Reports indicate China's Internet user base has exceeded 1.079 billion [3]. A significant 73% of these users access health science content via mobile devices [4].

The Internet has made the speed and patterns of health information dissemination unpredictable, and health care and information professionals no longer act as gatekeepers of such information nor verify its authenticity before it is disseminated to the public [5]. Unfortunately, misinformation often outstrips accurate information regarding dissemination speed, reach, and pervasiveness [6]. Such erroneous health information can incite public panic, exacerbate physical and mental health burdens, impede the delivery of medical services, and compromise the provision of high-quality and timely health care [7]. Consequently, the public is deeply concerned with the quality of health information online [8].

However, assessing the quality of online health information often leaves the public feeling overwhelmed and uncertain [9], highlighting the need for health information quality indicators [10]. Consequently, researchers have developed instruments to perform this function; however, the reliability of these instruments has rarely been validated. Jadad and Gagliardi [11] identified 47 rating instruments for health information in a survey but did not report interrater reliability or construct validity. In a follow-up review [12], Gagliardi and Jadad discovered that many of these tools were no longer available, and of the new tools identified, only five provided assessable information, with none having been validated. Further evaluation by Breckons et al. [13] of tools designed to assess complementary medicine information online reported only two tools had been tested for reliability.

Previous studies [11–13] have largely confined their scope to literature reviews, neglecting an in-depth examination of the psychometric properties of the instruments used to assess the quality of online health information. This gap is problematic, as without valid and comprehensive comparisons, identifying the most suitable tools for health information quality assessment is challenging. The Consensus-based Standards for Selecting Health Measurement Instruments (COSMIN) [14] guidelines provide a comprehensive framework for evaluating tool quality, focusing on methodology and measurement properties. This study applied the COSMIN guidelines to assess the quality of online health information evaluation tools. The objective was to identify appropriate evaluation tools and establish criteria for selecting assessment tools based on their functionality and relevance to the target content.

2. Methods

2.1. Study design

This systematic review was registered in the PROSPERO database (CRD42023486346).

2.2. Literature search strategy

This systematic review comprehensively searched both Chinese and international databases, including the China National Knowledge Infrastructure (CNKI), Wanfang, China Science and Technology Journal (VIP), SinoMed, PubMed, Web of Science, CINAHL, Embase, the Cochrane Library, PsycINFO, and Scopus databases. The search period spanned from the inception of the database to October 2023. Additionally, a supplementary search was conducted to identify related literature. The search strategy included keywords such as "web," "health information," "quality," and "evaluation tools" and was tailored to fit the specific features of each database. For the detailed search strategy, refer to Appendix A.

2.3. Literature inclusion and exclusion criteria

The following inclusion criteria were utilized: 1) the development, validation, and evaluation of measurement properties of online health information quality assessment scales; 2) the presentation of at least one measurement property; and 3) the availability of the full text of the scale. The following exclusion criteria were utilized: 1) using scale results solely as outcome measures; 2) literature that was not in English or Chinese; 3) secondary research, conference proceedings, and dissertations; and 4) literature on scale translation.

2.4. Literature selection and data extraction

The search, selection, and feature extraction of studies were performed independently by two researchers (Y. Li and H. Ouyang). Any discrepancies were resolved by discussion or consultation with a third researcher (G. Lin). All the identified literature was archived in the EndNote, and after duplicate elimination, titles and abstracts were screened against the inclusion-exclusion criteria. The full-text articles were then evaluated to identify studies for final inclusion. In addition, quality assessment tools from online health information quality assessment studies were included if they met the criteria for this study. The following data were extracted from the final included literature: first author, year of publication, region of study, name of the tool, target population, assessment objectives, sample size, time required to complete the scale, number of scale dimensions and items, description of scale dimensions, scoring method for each item, and measurement properties of the scale.

2.5. Quality assessment methods

The assessment tools included in the study were evaluated by two trained researchers (H. Ouyang and G. Lin), each working independently and using the COSMIN guidelines [14] to assess the measurement properties, methodological quality, and level of evidence. The final review of the evaluation outcomes was conducted by three researchers (Y. Li, H. Ouyang, and G. Lin).

2.5.1. Assessment of methodological quality

The COSMIN Risk of Bias tool [15] was employed to evaluate reliability and measurement error, and the COSMIN Risk of Bias checklist [16] was used to assess various aspects of the instruments, including tool development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, criterion validity, hypotheses testing for construct validity, and responsiveness. A five-point scale was used in both checklists to assess the methodological quality of each study: "very good," "adequate," "doubtful," "inadequate," and "not applicable." The worst rating principle determined the aggregate rating for each study.

2.5.2. Assessment of measurement properties

The tool (COSMIN methodology for assessing content validity) [16] was utilized to evaluate content validity, and an additional tool (the Chinese version of the criteria for good measurement properties) [16] was employed to assess structural validity, internal consistency, hypotheses testing for construct validity, criterion validity, reliability, cross-cultural validity/measurement invariance, measurement error, and responsiveness. The evaluation was based on the following ratings: "+" (sufficient), "-" (insufficient), or "?" (indeterminate). However, if a scale's measurement properties exhibit varying evaluation results across studies and these differences cannot be reconciled, the overall rating for the measurement properties is deemed "inconsistent."

2.5.3. Quality of evidence

The COSMIN guidelines adapt the Grading, Recommendations, Assessment, Development, and Evaluation (GRADE) system [17] to evaluate the quality of evidence for measurement properties. The guidelines presuppose that the initial quality of evidence for the measured properties is rated as "high." They then assess for the presence of risk of bias, inconsistency, imprecision, and indirectness in the studies of measurement properties. Depending on these factors, the quality of the evidence may be downgraded, leading to a final quality rating of "high," "moderate," "low," or "very low."

2.5.4. Overall recommendations

By the COSMIN guidelines, scales are categorized into three main groups [14]: 1) Category A scales, which are characterized by "sufficient" content validity (at any level of evidence) and "sufficient" internal consistency (with at least low-quality evidence), are recommended for use owing to their reliability. 2) Category B scales, which do not meet the criteria for Category A or C scales, are considered potentially useable but require further research to evaluate their quality. 3) Category C scales have high-quality evidence indicating "insufficient" measurement properties; thus, their use is not recommended.

3. Results

3.1. Search and study selection

A systematic search identified 36,672 potential studies, which were narrowed down to 2,349 by eliminating duplicate documents and screening the titles and abstracts. Upon reviewing the full texts, 11 studies were found to be eligible. An additional seven studies were discovered through literature tracking, resulting in 18 studies meeting this systematic review's criteria. The literature screening process is visualized in Fig. 1.

3.2. Study characteristics

The current review includes 17 health information quality assessment tools developed between 1999 and 2023 in countries including the United States, the United Kingdom, Australia, Canada, South Korea, and Turkey. These tools have been utilized in existing studies to assess the quality of online health information however, their target populations vary. The Centers for Disease Control and Prevention Clear Communication Index (CCI) [18], The Credible and Usable Evaluation of patient education tool for web-sites (CUE) [19], the Diabetes Quality of Internet Information (Diabetes QII) tool [20], the Ensuring Quality Information for Patients (EQIP) tool [21], and the QUality Evaluation Scoring Tool (QUEST) [22] target health care professionals with medical knowledge, whereas the Health Information Website Evaluation Tool (HIWET) [23], the Online Quality Assessment Tool (OQAT) [24], the Medical Quality Video Evaluation Tool (MQ-VET) [25], the Principles for Health-related Information on Social Media (PRHISM) [26], the Patient Education Materials Assessment Tool (PEMAT) [27], the DISCERN [28], the Health Information Quality Assessment Tool (HIQUAL) [29], the WebMedQual [30], the Tool for Measuring Consumer Perception of Quality Aspects (MCPQA) [31], the Online Patient Satisfaction Index (OPSI) [32], the Web Resource Rating (WRR) tool [33], and the Health Website and Health Information Rating index (version 2) (the Rating index(version2)) [34] are designed for the general public to assess health information. The Diabetes QII [20] had an average completion time of 30.26 min, whereas the HIWET [23] required 8.32 min. However, completion time data for the remaining scales were not provided. The basic characteristics of the included scales are presented in Appendix B.

3.3. Results of the evaluation of the quality of studies

None of the included studies addressed cross-cultural validity/measurement invariance, and responsiveness was also not evaluated. Appendix C presents the methodological quality and quality of measurement properties for the 17 health information quality assessment tools.

3.3.1. Tool development

The QUEST [22] and WRR [33] lacked reporting on their development processes. The MQ-VET [25], HIQUAL [29], and MCPQA [31] were deemed "inadequate" because it was unclear whether their target populations had been consulted on the comprehensibility of the scale during development. The PEMAT [27] and the Rating index (version 2) [34] also received an "inadequate" rating because they omitted cognitive interviews or other pretests. The WebMedQual [30] was also rated "inadequate" because it did not include pretesting for the scale items in its final form. The remaining nine studies [18–21,23,24,26,28,32] were given a "doubtful" rating owing to insufficient clarity regarding the involvement of at least two researchers in the analysis phase of the development process.

3.3.2. Content validity

The QUEST [22] and WRR [33] omitted a report on content validity. The HIQUAL [29], HIWET [23], and EQIP [21] provided data on the relevance and comprehensibility of the instruments used. The OQAT [24] and the Rating index (version 2) [34] reported on the relevance and comprehensiveness of their tools. The MQ-VET [25], MCPQA [31] and PEMAT [27] reported on relevance. The CCI [18], CUE [19], Diabetes QII [20], DISCERN [28], WebMedQual [30], OPSI [32] and PRHISM [26] all exhibited relevance, comprehensiveness, and comprehensibility. The methodological quality of these studies was considered "doubtful" due to insufficient detail in the data analysis methods or insufficient description by the experts involved.

3.3.3. Structural validity

The OQAT [24] used the Rasch model, which is considered an appropriate choice for addressing the research question. Consequently, its methodological quality was assessed as "adequate." However, the Outfit mean squares did not fall within the ideal range of 0.5–1.5, indicating "insufficient" measurement quality. The MQ-VET [25] employed exploratory factor analysis, which typically warranted an "adequate" methodological quality assessment. However, the measurement quality was deemed "insufficient" because the range of cumulative explained variance was only 32.11%–63.14%, which fell short of the recommended 50%–60% threshold. The MCPQA [31] utilized principal component analysis, which resulted in an "adequate" methodological quality assessment. However, the study's omission of values for the Kaiser–Meyer–Olkin measure and Bartlett's test of sphericity led to an "indeterminate" measurement quality rating.

3.3.4. Internal consistency

Cronbach's α coefficients for various dimensions were reported

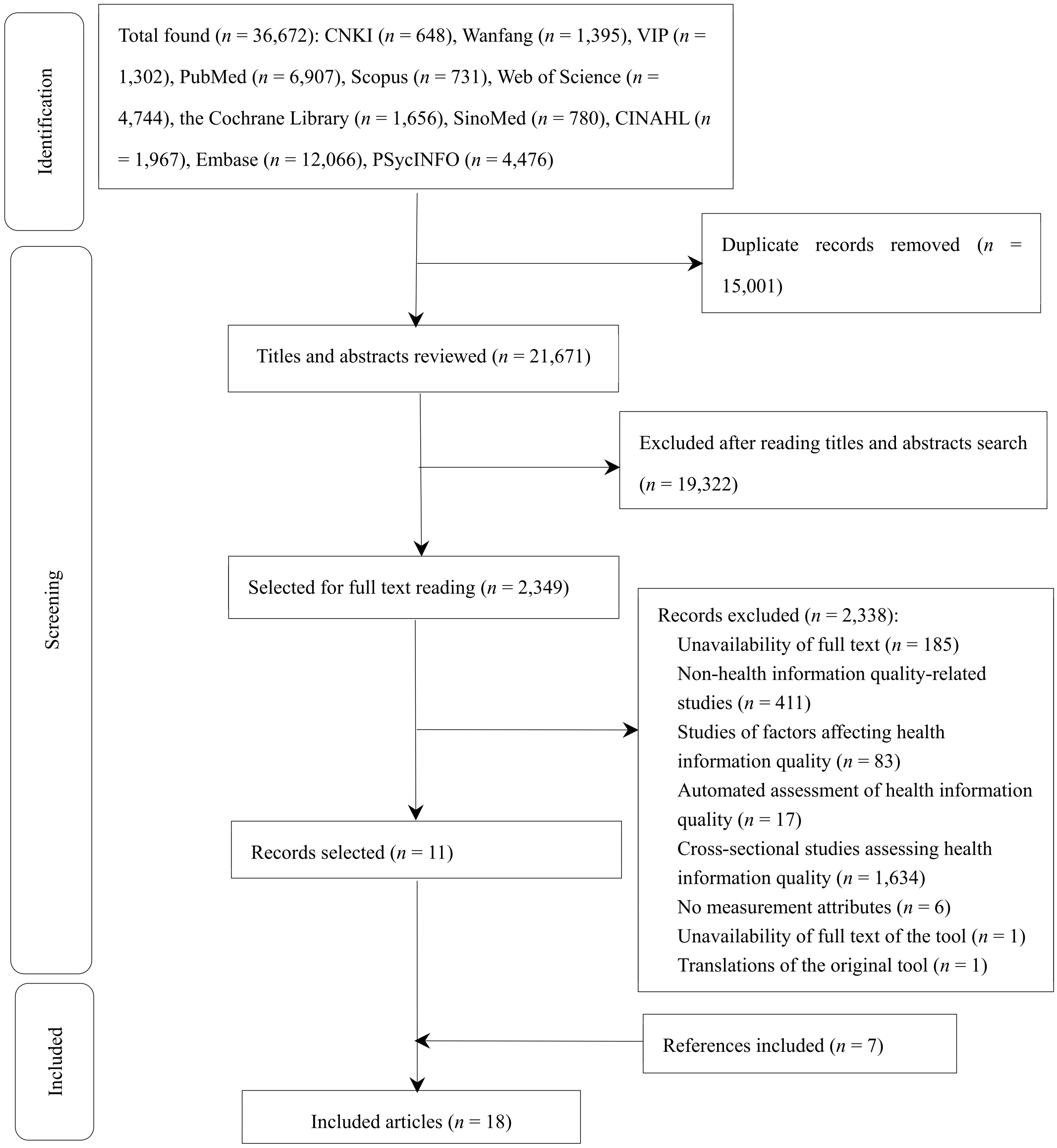

Fig. 1. Study selection flowchart based on PRISMA.

by the OQAT [24], MQ-VET [25], PEMAT [27], and MCPQA [31], which generally indicated "very good" methodological quality. The Cronbach's α coefficients of the OQAT [24] and MCPQA [31] were less than 0.7, and the measurement properties were rated as "insufficient." The Cronbach's α coefficients of the MQ-VET [25] all exceeded 0.7, and the measurement properties were rated as "sufficient." The PEMAT [27] omitted the structural validity report, resulting in an "indeterminate" rating of the measurement properties. The DISCERN [35] and EQIP [21] demonstrated Cronbach's α coefficients above 0.7, indicating acceptable internal consistency. However, owing to the lack of evidence for unidimensionality, their methodological quality was rated as "doubtful." Furthermore, the omission of construct validity reporting in these studies led to an "indeterminate" rating for the measurement properties.

3.3.5. Reliability

The test-retest reliability of the HIWET [23] and OPSI [32] was assessed as "inadequate" methodological quality. This rating stemmed from two factors: first, the time interval between retests deviated from the two weeks recommended by the COSMIN guidelines [36]; second, the scenario description was inadequate during the retest phase. The HIWET [23] and OPSI [32] exhibited ICCs greater than or equal to 0.7; as a result, their measurement properties were considered "sufficient."

Nine studies reported interrater reliability. The Pearson correlation coefficient was calculated for the Diabetes QII [20], but there was no evidence that no systematic difference between measurements had occurred, leading to a "doubtful" methodological quality assessment. The ICCs for the HIWET [23] and WRR [33] were calculated without specifying the model or formula, resulting in an "adequate" methodological quality rating. The DISCERN [35] and HIQUAL [29] clearly described the weighted kappa calculation, earning a "very good" methodological quality rating. However, the OQAT [24], PEMAT [27], EQIP [21], and QUEST [22] lacked specifications of the weighted kappa method, thus receiving an "adequate" methodological quality rating. The interrater reliability values of the Diabetes QII [20], WRR [33], HIQUAL [29], and QUEST [22] were greater than 0.7; the measurement properties were deemed "sufficient." The remaining five tools [21,23,24,27,28] were rated as having "insufficient" measurement properties.

3.3.6. Measurement error

The OPSI [32] reported the smallest detectable change (SDC) value of 4.71. However, the time interval between measurements spanned from 7 days to 4 weeks, which exceeded the 2 weeks recommended by the COSMIN guidelines [36]. As a result, the methodological quality of the scale was assessed as "inadequate." Additionally, the study did not report the value of the minimum important change (MIC) or provide a basis for comparing the relationships between the SDC and MIC. Therefore, the measurement properties were considered "indeterminate."

3.3.7. Criterion validity

Six studies presented correlation coefficients compared with other health information quality assessment scales. Upon review, no significant methodological or statistical flaws were detected; hence, the methodological quality was deemed "very good." However, the HIWET [23], MQ-VET [25], DISCERN [35], EQIP [21], and QUEST [22] demonstrated correlation coefficients below 0.7 with other health information quality assessment tools, resulting in an "insufficient" rating for their measurement attributes. In contrast, the HIQUAL [29] showed a correlation coefficient above 0.7 with other scales, and its measurement attribute was deemed "sufficient."

3.3.8. Hypotheses testing for construct validity

The PEMAT [27] and EQIP [21] were compared with other outcome measures, but their methodological quality was deemed "inadequate," primarily due to the unclear measurement structure of the comparators. Furthermore, the measurement properties of these tools were labeled "indeterminate" because the studies lacked explicit hypotheses, which is a critical component of the research methodology. The OPSI [32] reported convergent and discriminant validity and included defined hypotheses. For convergent validity, the measurement structure of the comparators was insufficiently detailed, leading to an "inadequate" methodological quality rating. Nevertheless, the findings were consistent with the hypotheses, and the measurement attributes were considered "sufficient." In contrast, for discriminant validity, the essential characteristics of the subgroups were not sufficiently defined, resulting in a "doubtful" methodological quality assessment. The results did not correspond with the hypothesized outcomes, rendering the measurement attributes "insufficient."

3.3.9. Evaluation and recommendation of the evidence

The content validity of 15 tools [18–21,23–32,34] was rated as "doubtful," suggesting a risk of bias. This led to a one-level downgrade of the evidence for all the studies. However, for the HIWET [23], OQAT [24], PRHISM [26], PEMAT [27], and the Rating index (version 2) [34], the included study populations were not representative of the target populations, resulting in an additional one-level downgrade of the evidence.

The structural validity ratings for the OQAT [24], MQ-VET [25], and MCPQA [31] were "adequate," which indicates a risk of bias. Consequently, the evidence for these tools was lowered by one level. However, the OQAT [24] exhibited a discrepancy between the study and target populations, necessitating a further one-level reduction in the evidence.

The MQ-VET [25] and MCPQA [31] demonstrated a "high" level of evidence for internal consistency. The evidence was rated as "very low" by the DISCERN [28] and EQIP [21] due to the risk of bias, underrepresentation of the target population, and small sample size. The OQAT's [24] level of evidence for internal consistency was rated as "very low" because of discrepancies between the study and target populations and a sample size of less than 50. The PEMAT's [27] internal consistency evidence level was classified as "low" because of an inadequate description of the study population and a small sample size.

The HIWET [23] and OPSI [32] demonstrated a "very low" level of evidence for reliability, as their retest reliability methodological quality was deemed "inadequate," suggesting a significant risk of bias. The DISCERN's [28] and HIQUAL's [29] reliability evidence was rated "low" due to their small sample sizes, both under 50. The OQAT's [24] and the WRR's [33] reliability were also categorized as "low" because their methodological quality was "adequate" and their study populations were part of their target populations. The QUEST [22], EQIP [21], and PEMAT [27] were assigned a "very low" level of evidence for reliability, influenced by "adequate" methodological quality, insufficient sample sizes, and partial representation of the target population. The reliability of the Diabetes QII [20] was also rated "very low" because of "doubtful" methodological quality and a sample size of less than 50.

The OPSI's [32] measurement error methodological quality was deemed "inadequate," resulting in a "very low" level of evidence. The MQ-VET's [25] criterion validity evidence level was "high." In contrast, the HIWET [23], HIQUAL [29], and QUEST [22] had a "low" level of evidence for criterion validity, primarily due to sample sizes under 50. The criterion validity of the DISCERN [28] and EQIP [21] was rated "low" because they included studies with sample sizes under 100 and study populations that were part of the target populations.

The PEMAT [27] and EQIP [21], along with the OPSI [32] for convergent validity, were assigned a "very low" level of evidence for hypotheses testing for construct validity, attributed to "inadequate" methodological quality. The discriminant validity of the OPSI [32] was classified as "low" because of "doubtful" methodological quality. After the measurement properties and quality of the associated evidence were evaluated, fifteen studies were assigned a recommendation at Level B. Two studies demonstrated "high" evidence of "insufficient" measurement properties and, as a result, were recommended at Level C. Appendix D presents detailed results of the evidence levels and corresponding recommendations.

4. Discussion

4.1. Applicability of these scales is limited, and their use requires careful consideration

The health information materials targeted by the 17 tools included in this study exhibited considerable variation. Specifically, the DISCERN [28] and EQIP [21] are designed to evaluate written health information. The WRR [33], CUE [19], HIWET [23], WebMedQual [30], MCPQA [31], and the Rating index (version 2) [34] are designed for evaluating health information on websites. The CCI [18], PRHISM [26], QUEST [22], and HIQUAL [29] focus on assessing the quality of online text-based health information. The PEMAT [27] is unique because it evaluates printable and audiovisual health information materials. The OPSI [32] is specialized for assessing the quality of information related to low back pain. The Diabetes QII [20] is used to evaluate the quality of diabetes-focused websites, whereas the OQAT [24] assesses online nutritional information. Finally, the MQ-VET [25] is intended for the quality assessment of medical videos. Although they are used to assess the quality of online health information, a comprehensive assessment of their suitability and reliability across different materials and populations remains lacking [37]. It is essential to consider the impact of subject characteristics and research context on the measurements derived from these tools, necessitating a re-evaluation of their reliability and validity. The selection of assessment tools must be guided by their functionality and alignment with the target material, ensuring that health information's true quality and reliability are not obscured.

4.2. Methodological quality of health information assessment tools requires further improvement

Instrument development was often rated as "doubtful" or "inadequate," with "doubtful" indicating a lack of clarity in data coding, analysis, and analyst numbers and "inadequate" indicating the absence of cognitive interviews or pretests, target population feedback on scale comprehensibility, or final evaluations. Tools such as the HIWET [23], OQAT [24], and the Rating index (version 2) [34], which are utilized by the general public, were found to have development processes that lacked public involvement. Integrating cognitive interviews or pilot studies during development and assessing their relevance, comprehensiveness, and comprehensibility is vital for instruments accurately reflecting the target audiences' perspectives and needs [38–40]. Exploratory factor analysis (EFA) was applied to the MQ-VET [25] and the MCPQA [31] to explore the underlying structure of their items. However, the COSMIN guidelines recommend confirmatory factor analysis (CFA) as the preferred method for establishing the construct validity of health information quality assessment scales [41]. CFA's superiority is attributed to its foundation in preexisting theoretical frameworks, which allows for more rigorous testing of the hypothesized relationships among items and constructs. To ensure a theoretically sound and empirically valid scale, future scale development initiatives should begin with CFA to delineate the internal structure, followed by EFA to refine the relationships between items and factors further.

4.3. The measurement properties of health information quality assessment tools are incompletely reported and require further validation

This study found no instances where the full range of measurement properties was addressed. The QUEST [22] and WRR [33] lacked content validity reporting, which is deemed the most crucial measurement property [14,36,41]. Future studies should include this measurement property. Other measurement properties are also important for studies [18,19,26,30,34] that only reported content validity. Additional studies are needed to evaluate these measurement properties and verify their feasibility. Structural validity, a prerequisite to assessing internal consistency [36], was not reported for most of the scales in this study. Future studies on online health information quality assessment scales should aim to increase the reporting of other measurement properties by the COSMIN guidelines, thereby increasing the scientific rigor of these instruments.

4.4. Implications for practice

Among the tools for website-based health information, the HIWET [23] can be tentatively recommended because of its demonstrated reliability, criterion validity, and content validity. The MCPQA [31] was assigned a Category C recommendation, indicating that it is not recommended for use. The CUE [19], WebMedQual [30], and the Rating index (version 2) [34] only reported content validity, which limited the ability to recommend them. The HIQUAL [29] is provisionally recommended for its "sufficient" interrater reliability and criterion validity to assess online text-based health information. The MQ-VET [25], with a Category C classification, is not currently recommended for video-based health information tools. However, the PEMAT [27] can be tentatively recommended for potential application.

These tools facilitate the assessment of health information quality for healthcare professionals, thereby supplying the public with high-quality health information and mitigating the adverse effects of inaccurate health information. Concurrently, healthcare professionals can craft appropriate health information by the criteria within health information quality assessment tools, fostering the spread of health information and enhancing public health literacy.

4.5. Limitations

The exclusion of literature on scale translation in the present study constrains the comprehensive assessment of cross-cultural validity. To enhance assessment, future research should incorporate translated scale versions to evaluate their cross-cultural applicability. Additionally, the absence of dissertations in this study might have led to the omission of key studies on scale development and validation, potentially affecting the evaluation of methodological quality and measurement properties. Therefore, future research requires a more exhaustive literature review. The COSMIN guidelines, which are used to assess the reliability and validity of patient-reported outcome measures, may not be entirely suitable for online health information quality assessment. Consequently, the findings should be interpreted with caution.

5. Conclusions

The results of this study can be divided into two sets of scales: B and C. The category B scales will undergo further validation to confirm their accuracy and reliability. The category C scales necessitate additional research to explore their full potential and application. The Health Information Website Evaluation Tool (HIWET) can be temporarily used to evaluate the quality of health information on websites. The Patient Education Materials Assessment Tool (PEMAT) can temporarily assess the quality of video-based health information. This study's findings establish a foundational basis for the evidence-based selection of suitable assessment tools and provide scientific guidance for their continuous refinement.

CRediT authorship contribution statement

Yating Li: Conceptualization, Methodology, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Validation. **Hui Ouyang:** Methodology, Formal analysis, Data curation, Writing – review & editing. **Gan Lin:** Methodology, Formal analysis, Data curation, Writing – review & editing. **Yichao Peng:** Methodology, Validation, Writing – review & editing. **Jinghui Yao:** Conceptualization, Methodology, Supervision, Project administration, Writing – review & editing. **Yun Chen:** Conceptualization, Methodology, Supervision, Project administration, Funding acquisition, Writing – review & editing.

Data availability statement

The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Funding

This study was supported by President Foundation of the Third Affiliated Hospital of Southern Medical University (YH202207). Funding organizations played no role in the design, implementation and analysis of the survey.

Declaration of competing interest

The authors declare that there are no conflicts of interest regarding the publication of this paper. We confirm that we do not have any financial or personal relationships with other people or organizations that could inappropriately influence (bias) our work presented in this manuscript.

Appendices. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijnss.2025.02.015.

References

[1] French-Lawyer J, Siano S, Ioerger M, Young V, Turk MA. Health information seeking and people with disability: a systematic search and scoping review. Disabil Health J 2021;14:100983. https://doi.org/10.1016/j.dhjo.2020.100983.

[2] Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: an analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017;3(1):1302785. https://doi.org/10.1080/23311886.2017.1302785.

[3] China Internet Network Information Center. The 52nd statistical report on China's internet development. https://www.cnnic.net.cn/n4/2023/0828/c88-10829.html. [Accessed 31 May 2024].

[4] Kantar Health, WebMD. Medical science short video and live streaming insight report. https://www.baogaoting.com/info/203693. [Accessed 31 May 2024].

[5] Dalmer NK. Questioning reliability assessments of health information on social media. J Med Libr Assoc 2017;105:61–8. https://doi.org/10.5195/jmla.2017.108.

[6] Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018;359(6380):1146–51. https://doi.org/10.1126/science.aap9559.

[7] Borges do Nascimento IJ, Pizarro AB, Almeida JM, Azzopardi-Muscat N, Gonçalves MA, Björklund M, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022;100(9):544–61. https://doi.org/10.2471/BLT.21.287654.

[8] Daraz L, Morrow AS, Ponce OJ, Beuschel B, Farah MH, Katabi A, et al. Can patients trust online health information? A meta-narrative systematic review addressing the quality of health information on the Internet. J Gen Intern Med 2019;34(9):1884–91. https://doi.org/10.1007/s11606-019-05109-0.

[9] Zhang Y, Kim Y. Consumers' evaluation of web-based health information quality: meta-analysis. J Med Internet Res 2022;24(4):e36463. https://doi.org/10.2196/36463.

[10] Khazaal Y, Chatton A, Zullino D, Khan R. HON label and DISCERN as content quality indicators of health-related websites. Psychiatr Q 2012;83(1):15–27. https://doi.org/10.1007/s11126-011-9179-x.

[11] Jadad AR, Gagliardi A. Rating health information on the Internet: navigating to knowledge or to babel? JAMA 1998;279(8):611–4. https://doi.org/10.1001/jama.279.8.611.

[12] Gagliardi A, Jadad AR. Examination of instruments used to rate quality of health information on the Internet: chronicle of a voyage with an unclear destination. BMJ 2002;324(7337):569–73. https://doi.org/10.1136/bmj.324.7337.569.

[13] Breckons M, Jones R, Morris J, Richardson J. What do evaluation instruments tell us about the quality of complementary medicine information on the Internet? J Med Internet Res 2008;10(1):e3. https://doi.org/10.2196/jmir.961.

[14] Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27(5):1147–57. https://doi.org/10.1007/s11136-018-1798-3.

[15] Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN risk of bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol 2020;20(1):293. https://doi.org/10.1186/s12874-020-01179-5.

[16] Consensus-based Standards for the selection of health measurement instruments (COSMIN). Checklists for assessing study qualities. https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/. [Accessed 2 June 2024].

[17] Akl E, Mustafa R, Wiercioch NSW, Schünemann H, Brożek J, Guyatt G, et al. GRADE handbook. https://gdt.gradepro.org/app/handbook/handbook.html. [Accessed 11 September 2024].

[18] Baur C, Prue C. The CDC Clear Communication Index is a new evidence-based tool to prepare and review health information. Health Promot Pract 2014;15(5):629–37. https://doi.org/10.1177/1524839914538969.

[19] Klompstra L, Liljeroos M, Lundgren J, Ingadottir B. A clinical tool (CUE-tool) for health care professionals to assess the usability and quality of the content of medical information websites: electronic Delphi study. J Med Internet Res 2021;23(2):e22668. https://doi.org/10.2196/22668.

[20] Seidman JJ, Steinwachs D, Rubin HR. Design and testing of a tool for evaluating the quality of diabetes consumer-information Web sites. J Med Internet Res 2003;5(4):e30. https://doi.org/10.2196/jmir.5.4.e30.

[21] Moult B, Franck LS, Brady H. Ensuring quality information for patients: development and preliminary validation of a new instrument to improve the quality of written health care information. Health Expect 2004;7(2):165–75. https://doi.org/10.1111/j.1369-7625.2004.00273.x.

[22] Robillard JM, Jun JH, Lai JA, Feng TL. The QUEST for quality online health information: validation of a short quantitative tool. BMC Med Inf Decis Making 2018;18(1):87. https://doi.org/10.1186/s12911-018-0668-9.

[23] Zubiena L, Lewin O, Coleman R, Phezulu J, Ogunfiditimi G, Blackburn T, et al. Development and testing of the health information website evaluation tool on neck pain websites - an analysis of reliability, validity, and utility. Patient Educ Counsel 2023;113:107762. https://doi.org/10.1016/j.pec.2023.107762.

[24] Ellis CH, Moore JB, Ho P, Evans CE. Development and validation of a quality assessment tool to assess online nutrition information. Digit Health 2023;9: 20552076231187249. https://doi.org/10.1177/20552076231187249.

[25] Guler MA, Aydın EO. Development and validation of a tool for evaluating YouTube-based medical videos. Ir J Med Sci 2022;191(5):1985–90. https://doi.org/10.1007/s11845-021-02864-0.

[26] Denniss E, Lindberg R, McNaughton SA. Development of principles for health-related information on social media: delphi study. J Med Internet Res 2022;24(9):e37337. https://doi.org/10.2196/37337.

[27] Shoemaker SJ, Wolf MS, Brach C. Development of the patient education materials assessment tool (PEMAT):A new measure of understandability and actionability for print and audiovisual patient information. Patient Educ Counsel 2014;96(3):395–403. https://doi.org/10.1016/j.pec.2014.05.027.

[28] Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999;53(2):105–11. https://doi.org/10.1136/jech.53.2.105.

[29] Lee N, Oh SW, Cho B, Myung SK, Hwang SS, Yoon GH. A health information quality assessment tool for Korean online newspaper articles: development study. J Med Internet Res 2021;23(7):e24436. https://doi.org/10.2196/24436.

[30] Provost M, Koompalum D, Dong DE, Martin BC. The initial development of the WebMedQual scale: domain assessment of the construct of quality of health web sites. Int J Med Inf 2006;75(1):42–57. https://doi.org/10.1016/j.ijmedinf.2005.07.034.

[31] Dubowicz A, Schulz PJ. Medical information on the Internet: a tool for measuring consumer perception of quality aspects. Interact J Med Res 2015;4(1):e8. https://doi.org/10.2196/ijmr.3144.

[32] Afzali T, Lauridsen HH, Thomsen JL, Hartvigsen J, Jensen MB, Riis A. The online patient satisfaction index for patients with low back pain: development, reliability, and validation study. JMIR Form Res 2021;5(11):e21462. https://doi.org/10.2196/21462.

[33] Dobbins M, Watson S, Read K, Graham K, Yousefi Nooraie R, Levinson AJ. A tool that assesses the evidence, transparency, and usability of online health information: development and reliability assessment. JMIR Aging 2018;1(1): e3. https://doi.org/10.2196/aging.9216.

[34] Bomba D. Evaluating the quality of health web sites: developing a validation method and rating instrument. Proceedings of the 38th annual Hawaii international conference on system sciences. January 6-6, 2005,Big island, HI, USA. IEEE; 2005. https://doi.org/10.1109/HICSS.2005.251.

[35] Ademiluyi G, Rees CE, Sheard CE. Evaluating the reliability and validity of three tools to assess the quality of health information on the Internet. Patient Educ Counsel 2003;50:151–5. https://doi.org/10.1016/S0738-3991(02)00124-6.

[36] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. Qual Life Res 2018;27(5):1171–9. https://doi.org/10.1007/s11136-017-1765-4.

[37] Afful-Dadzie E, Afful-Dadzie A, Egala SB. Social media in health communication: a literature review of information quality. Health Inf Manag 2023;52(1): 3–17. https://doi.org/10.1177/1833358321992683.

[38] Lasch KE, Marquis P, Vigneux M, Abetz L, Arnould B, Bayliss M, et al. PRO development: rigorous qualitative research as the crucial foundation. Qual Life Res 2010;19(8):1087–96. https://doi.org/10.1007/s11136-010-9677-6.

[39] Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity: establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1:Eliciting concepts for a new PRO instrument. Value Health 2011;14(8):967–77. https://doi.org/10.1016/j.jval.2011.06.014.

[40] Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS. Mayo/FDA patient-reported outcomes Consensus meeting group. Patient-Reported outcomes: instrument development and selection issues. Value Health 2007;10(Suppl 2):S86–93. https://doi.org/10.1111/j.1524-4733.2007.00271.x.

[41] Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Qual Life Res 2018;27(5):1159–70. https://doi.org/10.1007/s11136-018-1829-0.