Postprint of a Systematic Review of Symptom Assessment Tools for Heart Failure Patients

Authors: Li Jingjing, Zheng Gaigai, Wang Yu, Liu Yancun, Zhang Shuangqi, Yang Qiaofang, Yang Qiaofang

Date: 2024-01-24T00:00:00+00:00

Abstract

Background Both domestically and internationally, numerous symptom assessment tools are available for heart failure patients; however, standardized quality evaluation studies for such tools are relatively scarce, posing challenges for tool selection. Objective To evaluate the measurement properties and methodological quality of studies on symptom assessment tools for heart failure patients, and to provide a reference basis for relevant stakeholders in selecting symptom assessment tools. Methods We searched Chinese and English databases including PubMed, Embase, and CNKI (China National Knowledge Infrastructure) to retrieve studies evaluating the measurement properties of symptom assessment tools for heart failure patients from inception to July 30, 2023. Using the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines, we conducted a systematic evaluation of the included assessment tools and developed recommendations. Results A total of 11 studies were included for quality evaluation, encompassing 8 symptom assessment tools for heart failure patients: the Chinese version of the Memorial Symptom Assessment Scale-Heart Failure (MSAS-HF), Heart Failure-Somatic Perception Scale (HF-SPS), M.D. Anderson Symptom Inventory-Heart Failure (MDASI-HF), Heart Failure-Symptom Status Questionnaire (SSQ-HF), Shortness of Breath-Heart Failure scale (SOB-HF), Minnesota Living with Heart Failure Questionnaire (MLHFQ), Chronic Heart Failure Assessment Tool (CHAT), and Kansas City Cardiomyopathy Questionnaire (KCCQ). Regarding the measurement properties of the scales, the content validity of the Chinese versions of MSAS-HF, MDASI-HF, MLHFQ, and KCCQ was “sufficient,” that of HF-SPS, SSQ-HF, and SOB-HF was “uncertain,” and that of CHAT was “insufficient.” Finally, the Chinese versions of MSAS-HF and MLHFQ received an A-level recommendation, while the remaining six scales received a B-level recommendation. Conclusion The Chinese versions of MSAS-HF and MLHFQ have higher levels of evidence. Considering the specificity of assessment tools, the Chinese

version of MSAS-HF can be recommended for symptom assessment in heart failure patients.

Full Text

A Systematic Review of Symptom Assessment Tools for Patients with Heart Failure

LI Jingjing[1], ZHENG Gaigai[2], WANG Yu[3], LIU Yancun[4], ZHANG Shuangqi[1], YANG Qiaofang[3]*

[1]School of Nursing, Henan University of Chinese Medicine, Zhengzhou 450046, China
[2]Heart Failure Ward, Fuwai Central China Cardiovascular Hospital, Zhengzhou 450000, China
[3]Heart Center of Henan Provincial People's Hospital/Nursing Department, Fuwai Central China Cardiovascular Hospital, Zhengzhou 450000, China
[4]Structural Heart Disease Area, Fuwai Central China Cardiovascular Hospital, Zhengzhou 450000, China
*Corresponding author: YANG Qiaofang, Chief nurse; E-mail: eaam68@163.com

Abstract

Background: Numerous tools are available for symptom assessment in heart failure patients, but the lack of standardized quality evaluation studies for these tools creates difficulties in selection. **Objective:** To evaluate the measurement properties and methodological quality of symptom assessment tools for heart failure patients and provide evidence-based guidance for tool selection. **Methods:** We searched Chinese and English databases including PubMed, Embase, and CNKI for studies evaluating the measurement properties of heart failure symptom assessment tools from inception to July 30, 2023. The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guideline was employed to conduct a systematic review and formulate recommendations. **Results:** Eleven studies evaluating eight heart failure symptom assessment tools were included: Chinese version of Memorial Symptom Assessment Scale Heart Failure (MSAS-HF), Heart Failure Somatic Perception Scale (HF-SPS), M.D. Anderson Symptom Inventory Heart Failure (MDASI-HF), Symptom Status Questionnaire Heart Failure (SSQ-HF), Shortness of Breath in Heart Failure Instrument (SOB-HF), Minnesota Living with Heart Failure Questionnaire (MLHFQ), Chronic Heart Failure Assessment Tool (CHAT), and Kansas City Cardiomyopathy Questionnaire (KCCQ). Regarding measurement properties, the Chinese version of MSAS-HF, MDASI-HF, MLHFQ, and KCCQ demonstrated "sufficient" content validity, while HF-SPS, SSQ-HF, and SOB-HF showed "uncertain" content validity, and CHAT exhibited "insufficient" content validity. Ultimately, the Chinese version of MSAS-HF and MLHFQ received Grade A recommendations, while the remaining six scales received Grade B recommendations. **Conclusion:** The Chinese version of MSAS-HF

and MLHFQ have high-quality evidence. Considering tool specificity, the Chinese version of MSAS-HF is recommended for symptom assessment in heart failure patients.

Keywords: Heart failure; Symptom assessment; Measurement property; COSMIN guideline; Systematic reviews

Introduction

Heart failure represents the advanced stage of various cardiac diseases and has become a critical global public health challenge. Heart failure patients experience complex symptom burdens, and poor symptom control often leads to decreased disease coping capacity and increased hospital readmissions. Scientific and accurate measurement of symptoms through assessment scales is essential for preventing and reducing readmission rates. Currently, diverse heart failure symptom assessment tools exist with varying evaluation methods and content. However, systematic evaluations of these scales' measurement properties remain insufficient, making it difficult to determine whether these tools possess adequate psychometric qualities for clinical application. Measurement properties—including reliability, validity, and responsiveness—are critical indicators of scale quality, and robust measurement properties are prerequisites for clinical application.

The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guideline enables researchers to identify the most appropriate assessment tools by evaluating both methodological quality and measurement properties. This study systematically evaluated heart failure symptom assessment tools according to the COSMIN guideline to provide evidence-based guidance for tool selection and clinical application in China.

Methods

1.1 Literature Search Strategy

We systematically searched PubMed, Embase, Web of Science, CINAHL, CNKI, VIP, Wanfang Data, and the Chinese Biomedical Literature Database from inception to July 30, 2023, to identify studies evaluating the methodological and measurement properties of heart failure symptom assessment tools. We employed a combination of subject headings and free-text terms. Chinese search terms included: heart failure, cardiac insufficiency, symptom assessment, questionnaire, scale, reliability, and validity. English search terms included: heart failure, cardiac insufficiency, heart decompensation, symptom assessment, questionnaire, scale, tool, instrument*, reliability, and validation. The PubMed search strategy is illustrated in [Figure 1: see original paper].

1.2 Inclusion and Exclusion Criteria

Inclusion criteria: (1) Studies involving heart failure patients; (2) Studies focusing on symptom assessment tools; (3) Studies on tool development or validation; and (4) Studies evaluating measurement properties. **Exclusion criteria:** (1) Non-Chinese or non-English publications; (2) Reviews, commentaries, or editorials; (3) Duplicate publications; and (4) Studies using tools solely as outcome measures without psychometric evaluation.

1.3 Literature Screening and Data Extraction

Two researchers independently screened literature and extracted data, with discrepancies resolved by a third reviewer. Extracted information included first author, scale name, study population, country, sample size, dimensions, and measurement properties.

1.4 Quality Evaluation

We conducted the systematic review according to the COSMIN guideline.

1.4.1 Methodological Quality Assessment The COSMIN Risk of Bias checklist was used to evaluate methodological quality across three domains with ten items, assessing risk of bias in content validity, structural validity, and other measurement properties. Each item was rated as "very good," "good," "doubtful," "poor," or "inadequate," with the overall domain rating determined by the lowest score (worst-case principle).

1.4.2 Measurement Property Evaluation Measurement properties were evaluated using COSMIN quality criteria. Content validity was assessed through comprehensiveness, relevance, and comprehensibility. Final ratings were categorized as "sufficient (+)," "insufficient (-)," or "uncertain (?)."

1.4.3 Evidence Synthesis and Grading The modified GRADE system was used to grade evidence quality. Evidence for each measurement property began at "high" quality and could be downgraded based on risk of bias, indirectness, inconsistency, and imprecision, resulting in four levels: high, moderate, low, and very low.

Recommendation grades: Grade A (recommended) required sufficient content validity and internal consistency, plus high-quality evidence. Grade C (not recommended) applied when high-quality evidence showed insufficient content validity. All other cases received Grade B (recommended with reservations, requiring further research).

Results

2.1 Literature Search Results

The initial search yielded 1,247 records (611 Chinese and 636 English). After removing duplicates and screening titles, abstracts, and full texts, 11 studies were included for quality evaluation. The literature screening process is shown in [Figure 2: see original paper].

2.2 Characteristics and Quality Evaluation of Included Studies

The 11 included studies evaluated eight heart failure symptom assessment tools: Chinese version of Memorial Symptom Assessment Scale Heart Failure (MSAS-HF), Heart Failure Somatic Perception Scale (HF-SPS), M.D. Anderson Symptom Inventory Heart Failure (MDASI-HF), Symptom Status Questionnaire Heart Failure (SSQ-HF), Shortness of Breath in Heart Failure Instrument (SOB-HF), Minnesota Living with Heart Failure Questionnaire (MLHFQ), Chronic Heart Failure Assessment Tool (CHAT), and Kansas City Cardiomyopathy Questionnaire (KCCQ). Basic characteristics are presented in . All studies evaluated content validity and criterion validity, as detailed in .

2.2.1 Content Validity Five studies evaluated comprehensibility, relevance, and comprehensiveness while employing expert consultation methods. However, none conducted cognitive interviews, and descriptions of research processes and statistical methods were inadequate.

2.2.2 Structural Validity Three studies used confirmatory factor analysis (CFA) and received "very good" methodological quality ratings. Six studies used exploratory factor analysis (EFA), receiving "good" ratings. KCCQ did not specify the factor analysis method ("doubtful" rating), while KCCQ-12 did not use factor analysis ("poor" rating). In two studies, CFA showed good fit indices, yielding "sufficient" measurement property ratings. KCCQ demonstrated poor fit ("insufficient"), while other studies were rated "uncertain."

2.2.3 Internal Consistency Seven studies calculated internal consistency for each subscale ("very good" methodological quality). One study calculated only overall internal consistency ("poor"), and another failed to clearly describe whether the scale was unidimensional or multidimensional ("poor"). Two studies reported Cronbach's $\alpha < 0.7$ ("insufficient"), while seven demonstrated adequate internal consistency ("sufficient").

2.2.4 Stability Only three studies reported test-retest reliability. Two provided clear descriptions of measurement intervals and patient stability ("very good" methodological quality), with intraclass correlation coefficients (ICC) $\geq$ 0.76, yielding "sufficient" ratings. One study lacked clarity on patient stability and measurement intervals ("uncertain"). Another study had inappropriate

intervals and no correlation coefficients ("poor"), as COSMIN recommends 2-4 week intervals. Seven studies did not report stability.

2.2.5 Hypothesis Testing for Structural Validity Six studies did not conduct hypothesis testing. Five studies performed testing, with two providing incomplete subgroup descriptions ("good" quality) and three rated "very good." All five studies met expected results ("sufficient").

2.2.6 Criterion Validity All 11 studies described criterion validity. One study lacked clear description of indicators ("poor" methodological quality), while others were "very good." Five studies met all "sufficient" criteria, while six had partial non-compliance ("uncertain").

2.3 Evidence Quality Grading and Recommendations

Regarding risk of bias, HF-SPS, SOB-HF, CHAT, and KCCQ had questionable quality across multiple studies. Inconsistency was observed for Chinese MSAS-HF, MDASI-HF, SSQ-HF, MLHFQ, and KCCQ, resulting in downgraded evidence. Indirectness was present for MDASI-HF (heart failure patients with cancer) and SOB-HF (target population inconsistent with surveyed population), also leading to downgrades.

According to COSMIN criteria, Chinese MSAS-HF and MLHFQ demonstrated "sufficient" content validity and internal consistency, earning Grade A recommendations. No scales had high-quality evidence of "insufficient" content validity, so no Grade C recommendations were assigned. The remaining six scales received Grade B recommendations, as detailed in .

Discussion

3.1 Limitations of Included Studies

3.1.1 Incomplete Reporting of Content Validity Content validity is the most important measurement property, yet included studies reported it inadequately. Few studies consulted experts and patients regarding comprehensiveness, relevance, and comprehensibility. Additionally, quantitative studies lacked clear reporting of research processes and statistical methods. Future tool development or adaptation should follow COSMIN guidelines, incorporate cognitive interviews, and solicit patient and expert input to improve methodological quality.

3.1.2 Unreasonable Internal Structure COSMIN considers CFA superior to EFA, yet included studies provided unclear descriptions of structural validity methods. Future research could combine classical test theory with item response

theory to evaluate additional measurement properties such as unidimensionality, item discrimination, and item difficulty. Cross-cultural adaptation requires rigorous evaluation, which was lacking in included studies. This may reflect difficulties in assessing cultural adequacy and lack of consensus on cultural questionnaire adaptation. Future research should explore cross-cultural item differences and enhance sample comparability.

3.1.3 Potential Bias in Stability Evaluation Stability evaluation was inconsistent across included studies. Due to the dynamic nature of heart failure symptoms, determining measurement intervals relies on researcher judgment, potentially introducing bias. Researchers should explicitly justify their choice of retest intervals and provide rationale for variable stability during the interval to enhance transparency and reduce bias.

3.1.4 Incomplete Measurement Property Reporting No included studies evaluated measurement error or criterion validity comprehensively. Except for KCCQ, which described interpretability and responsiveness during development, other studies omitted these evaluations. Scale development should reference all measurement properties in COSMIN guidelines: for continuous scores, calculate standard error of measurement, minimal detectable change, or limits of agreement; for categorical or ordinal scores, calculate percentage agreement; for criterion validity, compute correlations, area under ROC curves, or sensitivity/specificity.

3.2 Insufficient Comprehensiveness of Assessment Tools

Current heart failure symptom assessment tools focus primarily on physical symptoms and psychological burden, lacking evaluation of social function or rehabilitation outcomes, thus failing to capture complete symptom experiences. Future tool development should follow established principles and processes to explore symptoms' impact on daily activities, social functioning, and rehabilitation. Comprehensive assessment will better understand heart failure's life impact and inform personalized treatment and rehabilitation plans. Assessment methods should consider self-report scales, clinician-rated scales, or combined approaches, supplemented by patient diaries and objective measures.

3.3 Tool Selection and Recommendations

Chinese MSAS-HF and MLHFQ have moderate-quality evidence supporting content validity with comprehensive methodological and measurement property evaluations, earning Grade A recommendations. However, MLHFQ primarily assesses quality of life and daily life impact rather than direct symptom assessment. Therefore, considering specificity, Chinese MSAS-HF is recommended for heart failure symptom assessment. The Chinese MSAS-HF requires 3-20 minutes to complete. Researchers should understand its features, dimensions, and psychometric properties to optimize administration efficiency.

This systematic review reveals considerable heterogeneity in methodological and measurement property reporting across studies. Chinese MSAS-HF and MLHFQ received Grade A recommendations, while six scales received Grade B. Chinese MSAS-HF demonstrates potential for heart failure symptom assessment. Some scales require additional psychometric evaluation. Clinicians should carefully consider tool applicability and limitations when making selection decisions. This review provides evidence-based guidance for choosing scientifically sound assessment tools, facilitating early detection and intervention of symptom changes, enabling more precise treatment and nursing care, and improving patient outcomes. Comparative analysis also highlights methodological issues for researchers to avoid, encouraging COSMIN-guided development of high-quality tools to advance heart failure symptom assessment.

Author Contributions: LI Jingjing conceived and designed the study, drafted and revised the manuscript; LI Jingjing and ZHANG Shuangqi conducted feasibility analysis and literature/data collection; ZHENG Gaigai, WANG Yu, and LIU Yancun performed quality control and manuscript review; YANG Qiaofang provided overall supervision.

Conflict of Interest: The authors declare no conflicts of interest.

ORCID: LI Jingjing: https://orcid.org/0009-0001-8829-2057

Note: Figure translations are in progress. See original paper for figures.