Journal of Pharmaceutical and Sciences
Electronic ISSN: 2656-3088 DOI: https://doi.org/10.36490/journal-jps.com
Homepage: https://journal-jps.com

REVIEW ARTICLE JPS. 2026, 9(3), 2473-2486

Psychometric Properties of Quality of Life Instruments for Patients with Hypertension: A Systematic Review

Karakteristik Psikometrik Alat Ukur Kualitas Hidup bagi Pasien Hipertensi: Tinjauan Sistematis

Afif Rakha Murtadha [a]*, Tri Murti Andayani [b], Dwi Endarti [c]

[a] *Master of Pharmacy Management Study Program, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta, Indonesia.*
[b] *Department of Pharmacology and Clinical Pharmacy, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta, Indonesia.*
[c] *Department of Pharmaceutics, Faculty of Pharmacy, Universitas Gadjah Mada, Yogyakarta, Indonesia.*

Corresponding Authors: afif.rakha.murtadha@mail.ugm.ac.id

Abstract

Background: Hypertension is a chronic condition significantly impacting physical and psychological well-being. HRQoL instruments must undergo cross-cultural adaptation and psychometric evaluation before clinical application. **Objective:** This systematic review evaluated the psychometric robustness (reliability, validity, responsiveness) of hypertension-specific HRQoL instruments. **Methods:** Eight databases were searched for original research published in English or Indonesian (2021–2026) following PRISMA 2020. Methodological quality and evidence certainty were assessed using COSMIN and modified GRADE. **Results:** 13 studies from 10 countries were included. Six instruments (46.1%) demonstrated strong psychometric profiles, including adequate content validity and robust reliability. While content validity (100%) and internal consistency (84.6%) were strong due to rigorous adaptation, divergent validity and responsiveness were severely underreported (84.6%). According to COSMIN guidelines, test-retest reliability studies with sample sizes below 50 participants are considered to have 'Low' GRADE evidence quality due to insufficient precision, despite demonstrating adequate ICC values. QLICD-HY, STI-HRQoL, and DASI provided the strongest evidence. **Conclusion:** Instruments are generally reliable, yet future research must evaluate clinical responsiveness through longitudinal designs. Instruments with the strongest psychometric profiles, such as QLICD-HY and DASI, are recommended for accurate patient assessment.

Keywords*: *Quality of Life, Hypertension, Psychometrics

Abstrak

Latar Belakang: Hipertensi adalah kondisi kronis yang berdampak signifikan pada kesejahteraan fisik dan psikologis. Instrumen HRQoL harus melalui adaptasi lintas budaya dan evaluasi psikometrik sebelum aplikasi klinis. **Tujuan:** Tinjauan sistematis ini mengevaluasi ketangguhan psikometrik (reliabilitas, validitas, responsivitas) dari instrumen HRQoL khusus hipertensi. **Metode:** Delapan basis data ditelusuri untuk penelitian orisinal berbahasa Inggris atau Indonesia (2021–2026) mengikuti standar PRISMA 2020. Kualitas metodologis dan kepastian bukti dinilai menggunakan COSMIN dan pendekatan GRADE yang dimodifikasi. **Hasil:** Tiga belas studi dari 10 negara disertakan. Enam instrumen (46,1%) menunjukkan profil psikometrik yang kuat, termasuk validitas isi yang memadai dan reliabilitas yang baik. Meskipun validitas isi (100%) dan konsistensi internal (84,6%) kuat karena adaptasi yang ketat, validitas divergen dan responsivitas sangat jarang dilaporkan (84,6%). Berdasarkan pedoman COSMIN, studi reliabilitas uji ulang dengan ukuran sampel di bawah 50 peserta dianggap memiliki kualitas bukti GRADE "Rendah" karena presisi yang tidak memadai, meskipun menunjukkan nilai ICC yang adekuat. QLICD-HY, STI-HRQoL, dan DASI memberikan bukti terkuat. **Kesimpulan:** Secara umum, instrumen menunjukkan reliabilitas yang baik, tetapi penelitian selanjutnya perlu mengevaluasi responsivitas klinis melalui desain longitudinal. Instrumen dengan profil psikometrik terkuat, seperti QLICD-HY dan DASI, direkomendasikan untuk mendukung penilaian pasien secara akurat.

Kata Kunci: Kualitas hidup, Hipertensi, Psikometrik.

Article History:
Received: 17/05/2026,
Revised: 08/08/2026
Accepted: 08/08/2026,
Available Online: 09/08/2026.

QR access this Article

https://doi.org/10.36490/journal-jps.com.v9i3.1724

Introduction

Hypertension is a serious non-communicable disease that adversely affects patients' quality of life (QoL) through lifelong management burden and associated complications [1,2]. QoL, defined as the overall perception of well-being across physical, psychological, social, and environmental domains [3], is increasingly measured using standardized generic or disease-specific instruments to support clinical decision-making and evaluate treatment efficacy [4,5]. Generic instruments, such as the EQ-5D and SF-36, are designed to compare HRQoL across different conditions and populations, but may lack sensitivity to condition-specific concerns. Disease-specific instruments, such as MINICHAL and emPHasis-10, are more sensitive to the unique impacts of hypertension but limit cross-condition comparability. Generic instruments evaluate broad dimensions of quality of life across different diseases [6]. Meanwhile, disease-specific instruments may better address the characteristics of particular patient groups, although their use can limit comparability between populations with similar conditions [7].

Accuracy in clinical application requires HRQoL instruments to demonstrate robust psychometric properties, specifically validity, reliability, and responsiveness, as defined by the COSMIN framework. [8]. The validity of an instrument indicates how well it reflects the underlying construct of interest [9]. Furthermore, responsiveness is a critical psychometric property, especially for evaluating the effectiveness of clinical interventions. According to the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) framework, responsiveness is defined as the ability of an instrument to detect clinically meaningful changes over time. This concept focuses on the detection of a minimal clinically important change that reflects a true improvement or deterioration in the patient's health status, rather than a mere statistical difference. Without established responsiveness, the clinical utility of a Patient-Reported Outcome Measure (PROM) for longitudinal monitoring remains limited [8].

Despite their widespread use, some generic HRQoL instruments still indicate insufficient evidence for key psychometric properties such as reliability [10]. Cross-cultural adaptation ensures linguistic and conceptual equivalence of an instrument for a new population, but does not guarantee its psychometric validity. Even with rigorous adaptation, each instrument must undergo formal psychometric evaluation to confirm its reliability, validity, and responsiveness in the target population. This systematic review systematically evaluates the psychometric properties of quality-of-life assessments for hypertensive patients, using the COSMIN methodology and the modified GRADE approach to address current evidence gaps. Therefore, this systematic review was conducted to identify, summarize, and compare the psychometric properties of QoL instruments for hypertension in many countries. The results of this study may provide evidence-based guidance for both researchers and healthcare professionals for assessing quality of life in future studies and clinical practice.

Method

Protocol

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 Statement guided this systematic review to ensure transparent, thorough, and reproducible reporting of the review process. The review protocol was established a priori. This pre-specified approach facilitated the systematic identification and appraisal of evidence regarding the psychometric properties of hypertension-related QoL instruments. Furthermore, it provided a structured framework for the synthesis of the findings. By defining

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

these steps in advance, we aimed to enhance methodological rigor, minimize potential bias, and ensure the dependability of the review's conclusions.

Quality Assessment

Quality assessment was performed using the COSMIN checklist framework. This framework was utilized to evaluate the methodological quality of the included studies. Additionally, it provided a standardized criteria for assessing the measurement properties of the various QoL instruments evaluated in this review. Following COSMIN criteria, internal consistency was rated as 'sufficient' (+) when Cronbach's alpha ≥ 0.70 AND the study pre-specified a hypothesis for acceptable reliability. Studies that reported alpha values without a priori hypotheses were rated as indeterminate (?). The assessment covered multiple psychometric domains, including PROM development, validity (content, structural, criterion, and construct), reliability, and responsiveness. In alignment with the COSMIN framework, construct validity was assessed through hypothesis testing, with a clear distinction between divergent validity and discriminative (known-groups) validity. Divergent validity evaluates the degree to which an instrument measures a construct that is theoretically distinct from unrelated constructs, while discriminative or known-groups validity assesses the instrument's capacity to differentiate between clinically distinct subgroups, such as those with varying levels of disease severity or demographic characteristics.

Internal consistency was one of the most frequently validated properties and generally demonstrated very good (V) results across the majority of instruments, including the English and Gujarati versions of MINICHAL, the Serbian SF-36, QLICD-HY, and STI-HRQoL. While high Cronbach's alpha indicates that items within a scale measure the same underlying construct, it does not provide evidence that the construct itself is validly measured. Validity requires separate evaluation of content validity, structural validity, and construct validity (including convergent and divergent validity). Most instruments demonstrated high methodological quality in terms of content validity, with QLICD-HY and STI-HRQoL receiving very good ratings due to rigorous development processes involving in-depth patient interviews. Findings related to reliability showed variation. QLICD-HY and DASI demonstrated very good reliability; several versions of emPHasis-10 (Thai) and MINICHAL (Gujarati and English) were rated as doubtful in this specific domain, primarily because their sample sizes were below the COSMIN standard of 50 participants.

Most studies reported satisfactory validity, particularly for known-groups validity. Large-scale studies evaluating the EQ-5D-5L in China and Vietnam demonstrated very good results in distinguishing quality of life across hypertensive subgroups.. Responsiveness, defined as the ability to detect clinically meaningful change over time according to COSMIN guidelines, was the least frequently reported psychometric property, with only two of 13 instruments (15.4%) providing evidence in this domain, with only QLICD-HY and the English version of MINICHAL providing very adequate evidence of the instruments' ability to detect clinical changes following intensive hypertension treatment.

Eligibility Criteria

The qualifying criteria for this systematic review were developed based on the Population, Intervention, Comparison, Outcome (PICO) framework. The review included individuals of all ages with hypertension, both essential hypertension and pulmonary arterial hypertension, from every geographic context. The intervention of interest was the use of disease-specific and general Quality of Life (QoL) questionnaires developed or validated for patients with hypertension. Where applicable, comparisons were undertaken between original and cross-culturally adapted versions, other versions of the same instrument, established generic QoL measures, and known-groups or test-retest analyses. The main results were the COSMIN assessed psychometric qualities of the QoL measures, including reliability, validity, responsiveness, measurement error and cross-cultural adaption.

Studies were eligible for inclusion if they were original primary research documenting the invention, modification, cross-cultural adaption or psychometric validation of QoL measures for patients with hypertension. Studies were eligible if they reported at least one psychometric feature according to the COSMIN framework: internal consistency, test–retest reliability, structural validity, convergent validity, divergent validity, discriminative validity, or responsiveness. The search was limited to items published in the English or Indonesian languages from 2021 to 2026. Studies were excluded if they failed to report psychometric evaluation of QoL instruments when used as clinical outcome measures, used QoL instruments for constructs other than QoL such as self-care behavior, health beliefs, medication adherence, included ineligible populations such as healthy individuals or patients without hypertension, were not original

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

publications, or did not provide enough methodological information or full-text availability to permit quality appraisal.

Search Strategy

A systematic literature search was performed across eight major databases, including Scopus (656 records), Springer (243 records), MEDLINE (200 records), PubMed (136 records), Google Scholar (119 records), Science Direct (117 records), Cochrane Library (37 records), and EBSCOhost (32 records), resulting in a total of 1,540 records identified. The search strategy applied across all databases employed a combination of specific keywords, Medical Subject Headings (MeSH), and Boolean operators there are ("hypertension" OR "hypertensive patients" OR "high blood pressure") AND ("quality of life" OR "health-related quality of life" OR HRQoL) AND ("instrument" OR questionnaire OR scale OR "measurement tool") AND ("psychometric properties" OR psychometric OR validity OR reliability OR responsiveness). To ensure the identification of the most current and relevant psychometric evaluation studies, the search was restricted to articles published in English or Indonesian between 2021 and 2026. Following the initial identification, 34 duplicate records were removed, and 101 records were marked as ineligible by automation tools, leaving 997 records for title and abstract screening. This rigorous search process aimed to capture all primary research studies documenting the validation and measurement properties of QoL instruments specifically for hypertensive clinical populations worldwide. The PRISMA flow diagram is presented in *Figure 1.*

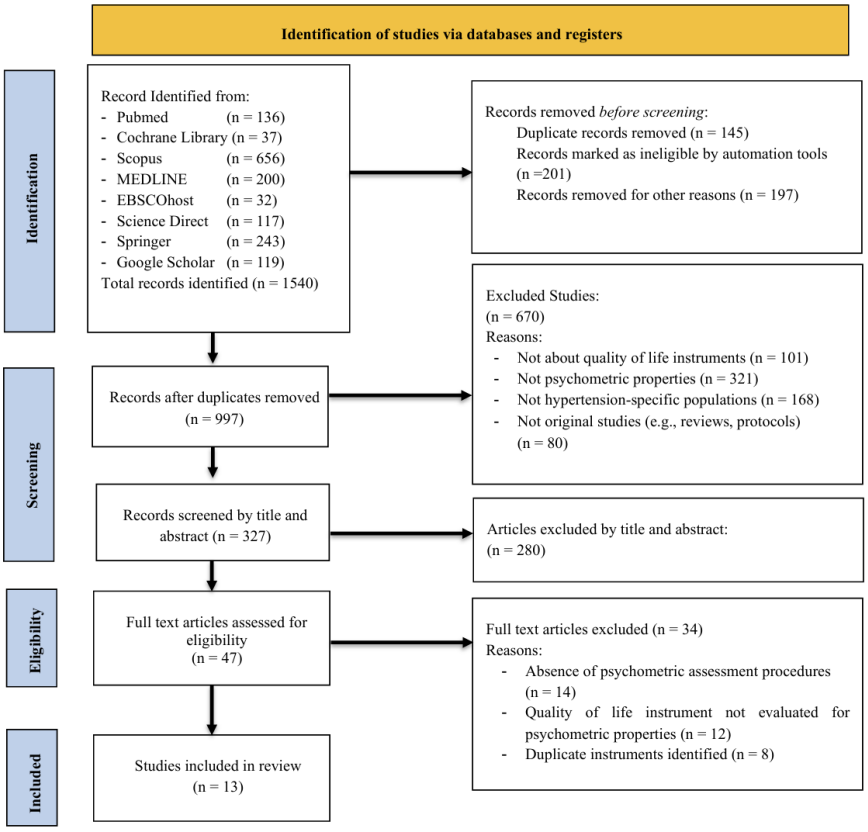

Figure 1. PRISMA diagram of retrieved and included

Data Retrieval

An initial total of 1540 records were found and imported into a reference management program. During deduplication, 145 duplicate records were detected and removed. In addition, 201 records were discarded as ineligible by automated tools, and 197 data points were excluded for various reasons, leaving 997 unique records for screening. Following title and abstract screening, 327 records were reviewed, and 280 papers were removed for not fulfilling the topic emphasis of this study. Subsequently, 47 full-text papers were retrieved and evaluated for eligibility. Studies were included if they met the pre-specified eligibility criteria, namely, patients diagnosed with hypertension (including essential and pulmonary arterial hypertension) and a formal psychometric evaluation of Quality of Life (QoL) questionnaires. Out of these, 34 full-text papers were

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

removed due to: (1) absence of psychometric assessment techniques (n=14); (2) QoL instruments not validated for psychometric qualities (n=12); (3) duplicate instruments (n=eight). A total of 13 studies were included in the systematic review.

Data Synthesis

This study utilizes a narrative synthesis method to integrate and summarize the findings of the included studies. This approach was chosen due to the identified heterogeneity in the types of Quality of Life (QoL) instruments evaluated ranging from generic tools like SF-36 and EQ-5D-5L to disease-specific measures such as MINICHAL, emPHasis-10, QLICD-HY, and STI-HRQoL and the diverse clinical sub-populations, including those with essential hypertension, pulmonary arterial hypertension (PAH), and connective tissue disease-associated PAH (CTD-PAH) across various global regions. The synthesis was performed by categorizing findings based on instrument characteristics, target populations, and specific measurement properties.

Quantitative data, including reliability coefficients (Cronbach's alpha and Intraclass Correlation Coefficients), validity indices (factor loadings, correlation coefficients, and KMO values), and responsiveness measures (Standardized Response Means or Cohen's d effect sizes), were summarized to enable a direct comparison of the psychometric robustness of the instruments across different studies. Following the COSMIN guidelines, the quality of evidence for each measurement property was assessed by evaluating the methodological risk of bias, accounting for factors and the rigor of the statistical methods used. Qualitative findings regarding the cross-cultural and linguistic adaptation processes encompassing forward-backward translation, expert committee reconciliation, and cognitive debriefing were analyzed descriptively to provide insights into the linguistic and cultural suitability of the instruments for the various global populations represented in the review.

Results and Discussion

Included Studies

This systematic review includes 13 studies, predominantly cross-sectional validation studies, alongside some longitudinal evaluations of Quality of Life (QoL) instruments. These studies were conducted in 10 countries across Asia, Europe, and Africa, including China, Thailand, Vietnam, Serbia, India, the United Kingdom, Turkey, Poland, Indonesia, and South Africa. The analyzed sample sizes showed substantial variation, ranging from small pilot samples of 20 patients with pulmonary arterial hypertension (PAH) in Thailand and 30 treatment-naïve hypertensive patients in the United Kingdom to larger cohorts involving 1,296 participants in Vietnam and a massive scale population of 11,412 patients in rural China [11–14]. While the majority of research focused on the general adult hypertensive population, specific studies targeted elderly patients in Indonesia, patients with connective tissue disease-associated PAH (CTD-PAH) in China, and specialized functional evaluations in Turkey [15–17].

The clinical focus of the 13 included studies primarily centered on individuals diagnosed with various forms of hypertension, including essential or arterial hypertension, pulmonary arterial hypertension (PAH), and specialized cases such as connective tissue disease-associated PAH (CTD-PAH) ([12,16,18–22]. Common inclusion criteria across these worldwide studies necessitated a confirmed medical diagnosis often confirmed via right heart catheterization for PAH populations and, for adult-focused research, a minimum age of 18 years [12,14,21,23,24]. Several studies specifically targeted elderly hypertensive populations or those residing in rural settings [13,17]. Exclusion criteria frequently filtered out individuals with cognitive impairments, severe mental illnesses, or unconsciousness, as these conditions could hinder the accurate completion of self-reported or interview-based instruments [13,15,20].

A diverse array of Health-Related Quality of Life (HRQoL) instruments was evaluated, comprising both generic and disease-specific tools. Generic instruments included the EQ-5D (3L and 5L versions), the SF-36 (including the Serbian and Chinese versions), and a modified version of the WHOQOL-BREF [11,13,16,17,19,20]. Disease-specific measures were extensively validated, most notably the emPHasis-10 for pulmonary hypertension across multiple languages including Thai, Polish, and Chinese [12,16,21]. Other validated specific tools included MINICHAL (Gujarati and English versions), the QLICD-HY (V2.0), the STI-HRQoL toolkit, and the Duke Activity Status Index (DASI) for assessing functional capacity related to QoL [14,15,18,20,25].

Data collection methodologies were tailored to the specific study settings and populations, employing face-to-face interactions, structured interviews conducted by trained researchers or clinicians, and self-

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

administered questionnaires [12,14,15,18,20]. For instance, large-scale studies in rural China and Vietnam primarily utilized interview-based administration to assist elderly or illiterate participants in differentiating between complex response levels and to ensure data completeness [11,13]. Conversely, studies in the UK and Poland often employed self-reporting or nurse-observed completions in clinical settings [14,21]. Furthermore, some evaluations utilized online platforms (such as Qualtrics) or digital dissemination to reach broader community-based samples [22,25].

Table 1. Characteristics of the Included Studies

Author (Year)	Country	QoL Instrument	Mode of Administration	No. Items	Data Collection Timepoint(s)	Target Population	Sample Size (n)
Aueyingsak (2026)	Thailand	emPHasis-10	Face-to-face	10	Baseline & 1 week	Pulmonary Arterial Hypertension (PAH)	20
Dave (2021)	India	MINICHAL (Gujarati)	Self-administered	17	Baseline & 48 hours	Essential Hypertension	36
Djuric (2023)	Serbia	SF-36 (Serbian)	Self-administered	36	Baseline & 2 weeks	Arterial Hypertension	169
Febriana (2023)	Indonesia	WHOQOL-BREF (Modified)	Interview	41	*Cross-sectional*	Elderly with Hypertension	30
Jiang (2021)	China	EQ-5D-5L & EQ-5D-3L	Face-to-face interview	5	*Cross-sectional*	Rural Hypertensive Patients	11,412
Jordan (2022)	UK	MINICHAL (English)	Self-administered	17	Weeks 0, 8, 10, 18	Treatment-naïve Grade II-III Hypertension	30
Liu (2023)	China	QLICD-HY (V2.0)	Inpatient questionnaire	41	Admission, 1-2 days, & Discharge	Hypertensive Inpatients	370
Mai (2022)	Vietnam	EQ-5D-5L	Face-to-face interview	5	*Cross-sectional*	Hypertensive population	1,296
Miłek (2026)	Poland	emPHasis-10 (Polish)	Observed by nurse	10	*Cross-sectional*	Pulmonary Arterial Hypertension (PAH)	120
Mustafaoglu (2023)	Turkey	DASI (Turkish)	Face-to-face interview	12	Baseline & 7 days	Pulmonary Hypertension (PH)	109
Rawlings (2024)	UK/Global	emPHasis-10	Online (Qualtrics)	10	*Cross-sectional*	Adults with Pulmonary Hypertension	263
Shi (2022)	China	emPHasis-10 (Chinese)	Administered	10	*Cross-sectional* (risk follow-up at 6 & 12 months)	CTD-associated PAH (CTD-PAH)	75
Smith (2023)	South Africa	STI-HRQoL	Paper & Digital	37 & 25	Baseline & 3 weeks	Hypertension, T2DM, & CVD	257

Established Psychometric Properties by Types of Hypertension

Psychometric adequacy, as presented in Table 2, was assessed according to the COSMIN criteria for good measurement properties, with a "sufficient (+)" rating required for each relevant domain, including content validity, internal consistency, test–retest reliability. Under these criteria, six of the 13 instruments (46.1%) demonstrated adequacy in the most frequently reported domains. However, it is important to emphasize that this does not imply validation, as none of the instruments met all COSMIN criteria across every measurement property evaluated. In the pulmonary hypertension (PH) and pulmonary arterial hypertension (PAH) category, the Thai emPHasis-10 and the Turkish Duke Activity Status Index (DASI) demonstrated established properties [12,15]. In patients with essential or arterial hypertension, instruments with established psychometric properties included the English and Gujarati versions of MINICHAL, the Serbian SF-36, the QLICD-HY (V2.0), and the STI-HRQoL toolkit [14,18–20,25].

Content Validity

All instruments underwent rigorous cross-cultural adaptation, but this does not automatically establish psychometric validity. The psychometric evaluations that followed these adaptations revealed varying levels of evidence across domains. The majority of the instruments (13, 100%) were rated as sufficient (+) because they followed rigorous cross-cultural adaptation or development procedures.

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

Table 2. Overall Rating and Quality of Evidence of Content Validity, Internal Consistency, and Test-Retest Reliability for Hypertension Instruments

Instrument / Study (Year)	Disease	Content Validity (Process/Expert Review)			A Priori Hypothesis	Internal Consistency			Test-retest Reliability		
		Description	Rating	GRADE		Alpha	Rating	GRADE	ICC	Rating	GRADE
emPHasis-10 / Aueyingsak (2026)	PAH	Forward-back translation, 5 experts, IOC	+	Low	No	0.925	?	Low	0.996	+	Low
MINICHAL / Dave (2021)	Essential HTN	Forward-back-forward, 6 experts, consensus	+	Low	Yes	0.879	+	Low	0.988	+	Low
SF-36 / Djuric (2023)	Arterial HTN	Serbian translation and validation	+	High	Yes	0.897	+	High	> 0.900	+	Low
WHOQOL-BREF Mod / Febriana (2023)	Elderly HTN	10 nursing experts, pilot test (n=5)	+	Low	No	> 0.60*	?	Low	NR	NR	NR
EQ-5D-5L & 3L / Jiang (2021)	Rural HTN	Official translations by EuroQol	+	High	-	NR	NR	NR	NR	NR	NR
MINICHAL / Jordan (2022)	Grade II-III HTN	Forward-back, panel consensus, cognitive debrief	+	Low	No	0.75–0.81	?	Low	0.71–0.96	+	Low
QLICD-HY / Liu (2023)	HTN Inpatients	Focus groups, nominal discussions, pilot	+	High	Yes	0.70–0.88	+	High	0.77–0.92	+	High
EQ-5D-5L / Mai (2022)	Hypertensive	Preference-based value set (Vietnam)	+	High	-	NR	NR	NR	NR	NR	NR
emPHasis-10 / Miłek (2026)	PAH	Forward-back, bilingual translator	+	High	Yes	0.94	+	High	NR	NR	NR
DASI / Mustafaoglu (2023)	PH	Forward-back, expert committee (n=5)	+	High	Yes	0.99	+	High	0.98	+	High
emPHasis-10 / Rawlings (2024)	PH	Factor analysis (EFA/CFA) support	+	High	No	0.85–0.93	?	High	NR	NR	NR
emPHasis-10 / Shi (2022)	CTD-PAH	Forward-back, medical translator	+	Mod.	No	0.95	?	Mod.	NR	NR	NR
STI-HRQoL / Smith (2023)	HTN, CVD	Delphi technique (12 HCPs), pilot	+	High	Yes	0.92–0.94	+	High	0.94	+	High

Notes: Content validity was defined as the degree to which the content of an instrument adequately reflected the construct to be measured and was evaluated in terms of relevance, comprehensiveness, and comprehensibility. Internal consistency was defined as the degree of interrelatedness among the items. A sufficient rating (+) for internal consistency required evidence of sufficient structural validity and a Cronbach's alpha value of ≥0.70 for each unidimensional scale or subscale. Test–retest reliability was defined as the extent to which scores remained stable among patients whose health status had not changed. A sufficient rating (+) for reliability required an intraclass correlation coefficient or weighted kappa value of ≥0.70.

Rating criteria: Sufficient (+), the result met the COSMIN criterion for good measurement properties; insufficient (−), the result did not meet the COSMIN criterion; indeterminate (?), insufficient information was available to determine the rating; not reported (NR), the measurement property was not evaluated or reported in the study.

Quality of evidence: The quality of evidence for each measurement property was evaluated using the modified GRADE approach and categorized as high, moderate, low, or very low. Evidence was downgraded based on risk of bias, inconsistency, imprecision, and indirectness.

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

Unlike reliability properties that require specific sample size thresholds for statistical precision, the quality of evidence for content validity was maintained as 'High' for instruments that followed comprehensive cross-cultural adaptation protocols. This approach recognizes that the rigor of the qualitative process, including expert committee review, forward–backward translation, and cognitive debriefing, is a stronger determinant of content validity than the total sample size. These procedures typically included forward–backward translation and expert panel reconciliation to ensure semantic and conceptual equivalence, as seen in the emPHasis-10 (Thai, Polish, and Chinese versions), MINICHAL, and DASI [12,14–16,18,21]. Furthermore, instruments such as QLICD-HY and STI-HRQoL received very good ratings due to rigorous development processes involving in-depth patient interviews, focus groups, or Delphi techniques with healthcare practitioners [20,25]. Cognitive debriefing was also utilized for tools like the English MINICHAL to ensure items were easily understood by the target population [14].

Internal Consistency

Internal consistency was assessed for 11 (84.6%) instruments, all of which were rated sufficient (+) based on Cronbach's Alpha values ≥ 0.70. High internal consistency was notably demonstrated by the Turkish DASI (0.99) and the Chinese emPHasis-10 (0.95) (Mustafaoglu et al. 2021; Shi et al. 2022). However, following the COSMIN Risk of Bias guidelines, only six instruments (46.1%) were rated as sufficient (+) because they established pre-specified a priori hypotheses for reliability. Several studies, such as Febriana et al. (2023), Aueyingsak et al. (2026), and Jordan et al. (2022), were assigned an indeterminate (?) rating because they failed to explicitly define reliability hypotheses before analysis, despite reporting high alpha coefficients. The only instruments not reporting internal consistency data were the EQ-5D-5L studies, which focused on preference-based values and known-groups validity (Jiang et al. 2021; Mai et al. 2022).

Test-retest Reliability

Test–retest reliability was assessed for six instruments. Among these, all six instruments were rated sufficient (+) and had ICC values ≥ 0.70. This included the Thai emPHasis-10, which demonstrated an exceptional ICC of 0.996, the Gujarati MINICHAL (0.988), and the DASI (0.98)[12,15,18]. The STI-HRQoL (0.94) and QLICD-HY (0.77–0.92) also showed high stability over time [20,25]. However, the quality of evidence for this property was frequently downgraded to a "Low" GRADE in studies such as Aueyingsak (2026), Dave (2021), and Jordan (2022) According to COSMIN guidelines, test-retest reliability studies with sample sizes below 50 participants are considered to have 'Low' GRADE evidence quality due to insufficient precision, despite demonstrating adequate ICC values. This threshold is critical because small cohorts may not provide the statistical power necessary to produce stable Intraclass Correlation Coefficient (ICC) estimates, thereby increasing the risk of bias in the assessment of measurement stability [12,14,18]. Consequently, Only the QLICD-HY and STI-HRQoL provided evidence with a "High" GRADE, as they employed larger samples of 121 and 257 patients, respectively [20,25].

Table 3. Study-Level Ratings and Quality of Evidence for Convergent Validity, Divergent Validity, Discriminative Validity, and Responsiveness

Author (Year)	Instrument	Convergent Validity		Divergent Validity		Discriminative Validity		Responsiveness	
		Rating	GRADE	Rating	GRADE	Rating	GRADE	Rating	GRADE
Aueyingsak (2026)	emPHasis-10	NR	-	NR	-	NR	-	NR	-
Dave (2021)	MINICHAL	NR	-	NR	-	NR	-	NR	-
Djuric (2023)	SF-36	+	High	NR	-	+	High	NR	-
Febriana (2023)	WHOQOL Mod.	NR	-	NR	-	NR	-	NR	-
Jiang (2021)	EQ-5D-5L/3L	NR	-	NR	-	+	High	NR	-
Jordan (2022)	MINICHAL	+	Low	NR	-	-	Low	+	Low
Liu (2023)	QLICD-HY	+	High	+	High	NR	-	+	High
Mai (2022)	EQ-5D-5L	NR	-	NR	-	+	High	NR	-
Miłek (2026)	emPHasis-10	+	High	NR	-	+	High	NR	-
Mustafaoglu (2023)	DASI	+	High	NR	-	NR	-	NR	-
Rawlings (2024)	emPHasis-10	+	High	NR	-	+	High	NR	-
Shi (2022)	emPHasis-10	+	Mod.	NR	-	+	Mod.	NR	-
Smith (2023)	STI-HRQoL	NR	-	NR	-	NR	-	NR	-

Notes: Convergent validity was defined as the extent to which instrument scores were consistent with predefined hypotheses regarding correlations with measures of related constructs. Divergent validity was defined as the extent to which instrument scores were consistent with predefined hypotheses regarding weak correlations with measures of theoretically unrelated constructs. Discriminative validity, also referred to as known-groups validity, was defined as the instrument's ability to distinguish between predefined groups expected to differ in the construct being measured. Responsiveness was defined

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

as the instrument's ability to detect change over time in the construct being measured, based on predefined hypotheses concerning change scores or relationships with external criteria.

Rating criteria: Sufficient (+), results were consistent with the predefined hypotheses; insufficient (−), results were inconsistent with the predefined hypotheses; indeterminate (?), no hypotheses were defined or insufficient information was available for assessment; not reported (NR), the measurement property was not evaluated or reported in the study.

Quality of evidence: The quality of evidence for each measurement property was evaluated using the modified GRADE approach and categorized as high, moderate, low, or very low. Evidence was downgraded based on risk of bias, inconsistency, imprecision, and indirectness.

Convergent Validity

Convergent validity was a well-established property in this review, with seven instruments (53.8%) receiving a sufficient (+) rating, predominantly supported by 'High' quality evidence. These instruments included the Serbian SF-36, the English version of MINICHAL, QLICD-HY, the Turkish DASI, and the Polish, Chinese, and Global versions of emPHasis-10 [14–16,19,21,22,26]. In evaluating this domain, researchers consistently utilized globally recognized generic standards specifically the EQ-5D (3L and 5L versions) and the Short Form-36 (SF-36) as reference measures to confirm their a priori hypotheses [14,16,20,21]. The high success rate in this category suggests that these hypertension-specific and adapted tools have strong correlations with established measures of general health status across diverse geographic regions.

Divergent Validity

Divergent validity which assesses whether an instrument measures a construct that is theoretically distinct from unrelated constructs was the least frequently reported property. A significant majority of the instruments either failed to define specific hypotheses or omitted reporting results for divergent assessment. Only one instrument, the QLICD-HY, was rated as sufficient (+) with 'High' quality evidence for this property [26]. This widespread evidence gap indicates that while many tools are known to correlate with similar constructs, their ability to prove they are measuring unique psychological dimensions remains largely unverified in the current hypertension literature. The divergent findings between the Gujarati MINICHAL, which demonstrated excellent psychometric properties, and the English MINICHAL, which failed to demonstrate discriminative validity, may be attributed to several factors: (1) differences in sample size and statistical power (n=36 vs n=30), (2) differences in disease severity (essential hypertension vs treatment-naïve Grade II-III hypertension), (3) differences in cultural context affecting item interpretation and response patterns, or (4) differences in the specific validation procedures used [14,18]. This highlights the critical importance of not assuming cross-cultural equivalence, even when instruments are translated using rigorous methods. The same instrument may function differently in different populations, necessitating separate validation studies for each cultural context.

Discriminative Validity

Discriminative validity also referred to as known-groups validity, was more widely established than divergent validity, with six instruments (46.1%) receiving sufficient (+) ratings. These instruments including the Serbian SF-36, the EQ-5D-5L/3L in China, the EQ-5D-5L in Vietnam, and various emPHasis-10 versions demonstrated a clear capacity to distinguish between sub-groups based on clinical severity, comorbidities, or demographic factors [11,13,19,22]. Notably, the English MINICHAL was the only tool rated as insufficient (-) for this property, as it failed to find statistically significant associations between quality-of-life scores and patient characteristics such as age, gender, and BMI. This finding suggests that the instrument may lack sensitivity in certain contexts or that its discriminative ability was hampered by a small sample size in that specific population [14].

Responsiveness

We evaluated whether there was evidence of responsiveness for the included instruments in detecting clinical changes over time. For disease-specific tools, detecting meaningful change following therapeutic interventions is a critical benchmark for clinical interpretation. Among the 13 instruments, 11 did not report any evidence of responsiveness. Among the remaining instruments, only one evaluated responsiveness and provided benchmarks for clinical change [14].

The QLICD-HY demonstrated high-quality evidence (+) of responsiveness, with a Standardized Response Mean (SRM) of 0.85 for its total score, indicating a strong ability to detect changes after inpatient treatment [20]. The English version of MINICHAL was rated with low-quality evidence (+) due to a small

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

sample size (n=30), showing a trend toward improvement in HRQoL following 18 weeks of intensive treatment, particularly in its State of Mind (StM) domain (Cohen's d = 0.32) [14]. No other instruments in this review provided quantitative data regarding their responsiveness to clinical interventions.

Discussion

The 13 studies included in this systematic review demonstrate a substantial global academic and clinical effort to refine and validate Patient-Reported Outcome Measures (PROMs) specifically for hypertensive populations. The broad geographical distribution across three continents underscores the global priority of validating hypertension-specific QoL instruments[11–13,15,21]. Most of these studies were conducted in tertiary referral hospitals or university clinics, ensuring access to verified clinical diagnoses ranging from essential hypertension to complex pulmonary arterial hypertension (PAH) [12,20,21].

The target populations were clinically heterogeneous, reflecting the extensive global burden of hypertension as a major non-communicable disease [20]. While a substantial portion of the research focuses on general essential or arterial hypertension, there is a notable expansion into specific chronic domains such as pulmonary arterial hypertension (PAH) and connective tissue disease-associated PAH (CTD-PAH) [12,16,21]. The clinical heterogeneity of the included populations (essential hypertension, pulmonary arterial hypertension, and CTD-associated PAH) raises important questions about the generalizability of findings. Instruments validated in PAH populations, such as emPHasis-10 and DASI, may not capture the distinct HRQoL concerns of patients with essential hypertension, who face different symptom profiles, treatment burdens, and disease trajectories. QLICD-HY, validated on hypertensive inpatients in China, and STI-HRQoL, validated on outpatients with hypertension/T2DM/CVD in South Africa, showed strong psychometric properties, but their applicability to other hypertensive populations remains uncertain. This review also notes a shift toward evaluating demographics, such as elderly patients in Indonesia and rural populations in China and Vietnam, who often face distinct literacy and healthcare access barriers compared to urban groups [13,17]. The widespread use of interviewer administration reflects the need to adapt data collection strategies to participant characteristics, thereby supporting higher-quality psychometric assessment. [11,13,20].

A primary finding of this review is that the majority of instruments demonstrate a strong psychometric profile regarding content validity and internal consistency. For content validity, all instruments received sufficient (+) ratings because they adhered to rigorous cross-cultural adaptation procedures, including forward-backward translation and expert committee reconciliation to ensure semantic and conceptual equivalence [12,14,18,21]. The high content validity of the QLICD-HY and STI-HRQoL may be associated with the integration of patient interviews, focus groups, and Delphi consensus throughout the development process. [20,25].

The universal finding of adequate content validity (100%) in this review, driven by rigorous cross-cultural adaptation, presents a notable contrast to other chronic disease areas. For instance, a systematic review of PROMs specifically developed for Type 2 Diabetes Mellitus found that only 27% of 150 subscales demonstrated sufficient content validity, with the development process for 85% of the instruments rated as inadequate due to a lack of patient involvement. This suggests that while the hypertension research community has successfully standardized the adaptation and expert review of existing tools, other chronic disease fields continue to struggle with the initial qualitative rigor required to ensure instruments are truly relevant to patients' lived experiences [27].

Internal consistency was the most successfully established measurement property, with most instruments achieving Cronbach's alpha values above the recommended threshold of 0.70. Exceptional reliability was observed in the Turkish version of the DASI (0.99) and the Chinese emPHasis-10 (0.95), indicating very strong inter-item relatedness within their respective domains [15,16]. As demonstrated in the studies by Aueyingsak et al. (2026) and Jordan et al. (2022) Jordan et al. (2022), retest sample sizes below 50 participants can limit the certainty of evidence for test–retest reliability. This suggests that while initial validation is often successful, larger longitudinal studies like the South African STI-HRQoL and the Chinese QLICD-HY are still required to provide high-certainty evidence of measurement stability over time [20,25].

The evaluation of validity through hypothesis testing showed high success in convergent and discriminative (known-groups) assessments. Regarding convergent validity, seven instruments demonstrated sufficient (+) results, predominantly supported by 'High' quality evidence. Researchers consistently utilized generic standards, specifically the SF-36 or EQ-5D, as reference measures to confirm their hypotheses, a strategy that provide evidence for effective across different geographic regions [13,19,20]. Further support for these findings was provided by large-scale studies in China and Vietnam provided strong evidence for

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

known-groups validity, showing that the EQ-5D-5L could significantly distinguish quality of life based on disease severity, age, and comorbidities [11,13]. The finding that generic instruments (EQ-5D, SF-36) demonstrated strong known-groups validity but lacked condition-specific content raises questions about whether they adequately capture the full impact of hypertension on quality of life. In contrast, discriminative or known-groups validity, which evaluates the instrument's ability to distinguish between clinically distinct groups (e.g., disease severity levels), was successfully reported in 46.1% of instruments. This distinction is critical as it highlights that while most tools can differentiate between patient severities, their ability to confirm they are measuring a unique psychological construct remains less verified.

The failure of the English MINICHAL to demonstrate discriminative validity based on age, gender, and BMI suggests several theoretical and methodological implications. This discrepancy indicates that either (1) the instrument lacks sensitivity to capture clinically meaningful differences in these specific subgroups, (2) the sample size (n=30) was insufficient to detect modest but clinically relevant differences, or (3) the construct of HRQoL in treatment-naïve hypertensive patients may be less influenced by these demographic factors than previously assumed [14]. In contrast, the Gujarati version of MINICHAL demonstrated better psychometric properties with a slightly larger sample (n=36), highlighting that differences in clinical profiles such as the more severe Grade II-III hypertension in the UK study and cultural contexts significantly influence an instrument's discriminative capability [18].

Divergent validity remains the least reported property, as most studies either omitted the results or did not define specific hypotheses. Responsiveness also requires further attention, as most instruments did not provide evidence of their ability to detect clinically meaningful changes over time. Only the QLICD-HY and the English MINICHAL provided quantitative data on an instrument's ability to detect meaningful clinical improvements following intensive therapeutic interventions [20] [14]. Without robust responsiveness data and benchmarks like the Minimal Clinically Important Difference (MCID), the clinical utility of many current instruments remains largely limited to cross-sectional assessment rather than long-term therapeutic monitoring [14].

The pattern of strong internal consistency but limited responsiveness observed in this review mirrors findings from systematic reviews of HRQoL instruments in other chronic conditions. Evaluation of cardiovascular disease-specific (CVD) instruments reported that while 37.5% of tools demonstrated sufficient internal consistency with high-quality evidence, 40% of the instruments failed to provide any evidence of responsiveness, and none achieved high-quality evidence in that domain. This suggests that while cross-sectional validity is well-established across various chronic disease populations, the ability to detect clinically meaningful change essential for monitoring treatment efficacy remains a universal challenge in PROM development [28].

The absence of responsiveness data limits the clinical utility of these instruments for longitudinal monitoring of treatment efficacy, as clinicians cannot reliably determine whether changes in scores reflect true improvement or measurement error. Furthermore, responsiveness is a prerequisite for any instrument intended for 'evaluative applications,' where the objective is to measure change in health status over time. The widespread failure to report this property in hypertension PROMs means that these tools may not be suitable for their intended purpose in clinical trials or routine practice. From a broader perspective, utilizing instruments with unknown or poor responsiveness is considered unethical and a waste of research resources, as it leads to findings that cannot be trusted to reflect the actual efficacy of therapeutic interventions. This evidence gap ultimately prevents healthcare professionals from selecting the 'best PROM' for the target population in a methodologically sound and transparent manner [8].

The emergence of innovative tools like the STI-HRQoL in South Africa, which offers an integrated toolkit for hypertension, diabetes, and cardiovascular disease, indicates a trend toward more efficient patient assessment in resource-limited settings [25]. Additionally, deeper factor analysis of established scales like the emPHasis-10 identifying sub-domains such as "fatigue," "independence," and "breathlessness" provides a more nuanced understanding of patient functioning without increasing respondent burden [22]. Several methodological limitations in the included studies warrant consideration. The majority of studies (61.5%) used convenience sampling, which may introduce selection bias and limit generalizability. Additionally, six studies (46.1%) had sample sizes below COSMIN's recommended minimum of 100 participants for structural validity evaluation, increasing the risk of imprecise estimates. The absence of blinding in most validation studies may have introduced detection bias, particularly in interviewer-administered assessments. These methodological limitations reduce the certainty of evidence for several measurement properties and highlight the need for more rigorous, adequately powered validation studies.

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

Strengths of the Review

The primary strength of this review lies in its rigorous methodological framework, strictly adhering to the PRISMA 2020 statement for reporting and utilizing the COSMIN guidelines to evaluate the methodological quality of studies on measurement properties. This review provides a global perspective, identifying psychometric validations across 10 countries and three continents, including Asia (China, Thailand, Vietnam, India, Indonesia), Europe (UK, Poland, Serbia, Turkey), and Africa (South Africa). The clinical scope is notably diverse, covering a wide range of hypertensive conditions from essential and arterial hypertension to complex cases like pulmonary arterial hypertension (PAH) and connective tissue disease-associated PAH (CTD-PAH). Furthermore, the review is inclusive of specialized demographics, identifying instruments tailored for elderly populations and patients in rural settings. The primary strength of this review lies in its search strategy, which strictly adhered to PRISMA 2020 standards to ensure a systematic and exhaustive identification of the most relevant literature. Additionally, the use of a modified GRADE approach to assess the quality of evidence for each measurement property enhances the reliability of the findings for both clinical and research applications.

Limitations of the Review

Despite its strengths, this review faces several limitations, most notably geographic gaps outside the represented nations; while the scope is global, many regions remain underrepresented in hypertension-specific QoL validation research. The language restriction to articles published only in English or Indonesian may have resulted in the exclusion of relevant studies published in other languages, such as Mandarin, Thai, or Vietnamese. A critical consideration for future research is the limitation of cultural and linguistic generalization. Given that all instruments in this review were validated within specific cultural contexts, researchers and clinicians should prioritize tools with established cross-cultural validity for their specific target population. It is vital not to assume that an instrument validated in one country or clinical setting is directly applicable to another without rigorous local validation. Another limitation is the methodological heterogeneity of the included studies; the diverse array of instruments (ranging from generic tools like SF-36 to disease-specific measures like emPHasis-10) and varying psychometric methods made a meta-analysis unfeasible, requiring a narrative synthesis instead. Furthermore, many primary studies lacked complete reporting, particularly regarding divergent validity, which was the least reported property across the dataset. The quality of evidence for test-retest reliability was frequently downgraded due to the small sample sizes ($n < 50$) used in the retest phases of several pilot and preliminary validation studies. Finally, a major gap was identified in responsiveness data. The vast majority of the included instruments (11 out of 13) failed to provide evidence for this property, which is vital for detecting meaningful clinical changes over time.

Conclusion

Most instruments in this review demonstrated adequate reliability and content validity; however, evidence for convergent validity, discriminative validity, and responsiveness remains limited, and no single instrument fulfilled all COSMIN criteria for sufficient measurement properties across all domains. Based on the synthesized evidence, disease-specific instruments including QLICD-HY (V2.0) and STI-HRQoL demonstrated the highest quality evidence across multiple psychometric domains and are recommended for clinical and research use in their respective populations (Chinese hypertensive inpatients and South African patients with hypertension/T2DM/CVD). While the DASI and emPHasis-10 show promise for pulmonary hypertension populations, they require further validation before being applied to the broader essential hypertension population. The English MINICHAL should be used with caution given its insufficient discriminative validity, until further validation studies with larger samples confirm its psychometric adequacy. Ultimately, longitudinal validation and cross-cultural examination should be prioritized in future research to ensure these instruments are directly applicable to diverse hypertensive populations beyond their original cultural contexts. Future validation studies should explicitly test measurement invariance across different hypertensive subpopulations to confirm that instruments function equivalently across clinical contexts.

Conflict of Interest

The authors declare that there are no financial, personal, or professional conflicts of interest that could have influenced the conduct or reporting of this study.

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com

Acknowledgment

The authors sincerely acknowledge the Faculty of Pharmacy, Universitas Gadjah Mada, for providing academic support throughout the completion of this study. The authors also extend their appreciation to everyone who contributed to this research, either directly or indirectly.

Supplementary Materials

There are no supplementary materials accompanying this manuscript.

References

[1] Chantakeeree C, Sormunen M, Estola M, Jullamate P, Turunen H. Factors Affecting Quality of Life among Older Adults with Hypertension in Urban and Rural Areas in Thailand: A Cross-Sectional Study. Int J Aging Hum Dev 2022;95:222–44. https://doi.org/10.1177/00914150211050880.

[2] Amlak BT, Getie A, Messelu MA, Ayenew T, GebreEyesus FA, Emire MS, et al. Health-related quality of life among adult hypertensive patients globally: a systematic review and meta-analysis. BMJ Public Heal 2025;3:e001662. https://doi.org/10.1136/bmjph-2024-001662.

[3] Teoli D, Bhardwaj A. Quality of Life. National Institutes of Health.: StatPearls Publishing; 2023.

[4] Lightfoot CJ, Howell M, Smith AC. How to assess quality of life in persons with chronic kidney disease. Curr Opin Nephrol Hypertens 2021;30:547–54. https://doi.org/10.1097/MNH.0000000000000740.

[5] Schmidt M, Staemmler M. Comparison of instruments for measuring health-related quality of life and their applicability to ehealth applications. Stud Health Technol Inform 2021;283:202–8. https://doi.org/10.3233/SHTI210561.

[6] van Krugten FCW, Feskens K, Busschbach JJV, Hakkaart-van Roijen L, Brouwer WBF. Instruments to assess quality of life in people with mental health problems: a systematic review and dimension analysis of generic, domain- and disease-specific instruments. Health Qual Life Outcomes 2021;19:1–13. https://doi.org/10.1186/s12955-021-01883-w.

[7] Pequeno NPF, Pequeno NPF, Cabral NL de A, Marchioni DM, Lima SCVC, Lyra C de O. Quality of life assessment instruments for adults: a systematic review of population-based studies. Health Qual Life Outcomes 2020;18:1–13. https://doi.org/10.1186/s12955-020-01347-7.

[8] 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;27:1171–9. https://doi.org/10.1007/s11136-017-1765-4.

[9] Feng YS, Kohlmann T, Janssen MF, Buchholz I. Psychometric properties of the EQ-5D-5L: a systematic review of the literature. Qual Life Res 2021;30:647–73. https://doi.org/10.1007/s11136-020-02688-y.

[10] Hernández-Segura N, Marcos-Delgado A, Pinto-Carral A, Fernández-Villa T, Molina AJ. Health-Related Quality of Life (HRQOL) Instruments and Mobility: A Systematic Review. Int J Environ Res Public Health 2022;19. https://doi.org/10.3390/ijerph192416493.

[11] Mai VQ, Giang KB, Minh H Van, Lindholm L, Sun S, Sahlen KG. Reference data among general population and known-groups validity among hypertensive population of the EQ-5D-5L in Vietnam. Qual Life Res 2022;31:539–50. https://doi.org/10.1007/s11136-021-02959-2.

[12] Aueyingsak S, Khankaew K, Teerawattananant W, Singsangtam P, Sanmueang A, Yasud M, et al. Validity and reliability testing of the Thai version of the emPHasis-10 questionnaire for patients with pulmonary arterial hypertension. J Patient-Reported Outcomes 2026;10:1–8. https://doi.org/10.1186/s41687-026-01017-0.

[13] Jiang J, Hong Y, Zhang T, Yang Z, Lin T, Liang Z, et al. Comparing the measurement properties of the EQ-5D-5L and the EQ-5D-3L in hypertensive patients living in rural China. Qual Life Res 2021;30:2045–60. https://doi.org/10.1007/s11136-021-02786-5.

[14] Jordan AN, Anning C, Wilkes L, Ball C, Pamphilon N, Clark CE, et al. Cross-cultural adaptation of the Spanish MINICHAL instrument into English for use in the United Kingdom. Health Qual Life Outcomes 2022;20:1–12. https://doi.org/10.1186/s12955-022-01943-9.

[15] Mustafaoglu R, Demir R, Aslan GK, Sinan UY, Zeren M, Yildiz A, et al. Translation, cross-cultural adaptation, reliability, and validity of the Turkish version of the Duke Activity Status Index in patients with pulmonary hypertension. Pulmonology 2021;29:S18–24. https://doi.org/10.1016/j.pulmoe.2021.06.008.

[16] Shi Y, Dong X, Hu X, Weng L, Liu Y, Lai J, et al. Cross-cultural validation of the Chinese version of the EmPHasis-10 questionnaire in connective tissue disease patients with pulmonary arterial hypertension and its relationship with risk stratification. BMC Pulm Med 2022;22:1–7. https://doi.org/10.1186/s12890-022-02056-1.

[17] Febriana A, Pelfbrianti D, Ifansyah MN, Lestari DH. Validitas dan Reliabilitas Instrumen Kualitas Hidup Lansia dengan Hipertensi. Indones J Heal Promot 2023;6:1401–6.

[18] Dave Y, Sorani D, Patel M. Cross Cultural Adaptation, Translation, Validation and Reliability Analysis of Gujarati Version of Hypertension Quality of Life Questionnaire (MINICHAL). Int J Heal Sci Res 2021;11:10–6.

https://doi.org/10.52403/ijhsr.20211003.

[19] Nikolic A, Biocanin V, Rancic N, Duspara M, Djuric D. Serbian Translation and Validation of the SF-36 for the Assessment of Quality of Life in Patients with Diagnosed Arterial Hypertension. Exp Appl Biomed Res 2023;24:227–34. https://doi.org/10.2478/sjecr-2020-0073.

[20] Liu X, Bai G, Li H, Li S. Applying SF - 6D to measure health state utilities among the middle and old aged patients with hypertension in China. Health Qual Life Outcomes 2020;18:1–10. https://doi.org/10.1186/s12955-020-01598-4.

[21] Miłek MW, Tkaczyk D, Torbicki A, Orłowska J, Kurzyna M, Wo´zniak-Prus M. Validation Analysis of the Polish-Translated Version of EmPHasis-10 Health-Related Quality of Life Questionnaire in Patients with Pulmonary Arterial Hypertension. J Clin Med 2026;15:1–12.

[22] Rawlings GH, Gaskell C, Beail N, Thompson A, Armstrong I. Exploratory and confirmatory factor analysis of emPHasis-10: The health-related quality-of-life measure in pulmonary hypertension. Pulm Circ 2024;14:1–11. https://doi.org/10.1002/pul2.12378.

[23] Jin KH, Sun SG. The Development of a Specific Quality of Life Scale for Hypertensive Patients: Methodological Study. Korean J Adult Nurs 2022;34:523–36. https://doi.org/10.7475/KJAN.2022.34.6.523.

[24] Si Y, Guo L, Chen S, Zhang X, Dai X, Wang D, et al. Progressing towards the 2030 health-related SDGs in ASEAN: A systematic analysis. PLoS Med 2025;22:1–20. https://doi.org/10.1371/journal.pmed.1004551.

[25] Smith L, Morris-Eyton H. Development, validation and reliability of the Smith Toolkit for Integrated Health Related Quality of Life (STI-HRQoL). Heal Sport Rehabil Med 2023;24:4–10. https://doi.org/10.26659/pm3.2023.24.1.4.

[26] Liu Y, Chang Y, Wan D, Li W, Xu C, Wan C. Development and validation of a disease-specific quality of life measure QLICD-HY (V2.0) for patients with hypertension. Sci Rep 2023;13:1–12. https://doi.org/10.1038/s41598-023-39802-2.

[27] Terwee CB, Elders PJM, Langendoen-Gort M, Elsman EBM, Prinsen CAC, van der Heijden AA, et al. Content Validity of Patient-Reported Outcome Measures Developed for Assessing Health-Related Quality of Life in People with Type 2 Diabetes Mellitus: a Systematic Review. Curr Diab Rep 2022;22:405–21. https://doi.org/10.1007/s11892-022-01482-z.

[28] Li X, Li R, Li M, Hui X, Li J, Yao L, et al. A Systematic Review and Quality Assessment of Cardiovascular Disease–Specific Health-Related Quality-of-Life Instruments: Part II Psychometric Properties. Value Heal 2025;28:294–305. https://doi.org/10.1016/j.jval.2024.08.011.

Electronic ISSN : 2656-3088
Homepage: https://www.journal-jps.com