Respiratory Medicine 234 (2024) 107829

Contents lists available at ScienceDirect

Respiratory Medicine

journal homepage: www.elsevier.com/locate/rmed

Appraisal and evaluation of the quality of life in pulmonary arterial hypertension instruments: A systematic review using COSMIN methodology

Gangemi Irene [a], Cedrone Nadia [a], Lommi Marzia [b,*], Paolo Iovino [c], Vellone Ercole [d,e]

[a] Tor Vergata University, Rome, Italy
[b] Sapienza University, Rome, Italy
[c] Health Sciences Department, University of Florence, Florence, Italy
[d] Doctoral Program in Nursing Sciences and Public Health Department of Biomedicine and Prevention, Tor Vergata University of Rome, Italy
[e] Department of Nursing and Obstetrics, Wroclaw Medical University, Poland

ARTICLE INFO

Keywords:
COSMIN
Pulmonary arterial hypertension
Psychometrics
Quality of life
Validation study

ABSTRACT

Aim: The aim of this review was to identify specific instruments currently available for measuring quality of life in patients with pulmonary arterial hypertension (PAH) and to evaluate their psychometric properties in order to provide robust evidence for their application in clinical practice.
Background: Pulmonary arterial hypertension is a rare pulmonary vascular disorder predominantly affecting women aged 30–50 years. It leads to elevated pulmonary artery pressure, causing increased cardiac workload. Symptoms such as dyspnea and fatigue progressively deteriorate. Given the substantial impact on patient well-being, quality of life assessment is a critical concern. Generic quality of life measures often fail to capture the unique challenges associated with PAH. Therefore, identifying a PAH-specific quality of life instrument is essential for optimising patient management.
Design: A systematic literature review.
Methods: A systematic review was performed to assess the psychometric properties of quality of life instruments for PAH patients, following the 2018 Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines. The literature search was conducted across Pubmed, Scopus, CINAHL, EMBASE, and APA PsycINFO databases.
Results: This review included four quality of life instruments: CAMPHOR, LPHQ, emPHasis-10, and PAH-SYMPACT. CAMPHOR and PAH-SYMPACT received a GRADE A rating, while LPHQ and emPHasis-10 were rated GRADE B. Despite some sample size limitations, these instruments demonstrated varying degrees of internal reliability, validity, and content coverage for assessing quality of life in pulmonary arterial hypertension patients.
Conclusions: This review provides an overview of available tools for assessing quality of life in patients with pulmonary arterial hypertension. Critical evaluation of these tools highlights incomplete psychometric assessments and methodological limitations in reference studies. Future research should prioritise more rigorous methodologies to ensure comprehensive psychometric evaluations.

1. Background

Pulmonary arterial hypertension (PAH) is a rare, progressive condition characterised by pulmonary artery vasoconstriction and obstruction, leading to increased pulmonary pressure and right heart failure [1,2]. It primarily affects women aged 30–50, with an incidence of 1–9 per 100,000 individuals [3]. PAH is classified as Group 1 by the World Symposium on Pulmonary Hypertension (WSPH), which categorises pulmonary hypertension into five groups based on aetiology. Initially asymptomatic, the condition gradually progresses, with common symptoms such as dyspnoea and fatigue, followed by chest pain, syncope, and oedema [4,5]. The progressive nature of PAH severely affects patients' quality of life (QoL), particularly in physical, mental, and social domains. Improving QoL has become a critical focus in

* Corresponding author.
E-mail addresses: irene.16gangemi@gmail.com (G. Irene), nadiacedrone@gmail.com (C. Nadia), marzia.lommi@uniroma1.com (L. Marzia), paolo.iovino@unifi.it (P. Iovino), ercole.vellone@uniroma2.it (V. Ercole).

https://doi.org/10.1016/j.rmed.2024.107829
Received 21 June 2024; Received in revised form 23 September 2024; Accepted 2 October 2024
Available online 6 October 2024

clinical research, especially for long-term PAH care [6]. Generic instruments like the SF-36 and Nottingham Health Profile (NHP) have been used to assess QoL, but they often lack specificity for PAH-related concerns [7,8]. To address this, PAH-specific tools, such as the Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR), PAH-SYMPACT, Living with Pulmonary Hypertension Questionnaire (LPHQ), and emPHasis-10, have been developed [7,9–11]. Patient-Reported Outcome Measures (PROMs) are key instruments reflecting patients' perspectives on their condition. The COSMIN initiative provides standards for selecting appropriate measurement tools in research and clinical practice [12]. There is currently no gold standard for QoL assessment in PAH patients, though this would significantly enhance care. No systematic reviews have rigorously evaluated QoL instruments in PAH, making this review essential for identifying the most valid and reliable tool for clinical use.

This review aims to identify PAH-specific QoL instruments and summarise their measurement properties to provide evidence for their use in clinical practice.

2. Methods

2.1. Design

A systematic literature review was conducted according to COnsensus-based standards for the selection of health measurement instruments (COSMIN guidelines) [13]. All research steps were conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) [14]. The protocol of the systematic review was registered on PROSPERO International prospective register of systematic reviews under number CRD42023450683.

2.2. Information sources and search strategy

This review was conducted by enquiring the following databases: Scopus, CINAHL, EMBASE, APA PsycINFO and PubMed in March 2024. No restrictions were placed on study design, language, disease or participants. The year of publication was restricted by the databases we searched: 1980–March 2024 CINAHL, 1947–March 2024 EMBASE, 1806–March 2024 APA PsycINFO, 1996–March 2024 PubMed.

The PICO was formulated as follows: a) Population: Patients with pulmonary arterial hypertension; b) Intervention: self-report tools assessing quality of life through validation studies and psychometric measurement tests; c) Comparison of instruments through 2018 COSMIN criteria guidelines and d) Outcome: Establishing a recommendation level based on measurement properties. Specific queries were constructed for each database according to the 2018 COSMIN guidelines. The elements that constituted the query were: 1) the population, 2) the construct, 3) the instrument type and 4) the measurement properties of interest. These elements were combined with the Boolean operators AND, OR and NOT. Appendix A shows the query used for the PubMed database as an example.

2.3. Eligibility criteria

The inclusion criteria were the following: a) studies on the development and validation of quality-of-life measurement instruments in patients with pulmonary arterial hypertension, b) articles published in English, Italian and Spanish, c) peer-reviewed articles published in academic journals. No time limit was imposed on the search.

Exclusion criteria were as follows: a) qualitative and/or quantitative studies that did not have as their main objective the development, psychometric testing and validation of a new scale measuring quality of life in patients with PAH (e.g. surveys, cross-sectional or phenomenological designs, protocols or reviews), b) studies that tested the instrument in populations other than patients with PAH, c) studies that did not publish the instrument in the article. In these cases, the researchers would contact the authors to request the instrument. The study was excluded if no response was received.

2.4. Data extraction

In order to facilitate the evaluation process, two researchers (ML and IG) extracted data from studies. This data included: instrument name, authors and publication details (year & country), study type (validity or developmental), sample characteristics, number of items, response system and psychometric properties measured. These data were used by the review team to describe the characteristics of the studies and the psychometric characteristics of each instrument (Table 1).

2.5. Data synthesis and assessing evidence

The evaluation and synthesis process was divided into two phases: the first evaluating and summarising evidence on the quality of the development and validation studies and a second phase evaluating and summarising evidence on the measurement properties of the instruments. COSMIN checklists and the Excel® file made available by them were used to conduct the data synthesis.

The first phase was divided into 4 steps 1) two reviewers independently assessed the methodological quality of each study (COSMIN Box 1 of the Excel® file), examining the relevance of the instrument items and the completeness and comprehensibility of the cognitive interview or pilot study used in the study; 2) two reviewers, separately, assessed the quality of the validation studies (COSMIN Box 2 of the Excel® file), examining the relevance, completeness and comprehensibility of the PROM items, by patients or experts; 3) the evidence of the studies was summarised and the instruments were evaluated to determine an overall score on relevance, comprehensiveness, comprehensibility and content validity (from sufficient to indeterminate); 4) finally, the confidence in the quality of evidence the reliability of the overall scores (high, moderate, low or very low) on was determined using the modified Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach.

The second phase comprised 4 steps: 1) two reviewers independently assessed the methodological quality of each study (COSMIN Box 3 of the Excel file); 2) two reviewers, separately, assessed each measurement property according to the criteria of the COSMIN checklist (COSMIN Box 4 of the Excel® file was used for this purpose); 3) the evidence for each instrument was finally summarised with an assessment of their psychometric properties (from sufficient to indeterminate) and the quality of the evidence (high, moderate, low and very low) using the GRADE approach; 4) finally, according to the quality of the evidence, levels of evidence A, B and C were assigned, according to the criteria of the Cosmin guidelines 2018 as follows: (i) level C when a PROMs has high quality evidence for an insufficient measurement property; (ii) level A when the PROMs shows evidence for sufficient content validity (any level) and at least low quality evidence for sufficient internal consistency, and (iii) level B when the PROMs cannot be classified with either level C or level A.

3. Results

3.1. Study characteristics

3.1.1. Characteristics and methodological quality of studies

A total of 591 records were retrieved. After removing duplicates (n. 45) 546 records remained. A further 521 records were removed from the title and abstract because they did not meet the inclusion criteria. A total of 25 full-text articles were found and 5 of these were eliminated. A total of 20 studies were included in the review (see Fig. 1).

These articles were 2 development, 16 validation and 2 mixed (both development and validation).

Fourteen studies were included for the CAMPHOR scale, consisting

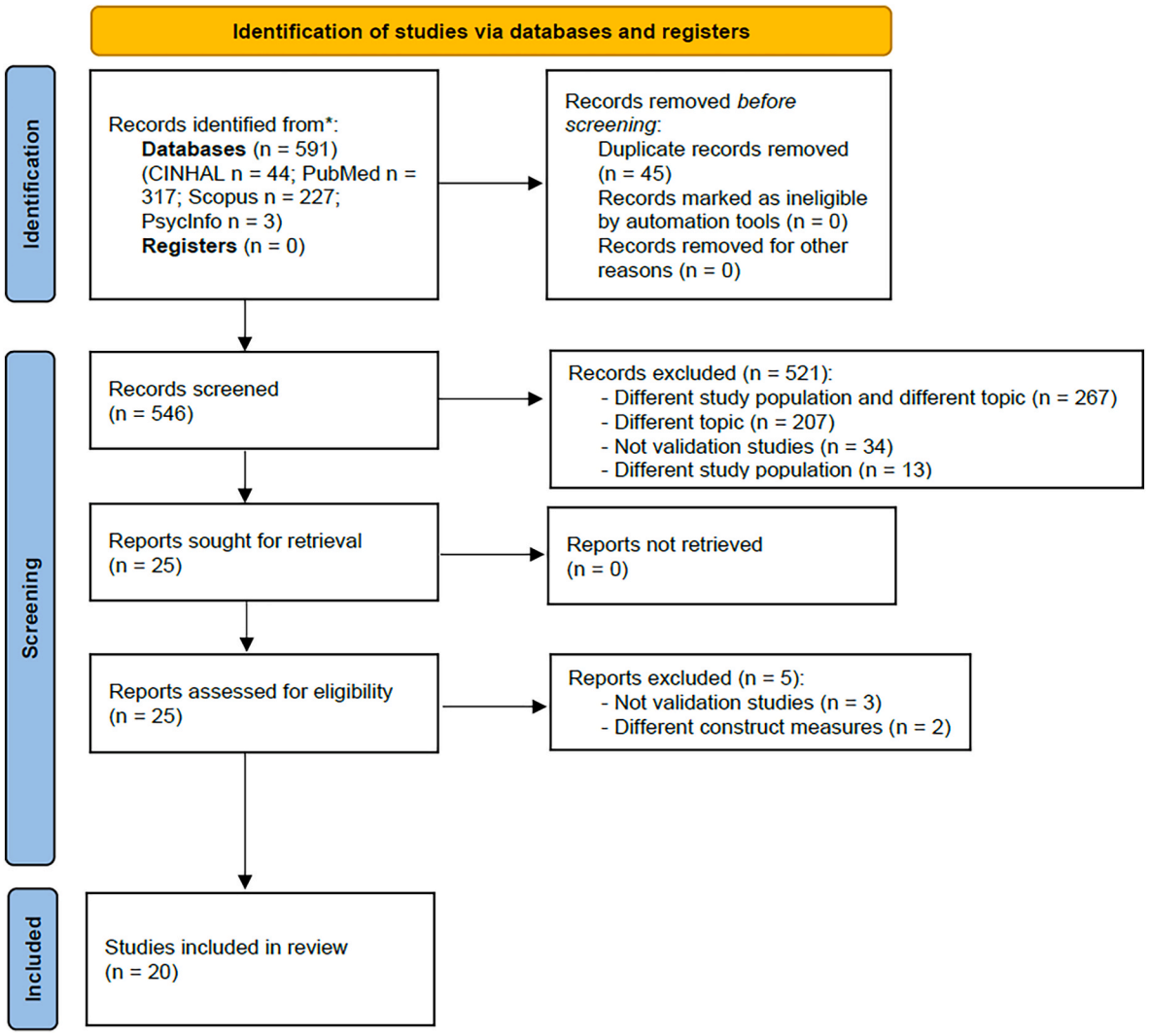

Fig. 1. PRISMA flowchart.

of one for development and validation, and thirteen for validation. The LPHQ scale had one article for development and validation, while the emPHasis-10 scale had two studies (one for development and one for validation). Lastly, the PAH-SYMPACT scale had three studies (one for development and two for validation).

Twelve studies were conducted in Europe (three in the United Kingdom, one in Austria, Germany and Switzerland, one in Spain, one in Croatia, two in Poland, one in Portugal, one in Sweden, one in Netherlands, one in Germany and France); one study was conducted in Asia (Turkey), six studies were conducted in America (one in Canada, four in the United States, one in Brazil) and one in Oceania (Australia and New Zealand) (Table 1).

Overall, one study was of adequate quality (McKenna et al., 2006), while the other 19 studies were of doubtful quality due to areas of bias inherent in the instrument development procedures (doubtful presence of a trained moderator/interviewer; lack of an interview guide in the article; a doubtful process of recording/transcribing participants' responses; doubtful independence of the data decoding process and doubtful achievement of data saturation) (Table 1).

Furthermore, in the pilot tests, the bias was due to the questionable relevance, completeness, or clarity of the items for the interviewees, the low number of participants enrolled in the pilot test/panel of experts and the frequent absence of the assessment of relevance, completeness or clarity by professionals.

3.1.2. Content validity, psychometric property assessment and level of evidence

The tools included in this review were CAMPHOR, LPHQ, emPHasis and PAH-SYMPACT. The longest developed instrument was the CAMPHOR (2006) and the most recent the PAH-SYMPACT (2016). Overall, two instruments received a GRADE A rating (CAMPHOR, PAH-SYMPACT) and two instruments were assigned a GRADE B (LPHQ, emPHasis-10) because, in most of them, the sample size did not meet the requirement of at least 7 times the number of items and $\geq$100 or at least 5 times the number of items, but $<$100 (Table 2). The Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) was developed and validated by McKenna et al. (2006) [7]. This scale is derived from unstructured interviews conducted on patients and consists of 65 items divided into three subscales: General Symptoms Scale (25 items with a dichotomous response system, total score 0–25); Functional Scale (15 items with a three-point response system, total score 0–30); Quality of Life Scale (25 items with a dichotomous response system, total score 0–25). The Symptoms and Quality of Life scales have a dichotomous response options ('True'/'Not true'), whereas the Functional scale uses three-point response options ('Able to do it alone without difficulty'/'-Able to do it alone with difficulty'/'Not able to do it alone'). This scale has shown excellent internal reliability in different studies [8,15–24].

In the study by McKenna et al. (2006) [7] concurrent validity of the CAMPHOR was assessed with the Nottingham Health Profile (NHP). The scales that are more closely related show stronger associations, (NHP

Table 1
Studies included in the review and psychometric properties of the instruments evaluated.

Tools	Author/Year Publication Country Type of Study	Sample	N. of Items Subscale Response System	Validity	Internal Consistency	Other Psychometric Properties	Methodological quality of studies
CAMPHOR	McKenna et al., 2006 [7] United Kingdom (Europe) Development and validation study	869 patients with PAH (63 % female; mean age 56.6 yy, SD 15.4 yy) *	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive interviews with 35 PAH-patients for items generation Qualitative interviews with 15 PAH patients for measures relevant, comprehensible, easy and quick to complete items' questionnaires Rash analysis for assessing unidimensional scales (good fit) Convergent validity: a) with NHP: higher levels of association (≥0.84) with NHP Energy/CAMPHOR Energy, NHP Emotional reactions/CAMPHOR Mood and NHP Physical mobility/ CAMPHOR Functioning b) with EQ-5D: negative and closely correlated with CAMPHOR Functioning (−0.74) and with. CAMPHOR Energy and total symptoms (EQ-VAS = −.0.71)	α Cronbach 0.76–0.92	*Test–retest reliability (interval time 14 days) Spearman rank ≥ 0.85* Known groups validity (stratified on NYHA classification): correlation from 0.30 to 0.62	Adequate
CAMPHOR	Meads et al., 2008[26] United Kingdom (Europe) Validation study	869 patients with PAH (63 % female; mean age 56.6 yy, SD 15.4 yy) *	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)			Responsiveness (Interval time 365 days): effect sizes small, except for the symptom scale (moderate). With the exception of the CAMPHOR functioning and QoL scales, these changes were significant. Known groups validity (stratified on NYHA classification): significant ($p < 0.01$) class II and III in all scales and in the class IV only for Symptoms and Functioning scales	Doubtful
CAMPHOR	Coffin et al., 2008 [27] Canada (North America) Validation Study	93 patients (41 for French-Canadian [FC] version and 52 for English-Canadian [EC] version) (77 % female, mean age 54.6 yy, SD 16.4 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive interviews with 30 PAH patients for assessing face and content validity (comprehension, relevance and comprehensiveness), 15 PAH patients for each version of scale. Convergent validity: a) with the five scales for the FC CAMPHOR (correlation from 0.31 to 0.92) and for the EC CAMPHOR (from 0.17 to 0.87). All scales were statistically significant except for the mood and breathlessness scales (for both FC and EC versions). b) with the NHP (strong correlations between the CAMPHOR and NHP energy scales, the CAMPHOR mood scales and the NHP	α Cronbach 0.69–0.92	Bilingual and lay panel translation for FC and EC versions of CAMPHOR Test–retest reliability (interval time 14 days): adequate level of reliability (≥0.73) except the EC mood scale (=0.64) Known groups validity (stratified for general health, use oxygen): a) with the exception of the mood and breathlessness on the FC version, all subscales showed an ability to distinguish appropriately between groups, on the basis of self-perceived general health b) with the exception	Doubtful

(*continued on next page*)

Table 1 (*continued*)

Tools	Author/Year Publication Country Type of Study	Sample	N. of Items Subscale Response System	Validity	Internal Consistency	Other Psychometric Properties	Methodological quality of studies
				emotional reactions scales, and the CAMPHOR functioning scales and the NHP physical mobility scales).		of the energy subscale on the FC version and the mood subscale on the EC and FC versions showed an ability to distinguish between participants on the basis of whether or not they had to use oxygen	
CAMPHOR	Gomberg-Maitland et al., 2008 [20] USA (North America) Validation Study	147 patients (84.0 % women, mean age of 50 yy, SD 14.6 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive interviews with 15 PAH patients for assessing face and content validity (comprehension, relevance, and perceived redundancy) Convergent *validity*: a) with 6MWT: significantly correlation (p 0.01), largest correlation with Functioning (r 0.45), lower levels of association for Symptoms (r 0.35) and QoL (r 0.33) scales b) with SF-36: significant (p 0.01) and most strongly correlations with Symptoms, Functioning and QoL scale.	α Cronbach 0.78–0.95	Bilingual and lay panel translation Test–retest reliability: (Interval time 14 days) 0.81–0.84 Known groups validity: a) WHO classification: class III had worse scores on the Breathlessness and on overall Symptoms, Functioning, and QoL compared with those in class II b) with perception general health: patients who reported worse health scores had significantly higher CAMPHOR scores than those with better health (all p 0.001)	Doubtful
CAMPHOR	Ganderton et al., 2011 [21] Australia and New Zealand (Oceania) Validation Study	61 patients with PAH (79 % women, mean age 56.9 yy, SD 14.5 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive interviews with 15 PAH patients for assessing face and content validity (relevant, terminology and language appropriate and understandable, difficulties during completion) Convergent validity with SF-36: symptom and QoLscales had the strongest correlations with the vitality, mental health and social functioning domains of the SF-36. The Functioning scale had the strongest correlations with SF-36 functional domains (physical functioning and role physical).	α Cronbach 0.89–0.92	Test–retest reliability (Interval time 14 days): correlation sufficient and significant for all scales (symptoms 0.86, functioning 0.87 and QoL 0.94, all $p < 0.01$). Known groups validity (stratified for WHO functional class): III and IV classes, had significantly higher CAMPHOR scores for all three scales ($p < 0.05$) compared with WHO functional classes I and II	Doubtful
CAMPHOR	Cima et al., 2012 [16] Austria, Germany and Switzerland. (Europe) Validation Study	107 patients with PAH (56.6 female, mean age 60.3 yy, SD 15.1 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous	Cognitive debriefing interviews with 12 PAH patients for assessing face and content validity (applicability, comprehension, relevance and comprehensiveness) Convergent validity: a) with NHP: the Symptoms scale showed strongest correlations with the emotional reactions, energy level and physical mobility sections of the NHP. The Functioning scale showed most closely related to physical mobility and energy level of the NHP. The QoL scale was less	α Cronbach 0.93–0.94	Bilingual and lay panel translation Test–retest reliability (Interval time 14 days): Spearman's rank from 0.90 to 0.91. Known group validity a) with perception of general health and severity of PH: individuals with worse general health and worse PH had higher scores for the Symptoms, Functioning and QoL scales ($p < 0.001$) b) with oxygen use: all three CAMPHOR scales	Doubtful

(*continued on next page*)

Table 1 (*continued*)

Tools	Author/Year Publication Country Type of Study	Sample	N. of Items Subscale Response System	Validity	Internal Consistency	Other Psychometric Properties	Methodological quality of studies
			response, total score 0–25)	influenced by pain and social isolation of the NHP. b) with 6MWT: had the highest correlation with the Functioning scale.		were able to distinguish between participants based on whether or not participants received oxygen for their PH ($p \leq 0.003$) c) with NYHA classification: individuals in groups NHYA III and IV had higher CAMPHOR scores than those in NYHA I and II (Symptoms $p < 0.001$; Activity limitations $p < 0$ 0.001; QoL $p < 0.05$).	
CAMPHOR	Selimovic et al., 2012 [19] Sweden (Europe) Validation Study	38 patients with PH (52.6 % female, mean age 61.2 yy, SD 13.3 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive debriefing interviews with 14 PAH patients for assessing face and content validity (applicability, relevance and comprehensiveness) Convergent *validity:* a) with NHP: Overall Symptoms scale correlated most strongly with the NHP energy scale. Functioning scale correlated most strongly with the NHP scale to which is most similar. Several of the NHP sections had moderate correlations with the QoL scale indicating that multiple factors contribute to patients overall QoL. b) with 6MWT: Overall Symptoms scale $r = -62$, $p = 0.001$, Functioning scale $r = -0.53$, $p = 0.008$ and QoL scale $r = -0.62$, $p = 0.001$	α Cronbach 0.92–0.95	Bilingual and lay panel translation Known groups validity: a) with age: the Overall Symptoms scale scores were significantly higher in elderly patients ($p = 0.028$) b) with severity PAH: patients with APAH had significantly higher Symptoms scores than patients with IPAH ($p = 0.032$). c) with WHO functional class: patients with a higher WHO functional class (III and IV) had significantly worse scores on symptoms, Functioning and QoL when compared with patients in WHO functional class I and II ($p < 0.001$, $p = 0.001$ and $p = 0.004$)	Doubtful
CAMPHOR	Aguirre et al., 2016 [15] España (Europe) Validation Study	70 patients with PAH (80 % females; mean age, 49.2 yy, SD 13.30 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive debriefing interviews with 23 PAH patients for assessing face and content validity (relevance, comprehensiveness and acceptability) Convergent *validity with* NHP: Overall Symptoms scale showed a strong correlation with the NHP energy level (0,79, $p < 0.01$) and physical mobility sections (0,82, $p < 0.01$). Functioning scale showed the strongest correlation with the NHP physical mobility section (0.86, $p < 0.01$). QoL scores were associated with NHP-D (0.74, $p < 0.01$)	α Cronbach 0.90–0.91	Bilingual and lay panel translation Test–retest reliability (Interval time 14 days): Spearman's rank >0.87. Known groups validity: a) with perception of general health and severity of PH: participants who rated their health status as "very good/good" had significantly lower levels of symptoms and disability, as well as better QoL, than participants who rated their health status as "fair/poor". Similar differences were observed between participants who perceived their PH as "mild/moderate" and those who perceived it as "severe/very severe" c) with WHO	Doubtful

(*continued on next page*)

Table 1 (*continued*)

Tools	Author/Year Publication Country Type of Study	Sample	N. of Items Subscale Response System	Validity	Internal Consistency	Other Psychometric Properties	Methodological quality of studies
						functional class: participants in functional classes III-IV had the highest scores on all CAMPHOR scales, indicating they had more symptoms, disability, and poorer QoL	
CAMPHOR	Reis et al., 2016 [18] Portugal (Europe) Validation Study	50 patients with PAH (74 % female, mean age 47 yy, SD 14 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive debriefing interviews with 10 PAH patients for assessing face and content validity (relevance, acceptability, comprehensiveness and understandability) Convergent *validity* with NHP: Overall Symptoms scale showed a strong correlation with the NHP energy level, emotional reactions, physical mobility sections and NHP-D (≥0.78, $p < 0.01$). Functioning scale showed the strongest correlation with the NHP energy level and physical mobility section (≥0.76, $p < 0.01$). QoL scores showed the strongest correlation with the NHP emotional reactions, physical mobility section and NHP-D (≥0.77, $p < 0.01$)	α Cronbach 0.93–0.95	Bilingual and lay panel translation Test–retest reliability (Interval time 14 days): Spearman's rank ≥0.89. Known groups validity with perception of general health and severity of diseases: participants who rated their health status as "very good/good" had significantly lower levels of symptoms and disability, as well as better QoL, than participants who rated their health status as "fair/poor". Similar differences were observed between participants who perceived their PH as "mild/moderate" and those who perceived it as "severe/very severe"	Doubtful
CAMPHOR	Wapenaar et al., 2016 [28] Rotterdam and Amsterdam, Netherlands (Europe) Validation Study	76 patients (77.7 % female, mean age 56 yy; SD not indicated)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive debriefing interviews with 10 PAH patients for assessing face and content validity (comprehension and relevance) Convergent validity: a) with NHP: Overall Symptoms scale showed a strong correlation with the NHP energy level (0.71, $p < 0.01$). Functioning scale showed the strongest correlation with the NHP physical mobility section (0.76, $p < 0.01$). b) with Borg dyspnoea scores: Overall Symptoms scale showed a moderate correlation (0.51, $p < 0.01$) c) with 6MWT:Functioning scale showed the moderate correlation (−0.47, $p < 0.01$)	α Cronbach 0.89–0.91	Bilingual and lay panel translation *Test–retest reliability: (*Interval time 14 days): Spearman's rank ≥0.87 Known groups validity: a) with perception of general health and severity of diseases: patients with worse perceived general health and more severe PAH had higher scores for all three scales of the CAMPHOR ($p < 0.01$) b) with NYHA classification: class 3 showed significantly higher scores on all three CAMPHOR scales compared with patients in NYHA class 2 ($p < 0.01$)	Doubtful
CAMPHOR	Hećimović et al., 2019 [8] Croatia (Europe) Validation Study	50 patients (70 % female; mean age 52.8 yy; SD 14.4 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale	Cognitive debriefing interviews with 10 PAH patients for assessing face and content validity (acceptability, relevant and comprehensiveness) Convergent validity: a) with SF-36: Overall Symptoms scale showed a strong correlation with the Physical role and Vitality of	α Cronbach 0.92–0.95	Bilingual and lay panel translation *Test–retest reliability: (*Interval time 14 days): Spearman's rank ≥0.90 Known groups validity with perception of general health and severity of diseases:	Doubtful

(*continued on next page*)

Table 1 (*continued*)

Tools	Author/Year Publication Country Type of Study	Sample	N. of Items Subscale Response System	Validity	Internal Consistency	Other Psychometric Properties	Methodological quality of studies
			(15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	SF-36 ($r > -0,70$, $p < 0.01$); Functioning scale showed a strong correlation with Physical functioning ($r > -0,70$, $p < 0.01$); QoL scale showed a strong correlation with Vitality, Social functioning and Emotional well-being ($r > -0,70$, $p < 0.01$)		statistically significant differences in scores on all three CAMPHOR scales related to both self-perceived disease severity and perceived overall health ($p < 0.01$)	
CAMPHOR	Villaquirán et al., 2019 [29] USA (North America) Validation Study	81 patients with PAH (84 % female; mean age 49 yy; SD 15 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Cognitive debriefing interviews with 11 PAH patients for assessing face and content validity (applicability, relevance, comprehensibility) Convergent validity with SF-36: the Overall Symptoms and functioning scales had strong correlations with the SF-36 physical functioning and role-physical scales. The QoL scale had moderate to strong associations with the SF-36 scale scores	α Cronbach 0.89–0.92	Bilingual and lay panel translation *Test–retest reliability:* (Interval time 15 days): Spearman's rank ≥0.79 Known groups validity with perception of general health and severity of diseases: with patients with more severe PH scoring higher on each scale ($p < 0.01$)	Doubtful
CAMPHOR	Corrêa et al., 2020 [17] Brazil (South America) Validation Study	102 patients with PAH (80,4 % female; mean age 48.8 yy; SD 14.5 yy)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	Field test interviews with 12 PAH patients for assessing face and content validity (relevant, acceptable and meaningful) Convergent validity with NHP: Overall Symptoms and Functioning scales correlated most highly with the Physical Mobility section of the NHP. The QoL scale was most strongly associated with scores on the Emotional Reactions and Social Isolation sections of the NHP.	α Cronbach 0.92 for each scales	Bilingual and lay panel translation Test–retest reliability: (Interval time 14 days): Spearman's rank ≥0.87 Known groups validity with perception of general health and severity of diseases: Patients who rated their disease severity as 'Quite severe/Very severe' had significantly worse scores on all CAMPHOR scales than patients who rated their disease severity as 'Mild/ Moderate' ($p < 0.001$). Also, patients who considered their general health to be 'Fair/Poor' had worse scores on all CAMPHOR scales than patients who rated their health as 'Very good/Good' ($p < 0.001$).	Doubtful
CAMPHOR	Malaczynska-Rajpold et al., 2020 [30] Poland (Europe) Validation Study	56 patients (69.6 % female; mean age 57.1, SD not indicated)	Subscales: a) Overall Symptoms scale (Energy, Breathlessness and Mood subscales) (25 items with dichotomous response, total score 0–25); b) Functioning scale (15 items with	Cognitive debriefing interviews with 15 PAH patients for assessing face and content validity (applicability, comprehension, relevance and comprehensiveness) Convergent validity with NHP: Overall Symptoms scales correlated most highly with the Energy	α Cronbach 0.89–0.94	Bilingual and lay panel translation Test–retest reliability: (Interval time 14 days): Spearman's rank ≥0.81 Known groups validity with perception of general health and severity of diseases: patients who rated their disease severity	Doubtful

(*continued on next page*)

Table 1 (*continued*)

Tools	Author/Year Publication Country Type of Study	Sample	N. of Items Subscale Response System	Validity	Internal Consistency	Other Psychometric Properties	Methodological quality of studies
			three-point response options, total score 0–30); c) QoL scale (25 items with dichotomous response, total score 0–25)	section of the NHP (0.75, $p = 0.01$). The Functioning scale correlated most highly with the Physical mobility section of the NHP (0.86, $p = 0.01$). The QoL scale was most strongly associated with the Energy and Emotional Reactions sections of the NHP (both 0.72, $p = 0.01$)		as quite or very severe had significantly worse scores on all CAMPHOR scales than patients who rated their disease severity as mild or moderate. Respondents who considered their general health to be fair or poor had significantly worse CAMPHOR scores than patients who rated their health as good or very good.	
LPHQ	Bonner et al., 2013 USA [10] Germany and France (Europe) Development and validation study	190 patients with PAH (mean age: 48.1, SD = 16.3; Female: 76.8 %)	21 items with 6-point Likert scale (from 0 'No' to 5 'Very much) Total scale (score range 0–105) and 2 subscales: Physical dimension (score range 0–40, 8 items) and Emotional dimension (score range 0–25, 5 items). Higher score in total scale and subscales indicated that patients are more affected by their medical conditions.	Qualitative interviews with 38 PAH patients for item generation Cognitive debriefing interviews with 38 PAH patients for evaluate the content validity (understanding and relevance) CFA with moderate fit for emotional and physical subscales and poor fit for total scale (no further data indicated) Convergent validity with EQ-5D: LPH Emotional with the EQ-5D anxiety/depression is 0.59, with EQ-5D self-care is 0.24.	α Cronbach 0.87 for total scale and 0.89 and 0.92 for physical and emotional subscale	Responsiveness (Interval time 12 weeks) with WHO Functional Class, Six-minute walking test, Borg score. Difference between change group for LPH physical ($p = 0.0073$) and total score ($p = 0.0415$). Known groups validity with a) WHO functional class: LPH Emotional, Physical and Total scores were broadly worse for those subjects with more severe disease across clinical criteria b) Borg score: The highest Borg correlations were with the LPH Physical score ($r = 0.36$ and $r = 0.34$ respectively). Correlations with the LPH Emotional Score ($r = 0.11$ and $r = 0.15$), and the LPH Total Score ($r = 0.21$ and $r = 0.23$) were low.	Doubtful
emPHasis-10	Yorke et al., 2014 [11] United Kingdom (Europe) Development study	226 patients with PAH (mean age: 55, SD = 14; Female: 69 %)	10 items with a semantic 6-point differential scale (from 0 to 5). Total score from 0 to 50 and high scores indicate poor quality of life in PAH.	Qualitative interviews with 30 PAH patients for item generation Cognitive debriefing interviews with 14 patients and two family members to ensure that item was clear and easy to understand *Rash analysis* (item residual ± 2.5; Chi-squared p-value >0.05) Convergent *validity* with: a) MLHFQ modified ($r = 0.61$, $p < 0.001$) b) HAD tools ($r = 0.77$, $p < 0.001$) c) D-12 ($r = 0.74$, $p < 0.001$) d) 6MWD ($r = -0.40$, $p < 0.001$)	α Cronbach 0.90	Test-retest (interval time 7 days) ICC = 0.95 Known groups validity with WHO functional class: class II and III mean difference 10.9, 95%CI 7.3–14.5, $p < 0.001$	Doubtful
emPHasis-10	Odevoglu et al., 2019 [22] Turkey (Asia)	101 patients with PAH (mean age: 52.5, SD = 16.1; Female: 81.2 %)	10 items with a semantic 6-point differential scale (from 0 to 5). Total score from 0 to 50 and high scores	Pilot test for Turkey version: 5 patients PAH for comprehensibility Criterion validity with gold standard MLHFQ: Spearman coefficient 0.85, $p = 0.001$	α Cronbach 0.98	Backward and forward translation Test-retest (interval time 7 days) ICC = 0.97	Doubtful

(*continued on next page*)

Table 1 (*continued*)

Tools	Author/Year Publication Country Type of Study	Sample	N. of Items Subscale Response System	Validity	Internal Consistency	Other Psychometric Properties	Methodological quality of studies
	Validation study		indicate poor quality of life in PAH.				
PAH-SYMPACT	Mc Collister et al., 2016 [9] USA (North America) Development study	55 patients with PAH (mean age: 53.1, SD = 15.8; Female: 93 %)	41 items with 5-point Likert scales indicated severity of symptom or level of severity of impact or level of difficulty of action to do. 2 Subscales: Symptom's domains (16 items) and Impacts domains (25 items)	Cognitive interviews with 45 PAH-patients for concept elicitation Cognitive debriefing interviews with 10 patients for evaluate the content validity (clearness, comprehensibility, and relevance)			Doubtful
PAH-SYMPACT	Chin et al., 2018 [24] USA (North America) Validation study	278 patients with PAH (mean age: 60, SD = 13.4; Female: 79 %)	22 items with 5-point Likert plus one item on oxygen use Subscales: Symptom's domains (11 items) and Impacts domains (11 items)	Content validity: test for floor and ceiling effects and item-to-item correlations (correlation ≥0.7 indicating potential redundancy) CFA: acceptable model fit for symptom items (CFI = 0.861) and good model fit for impact items (CFI = 0.960). Convergent validity with SF-36: from −0.36 to −0.56 ($p < 0.001$) for symptom items and from −0.34 to −0.73 ($p < 0.001$) *for impacts items* *Concurrent validity with* CAMPHOR: from +0.38 to +0.68 ($p < 0.0001$) for symptom items and from +0.46 to +0.80 ($p < 0.0001$) for impacts items	Cronbach's alpha for 4 domains >0.80	Test-retest reliability analysis (24h for symptom items and 7 days for impacts items): ICC = 0.84–0.94 Known groups validity with WHO functional class): p = 0.0039 for Cardiovascular Symptoms, $P < 0.0001$ for the three other domains)	Doubtful
PAH-SYMPACT	Sarzyńska et al., 2021 [23] Poland (Europe) Validation study	55 patients with PAH (mean age: 56, SD = 17.25; Female: 78.8 %)	22 items with 5-point Likert plus one item on oxygen use 2 Subscales: Symptom's domains (11 items) and Impacts domains (11 items)	Cognitive debriefing interviews (declared but results were not reported Convergent validity with CAMPHOR (the authors declare that every domain of the PAH-SYMPACT, correlates with all domains of the CAMPHOR) and WHO-QOL BREF (results not declared)	Cronbach's alpha for 4 domains >0.70	*Forward and backward translation* *Test–retest reliability:* (Interval time not indicated) Spearman coefficient >0.9 ($P < 0,05$)	Doubtful

Note: *same sample in the study; CAMPHOR= Cambridge Pulmonary Hypertension Outcome Review; PAH: Pulmonary arterial hypertension; QoL: Quality of life; NHP: Nottingham Health Profile; EQ-5D = EuroQoL-5D; NYHA: New York Heart Association; EQ-VAS: European Quality of Life Visual Analog Scale; 6MWT: 6 Minute Walking Test; SF-36: Short Form Health Survey 36; WHO: World health organisation; APAH: associated pulmonary arterial hypertension; IPAH = idiopathic pulmonary arterial hypertension; NHP-D: Nottingham Health Profile index of Distress; LPHQ: Living with pulmonary hypertension; CFA: Confirmatory Factor Analysis; MLHFQ = Minnesota living with heart failure; HAD tools: Hospital Anxiety and Depression scale; D-12: Dyspnoea-12 questionnaire; 6MWD: 6-min walk distance; PAH SYMPACT: Pulmonary Arterial Hypertension-Symptoms and Impact; CFI: Confirmatory Factor Index.

Energy level with CAMPHOR Energy = 0.84; NHP Emotional reactions and CAMPHOR Mood = 0.84 and NHP Physical mobility and CAMPHOR Functioning = 0.85). Similarly, in the study by Aguirre and colleagues (2016) [15], concurrent validity was measured with NHP, which showed a strong correlation with energy levels (=0.79, $p < 0.01$); and physical mobility sections (=0.82, $p < 0.01$). Functioning scale showed the strongest correlation with the NHP physical mobility section (=0.86, $p < 0.01$). QoL scores were associated with Nottingham Health Profile index of Distress (NHP-D) (=0.74, $p < 0.01$). In the study by Gomberg-Maitland et al. (2008) [25], the CAMPHOR scores correlated with those of the SF-36 ($p = 0.01$; $p < 0.01$), as in the study by Ganderton et al. (2011) [26]. Hećimović and colleagues (2019) [8] for the assessment of concurrent validity also used the SF-36 reporting a statistically significant result ($r > -0.70$, $p < 0.01$). The GRADE rating of the CAMPHOR is A due to moderate evidence for sufficient content validity and low quality evidence for sufficient internal consistency.

The Living with Pulmonary Hypertension Questionnaire (LPHQ) was developed and validated by Bonner et al. (2013) [10].This scale is derived from the Minnesota Living with Heart Failure (MLHF), a scale specifically designed to assess QoL in patients with myocardial infarction. The LPHQ consists of 21 items on a 6-point Likert response options (0 = 'No' to 5 = 'Very much'). The total score of the questionnaire ranges from 0 to 105, with higher score meanings that patients are more negatively influenced by their medical condition. In addition to a total score, the scale has scores for the subscales emotional with a score from 0 to 25 (comprising 5 items) and physical with a total score from 0 to 40 (comprising 8 items). Similar to the total score, higher scores on the subscales indicate that the medical condition influences these patients.

Table 2
Evaluation of content validity and psychometric properties and development of recommendations for the development of the instruments.

Tool	Relevance	Comprehensiveness	Comprehensibility	Overall Content Validity	Structural Validity	Internal Consistency	Other Measurement	Recommendation
CAMPHOR	+/Moderate	+/Moderate	+/Moderate	+/Moderate	+/Low	+/Low	Reliability +/Low Construct validity +/Low Responsiveness +/Low	A
LPH Questionnaire	+/Moderate	+/Moderate	+/Moderate	+/Moderate	-/Low	?/Low	Construct validity +/Low Responsiveness +/Low	B
emPHAsis-10	+/Moderate	+/Moderate	+/Moderate	+/Moderate	-/Moderate	?/Moderate	Reliability +/Moderate Criterion validity +/Moderate Construct validity +/Moderate	B
PAH-SYMPACT	+/Moderate	+/Moderate	+/Moderate	+/Moderate	+/Low	+/Low	Reliability +/Low Construct validity ?/Low	A

Note: + = sufficient; − = insufficient; ? = indeterminate; M = moderate; L = low.

Bonner et al. (2013) [10] conducted a confirmatory factor analysis (CFA) which exhibited moderate fit for the emotional and physical subscales and a poor fit for the total scale. However, the fit indices were not reported by the authors. The internal consistency was satisfactory (Cronbach's $\alpha = 0.87$). For this reason, the GRADE of the instrument is level B, low quality evidence for an indeterminate internal consistency and negative structural validity of the instrument.

The emPHasis-10 scale, developed by York and colleagues (2014) [11], is a one-dimensional scale consisting of 10 items with a semantic 6-point differential scale (from 0 to 5). The total scale score ranges from 0 to 50 and high scores indicate a poor disease specific quality of life. This scale was developed in response to the need to use an instrument with fewer items, compared to CAMPHOR and LPH, which require a lot of time to administer, and find an instrument with strong discriminative abilities between subgroups of patients stratified according to World Health Organization functional class (WHO). EmPHasis-10 has been validated in Spanish, English, French, German and Italian, but none of the translation studies have been published and therefore could not be included in this review. EmPHasis-10 assesses important components of PAH, such as dyspnoea, fatigue and lack of energy, social restrictions and concerns about effects on others (e.g. family and friends). EmPHasis-10 has demonstrated excellent internal reliability Cronbach's $\alpha = 0.98$ (Odevoglu et al., 2019); Cronbach's $\alpha = 0.90$ (Yorke et al., 2014). In the study by York and colleagues (2014) concurrent validity was demonstrated by correlating its scores with those of the modified MLHFQ ($\rho = 0.61$, $p < 0.001$), Hospital Anxiety and Depression scale (HADs) ($\rho = 0.77$, $p < 0.001$) and the Dyspnoea-12 questionnaire (D-12) ($\rho = 0.74$, $p < 0.001$). The study by Odevoglu et al. (2019) [27] used the MLHFQ as the gold standard to assess criterion validity with EmPHasis-10, which showed a strong correlation ($\rho = 0.85$; $p < 0.001$); therefore, the criterion validity of the EmPHasis-10 questionnaire was considered high. However, the GRADE of the instrument is level B, low quality evidence for an indeterminate internal consistency and structural validity of the instrument, since the authors did not fully state the fit indices of the Rasch analysis.

The Pulmonary Arterial Hypertension-Symptoms and Impact (PAH-SYMPACT) was developed by Mc Collister and colleagues (2016) [9] with the aim of developing an instrument to assess PAH symptoms and their impact on quality of life. The scale was translated and validated in Polish [28] and originally consisted of 41 items [9] divided into two subscales (Symptoms domains and Impacts domains) with a 5-point Likert response options. In the study by Chin and colleagues (2018) [29], the items were reduced to 22, plus one investigating oxygen use. The scores for the subscales Symptoms and Impacts domains both range from 0 to 44. Higher scores indicating greater symptom severity or worse impact. This scale demonstrated good internal consistency, test-retest reliability and convergent validity [28,29]. The GRADE rating of the instrument is A because with moderate evidence for sufficient content validity and low quality evidence for sufficient internal consistency.

4. Discussion

The aim of this systematic review was to summarise the characteristics and psychometric properties of existing instruments measuring quality of life in patients with PAH.

There are instruments in the literature that can assess the generic quality of life of non-disease-specific patients such as the EuroQoL-5D, the SF-36 and the Nottingham Health Profile Index of Distress (NHP-D). These instruments have often been used to evaluate the concurrent validity of instruments assessing PAH-specific QoL.

The researchers' desire to devise instruments that would assess quality of life in a specific disease such as PAH emerged in the 2000s, specifically between 2006 and 2016. Among the scales developed, those that demonstrated better psychometric properties were the CAMPHOR and the PAH-SYMPACT. These new instruments are able to capture and

assess specific domains such as symptoms and their impact on the quality of life of these patients. The recent development of these instruments and the limited number of instruments found, which, among others, differ in the domains and symptoms explored and the number of items, suggests that further study is needed for additional disease-specific instruments. Overall, the symptoms assessed by each instrument were respiratory distress (breathlessness, shortness of breath, difficulty breathing, cough), energy levels (fatigue, lack of energy,

Table 3
Domains and scopes of the scales analysed.

Domains		emPHAsis-10	CAMPHOR	PAH-SYMPACT	LPHQ
Symptoms	Respiratory distress	Breathlessness	Breathlessness	Shortness of breath	Shortness of breath
		Shortness of breath	Shortness of breath	Cough	Difficulty breathing
			Difficulty breathing		
			Cough		
			Breathlessness in walking		
			Breathlessness climbing a step		
			Breathlessness even without doing anything		
			Breathlessness standing up		
	Levels of energy	Fatigue	Fatigue	Fatigue	Fatigue
		Lack of energy	Lack of energy	Lack of energy	Lack of energy
		Exhaustion	Tired very quickly		Tiredness
			Weakness		
			Stamina levels low		
	Cardiovascular problems		Swelling in ankles or legs	Swelling in ankles or legs	Swelling in the ankles or legs
				Heart palpitations	
				Rapid heartbeat	
				Chest pain	
				Chest tightness	
				Light-headedness	
	Gastrointestinal problems			Swelling in stomach area	
Impacts	Impact on Physical Wellness	Rest during the day			Rest during the day
		Difficulty for climbing stairs			Rest during and after activities
			Walking short distances on level ground	Walking slowly	Difficulty for climbing stairs
			Standing for a short time	Walk on a flat surface	Difficulty for walking
				Walk quickly on a flat surface	
				Walk uphill	
			Difficulty lifting heavy objects		
			Not knowing how to do things on the spot		
					Problems getting to sleep
			Not having control of one's body		
					Side effects from treatments
					Impact on eating/diet
				Carry thing	
	Impact on Mental wellness			Think clearly	Memory
					Concentration
	Impact on Emotional wellness	Loss of control of own life	Loss of control of own life		Loss of control of own life
			Depression	Sad	Depression
			Anxious	Worried	Worry
			Darkness	Frustrated	
			Forgotten what it's like to enjoy myself		
			The loss of a role in life		
			Self-esteem needs		
			General fear of the future		
			Fears of being left alone		
			Need for security		
	Impact on Social wellness	Feeling a burden	Feeling a burden		Feeling a burden
		Dependence on others	Dependence on others	Need help from others	Dependence on other
			Difficulty in relating to others		Difficulty in relating to others
			Doing things with friends and family		Doing things with friends and family
			Social restrictions		
		Confident in going out of the house			Going out of the house
	Impact on Financial wellness				Treatments burdensome
					High cost of medication
			Need for financial independence		
					Difficult to work to earn a living
	Impact on Activities daily living		Dress self	Wash self	
				Dress self	
				Difficulty doing light housework	To work around the house or in the garden
					Difficult making leisure activities
					Difficult doing sport
					Difficult making hobby

exhaustion, tired) and cardiovascular problems (heart palpitations, rapid heartbeat, chest pain, swelling in the ankles or legs). Only in PAH-SYMPACT an additional item was added to assess gastrointestinal problems (swelling in the stomach area).

Regarding the impacts of quality of life, we identified several key areas. For the social well-being domain, the instruments primarily focus on dependence on others and the feeling of being a burden. Notably, only the CAMPHOR and LPHQ questionnaires delve into the difficulty of maintaining relationships with friends and family. For physical well-being, most instruments assess basic limitations like difficulty climbing stairs or walking slowly. However, the CAMPHOR goes a step further by measuring the need to plan activities in advance and the perception of losing control over one's body. The LPHQ also assesses additional aspects like the impact on eating habits or diet, sleep problems, and side effects experienced from treatments. Regarding emotional well-being the instruments assess fear of loneliness and for the future, depression, anxiety, frustration, need for self-esteem.

Mental well-being is only assessed in the PAH-SYMPACT ('think clearly') and the LPHQ ('memory' and 'concentration').

The only instruments investigating financial wellbeing are the CAMPHOR ('need for financial independence') and the LPHQ ('difficult to work to earn a living', 'Treatments burdensome' and 'High cost of medication').

Finally, with the exception of emPHasis-10, in all other included instruments the impact on activities of daily living was investigated, 'dressing alone', 'washing alone', 'doing sports with difficulty', 'doing hobbies with difficulty'. See Table 3.

5. Limitations

One limitation of this review is the inclusion of peer-reviewed studies of academic journals and those written exclusively in English, Italian and Spanish. This may have generated biases (i.e. publication selection) due to the exclusion of studies that were published as grey literature and articles for the development or validation of PAH-specific QOL instruments published in languages other than those included in the review. The evaluation of the studies was based on the 2018 COSMIN guidelines. Some criteria necessary for a "very good" or "adequate" rating may not have been considered by the authors of older studies. This applies to both the methodological quality assessment (e.g., unclear presence of a trained moderator/interviewer, absence of an interview guide mentioned in the article) and the evaluation of psychometric properties (e.g., missing calculation of intraclass correlation coefficient (ICC) and unspecified fit indices for confirmatory factor analysis). This may have influenced the final evaluation of the tools. Finally, except for two studies, one on CAMPHOR and one on LPHQ, it was not possible to assess responsiveness, or the ability to detect a change in the construct measured over time, due to the lack of longitudinal studies among those included on the other tools.

Often the authors of the studies included in the review used Cronbach's alpha to assess internal consistency. However, according to recent guidelines for assessing internal consistency for multidimensional instruments, the use of the omega hierarchical coefficient is suggested [30]. This could be a limitation in assessing the psychometric properties of the instruments.

6. Conclusions

This review aims to provide an overview and understanding of the tools currently available to measure and assess the quality of life of patients with PAH. The clinical use of the included scales requires critical evaluation, in light of an incomplete psychometric evaluation and, sometimes, questionable or inadequate methodology in reference studies. However, two tools have been considered, according to the COSMIN methodology, to be Grade A (CAMPHOR and PAH-SYMPACT) and two to be Grade B (emPHasis-10 and LPHQ). The difference in the validation levels of the psychometric tools mainly depends on the rigour and completeness of the methodology used to evaluate the tool itself. It is necessary that future research focuses on the development of scales using a more rigorous methodology employing higher quality methods, and estimating all psychometric properties.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Funding statement

This research received no funding.

Ethics approval statement

The study was conducted in accordance with the standards and ethical principles of the Declaration of Helsinki and was approved by the board of directors of each participating degree site.

CRediT authorship contribution statement

Gangemi Irene: Writing – review & editing, Writing – original draft, Supervision, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Cedrone Nadia:** Visualization, Validation, Software, Project administration, Investigation, Conceptualization. **Lommi Marzia:** Writing – original draft, Resources, Methodology, Formal analysis, Conceptualization. **Paolo Iovino:** Writing – review & editing, Supervision. **Vellone Ercole:** Writing – review & editing, Validation, Supervision, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rmed.2024.107829.

References

[1] D.J. Levine, Pulmonary arterial hypertension: updates in epidemiology and evaluation of patients, Am. J. Manag. Care 27 (2021). https://doi:10.37765/ajmc.2021.88609.
[2] M. Humbert, G. Kovacs, M.M. Hoeper, R. Badagliacca, R.M.F. Berger, M. Brida, J. Carlsen, A.J.S. Coats, P. Escribano-Subias, P. Ferrari, D.S. Ferreira, H. A. Ghofrani, G. Giannakoulas, D.G. Kiely, E. Mayer, G. Meszaros, B. Nagavci, K. M. Olsson, J. Pepke-Zaba, J.K. Quint, J.K. Rådegran, G. Simonneau, O. Sitbon, T. Tonia, M. Toshner, J.L. Vachiery, A. Vonk Noordegraaf, M. Delcroix, S. Rosenkranz, ESC/ERS Scientific Document Group, ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension, Eur. Heart J. 43 (38) (2022) 3618–3731, 2022, https://doi:10.1093/eurheartj/ehac237.
[3] Orphanet. (n.d.). Epidemiology of PAH. https://www.orpha.net/it/disease/detail/182090?name=ipertensione%20arteriosa%20polmonare&mode=name.
[4] V.V. McLaughlin, S.J. Shah, R. Souza, M. Humbert, Management of pulmonary arterial hypertension, J. Am. Coll. Cardiol. 65 (18) (2015) 1976–1997, https://doi.org/10.1016/j.jacc.2015.03.540.
[5] A. Frost, D. Badesch, J.S.R. Gibbs, D. Gopalan, D. Khanna, A. Manes, R. Oudiz, T. Satoh, F. Torres, A. Torbicki, Diagnosis of pulmonary hypertension, Eur. Respir. J. 53 (1) (2019). https://doi:10.1183/13993003.01904-2018.
[6] C. Wan, R. Jiang, X. Tu, M. Tang, W. Pan, J. Yang, R. Li, X. Yang, Z. Zhang, The hypertension scale of the system of Quality of Life Instruments for Chronic Diseases, QLICD-HY: a development and validation study, Int. J. Nurs. Stud. 49 (4) (2012) 465–480, https://doi.org/10.1016/j.ijnurstu.2011.10.010.
[7] S.P. McKenna, N. Doughty, D.M. Meads, L.C. Doward, J. Pepke-Zaba, The Cambridge pulmonary hypertension outcome review (CAMPHOR): a measure of health-related quality of life and quality of life for patients with pulmonary

hypertension, Qual. Life Res. 15 (1) (2006) 103–115, https://doi.org/10.1007/s11136-005-3513-4.
[8] A. Hećimović, A. Heaney, S.P. McKenna, L. Basara, M. Jakopović, A. Vukić Dugac, G. Redžepi, C. Rotim, M. Samaržija, N. Jokić-Begić, S. Popović-Grle, Adaptation and validation of the cambridge pulmonary hypertension outcome review (CAMPHOR) for Croatia, Acta Clin. Croat. 58 (1) (2019) 3–12. https://doi:10.20471/acc.2019.58.01.01.
[9] D. McCollister, S. Shaffer, D.B. Badesch, A. Filusch, E. Hunsche, R. Schüler, I. Wiklund, A. Peacock, Development of the Pulmonary Arterial Hypertension-Symptoms and Impact (PAH-SYMPACT®) questionnaire: a new patient-reported outcome instrument for PAH, Respir. Res. 17 (1) (2016) 72, https://doi.org/10.1186/s12931-016-0388-6.
[10] N. Bonner, L. Abetz, J. Meunier, M. Sikirica, S.C. Mathai, Development and validation of the living with pulmonary hypertension questionnaire in pulmonary arterial hypertension patients, Health Qual. Life Outcome 11 (2013) 161, https://doi.org/10.1186/1477-7525-11-161.
[11] J. Yorke, P. Corris, S. Gaine, J.S.R. Gibbs, D.G. Kiely, C. Harries, V. Pollock, I. Armstrong, emPHasis-10: development of a health-related quality of life measure in pulmonary hypertension, Eur. Respir. J. 43 (4) (2014) 1106–1113, https://doi.org/10.1183/09031936.00127113.
[12] C.A.C. Prinsen, L.B. Mokkink, L.M. Bouter, J. Alonso, D.L. Patrick, H.C.W. De Vet, C.B. Terwee, COSMIN guideline for systematic reviews of patient-reported outcome measures, Qual. Life Res. (2018) accepted.
[13] C.B. Terwee, C.A.C. Prinsen, A. Chiarotto, M.J. Westerman, D.L. Patrick, J. Alonso, L.M. Bouter, H.C.W. de Vet, L.B. Mokkink, COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study, Qual. Life Res. 27 (5) (2018) 1159–1170. https://doi:10.1007/s11136-018-1829-0.
[14] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, E.A. Akl, S.E. Brennan, R. Chou, J. Glanville, J. M. Grimshaw, A. Hróbjartsson, M.M. Lalu, T. Li, E.W. Loder, E. Mayo-Wilson, S. McDonald, L.A. McGuinness, L.A. Stewart, J. Thomas, A.C. Tricco, V.A. Welch, P. Whiting, D. Moher, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews, BMJ 372 (2021) 71. https://doi:10.1136/bmj.n71.
[15] A. Aguirre-Camacho, J. Stepanous, L.M. Blanco-Donoso, B. Moreno-Jiménez, J. Wilburn, L. González-Saiz, S.P. McKenna, Adaptación y validación del cuestionario Cambridge Pulmonary Hypertension Outcome Review (CAMPHOR) para uso en España, Rev. Española Cardiol. 70 (6) (2017) 467–473, https://doi.org/10.1016/j.recesp.2016.11.006.
[16] K. Cima, J. Twiss, R. Speich, S.P. McKenna, E. Grünig, C.M. Kähler, N. Ehlken, U. Treder, S.R. Crawford, L.C. Huber, S. Ulrich, The German adaptation of the Cambridge pulmonary hypertension outcome review (CAMPHOR), Health Qual. Life Outcome 10 (2012) 110, https://doi.org/10.1186/1477-7525-10-110.
[17] R.A. Corrêa, M.C. Pereira, M.F. Bizzi, R.W.R. De Oliveira, C.F. Rezende, B.C.M. T. De Oliveira, A. Heaney, S.P. McKenna, A. Ribeiro-Oliveira, Adaptation and validation of the quality of life assessment of the Cambridge pulmonary hypertension outcome review (CAMPHOR) for Brazil, J. Patient-Report. Outcomes 4 (1) (2020) 43, https://doi.org/10.1186/s41687-020-00209-6.
[18] A. Reis, J. Twiss, M. Vicente, F. Gonçalves, L. Carvalho, J. Meireles, A. Melo, S. P. McKenna, L. Almeida, Portuguese validation of the Cambridge pulmonary hypertension outcome review (CAMPHOR) questionnaire, Health Qual. Life Outcome 14 (2016), https://doi.org/10.1186/s12955-016-0513-8.
[19] N. Selimovic, B. Rundqvist, E. Kjörk, J. Viriden, J. Twiss, S.P. McKenna, Adaptation and validation of the Cambridge pulmonary hypertension outcome review for Sweden, Scand. J. Publ. Health 40 (8) (2012) 777–783.
[20] D.M. Meads, S.P. McKenna, N. Doughty, C. Das, W. Gin-Sing, J. Langley, J. Pepke-Zaba, The responsiveness and validity of the CAMPHOR Utility Index, Eur. Respir. J. 32 (6) (2008) 1513–1519, https://doi.org/10.1183/09031936.00069708.
[21] D. Coffin, K. Duval, S. Martel, J. Granton, M.C. Lefebvre, D.M. Meads, J. Twiss, S. P. McKenna, Adaptation of the Cambridge pulmonary hypertension outcome review (CAMPHOR) into French-Canadian and English-Canadian, Can. Respir. J. J. Can. Thorac. Soc. 15 (2) (2008) 77–83, https://doi.org/10.1155/2008/767126.
[22] M. Wapenaar, J. Twiss, M. Wagenaar, P. Seijkens, L. van den Toorn, J. Stepanous, A. Heaney, A. van den Bosch, K.A. Boomars, Adaptation and validation of the Cambridge pulmonary hypertension outcome review (CAMPHOR) for The Netherlands, Neth. Heart J.: Monthly J. Netherlands Soci. Cardiol. Netherlands Heart Found. 24 (6) (2016) 417–424, https://doi.org/10.1007/s12471-016-0849-z.
[23] C. Villaquirán, S. Moreno, R. Dueñas, P. Acuña, J.R. Lutz, J. Wilburn, A. Heaney, Cross-cultural adaptation of the Cambridge pulmonary hypertension outcome review for use in patients with pulmonary hypertension in Colombia, J. Bras. Pneumol.: Publicacao Oficial Da Sociedade Brasileira De Pneumologia E Tisilogia 45 (6) (2019) e20180332, https://doi.org/10.1590/1806-3713/e20180332.
[24] K. Małaczyńska-Rajpold, A. Smukowska-Gorynia, A. Heaney, S.P. McKenna, M. Janus, A. Araszkiewicz, S. Jankiewicz, S. Slawek-Szmyt, I. Tomaszewska, T. Mularek-Kubzdela, The polish adaptation of the CAMbridge pulmonary hypertension outcome review (CAMPHOR), Cardiol. J. 27 (5) (2020) 608–615. https://doi:10.5603/CJ.a2018.0119.
[25] M. Gomberg-Maitland, T. Thenappan, K. Rizvi, S. Chandra, D.M. Meads, S. P. McKenna, United States validation of the Cambridge pulmonary hypertension outcome review (CAMPHOR), J. Heart Lung Transplant. 27 (1) (2008) 124–130, https://doi.org/10.1016/j.healun.2007.10.004.
[26] L. Ganderton, S. Jenkins, S.P. McKenna, K. Gain, R. Fowler, J. Twiss, E. Gabbay, Validation of the Cambridge pulmonary hypertension outcome review (CAMPHOR) for the Australian and New Zealand population, Respirology 16 (8) (2011) 1235–1240, https://doi.org/10.1111/j.1440-1843.2011.02030.x.
[27] P. Odevoglu, R. Demir, G. Okumus, M.S. Kucukoglu, G. Kuran Aslan, Validity and reliability of the Turkish version of the EMPHASIS-10 questionnaire in patients with pulmonary hypertension, J. Eval. Clin. Pract. 25 (5) (2019) 896–902, https://doi.org/10.1111/jep.13115.
[28] K. Sarzyńska, J. Polański, G. Kopeć, E. Mroczek, B. Jankowska-Polańska, Cultural adaptation and validation of the polish version of the pulmonary arterial hypertension-symptoms and impact (PAH-SYMPACT) questionnaire, Pol. Heart J./Kardiol. Pol. 79 (12) (2021) 1372–1374, https://doi.org/10.33963/KP.a2021.0149.
[29] K.M. Chin, M. Gomberg-Maitland, R.N. Channick, M.J. Cuttica, A. Fischer, R. P. Frantz, E. Hunsche, L. Kleinman, J.W. McConnell, V.V. McLaughlin, C.E. Miller, R.T. Zamanian, M.S. Zastrow, D.B. Badesch, Psychometric validation of the pulmonary arterial hypertension-symptoms and impact (PAH-SYMPACT) questionnaire: results of the SYMPHONY trial, Chest 154 (4) (2018) 848–861, https://doi.org/10.1016/j.chest.2018.04.027.
[30] E. Cho, Reliability and Omega Hierarchical in Multidimensional Data: A Comparison of Various Estimators, Psychological Methods, Advance online publication, 2022, https://doi.org/10.1037/met0000525.