ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

Patient-Reported Outcomes

Patient-Reported Outcome Measures of Quality of Life in People Affected by Diabetic Foot: A Psychometric Systematic Review

Àngel Romero-Collado, PhD, Elvira Hernández-Martínez-Esparza, PhD, Edurne Zabaleta-del-Olmo, PhD, Ana-María Urpí-Fernández, PhD, Rosalía Santesmases-Masana, PhD

Objectives: This psychometric systematic review aimed to identify the most suitable patient-reported outcome measures (PROMs) of quality of life (QoL) in people affected by diabetic foot.

Methods: We performed a literature search in MEDLINE (PubMed), CINAHL (EBSCOhost), and PsycINFO (EBSCOhost) databases from inception to February 1, 2022. We also searched gray literature databases. Eligible studies were full-text reports developing a QoL condition-specific PROM or assessing one or more of its measurement properties in people affected by diabetic foot. We assessed the methodological quality of included studies independently using the "Consensus-Based Standards for the Selection of Health Measurement Instruments Risk of Bias" checklist. The measurement properties were evaluated using specific criteria. We graded the quality of the evidence using a "Grading of Recommendations Assessment, Development and Evaluation" approach modified by Consensus-Based Standards for the Selection of Health Measurement Instruments.

Results: Forty-three reports (46 studies) providing information on the measurement properties of 10 different PROMs were included. We did not identify any instruments that could be recommended for use. We identified 2 PROMs that were not recommended for use and 8 that were potentially recommended but would require further investigation. Of these 8 PROMs, 4 had better evidence for content validity.

Conclusions: Available PROMs to measure QoL in people affected by diabetic foot have limited evidence for their measurement properties. There is no fully suitable PROM. Pending further evidence, 4 PROMs could potentially be recommended for use.

Keywords: diabetic foot, patient-reported outcome measures, psychometrics, quality of life, systematic review.

VALUE HEALTH. 2022; 25(9):1602–1618

Introduction

Diabetic foot is defined as the "infection, ulceration, or destruction of tissues of the foot of a person with currently or previously diagnosed diabetes mellitus, usually accompanied by neuropathy and peripheral arterHy disease in the lower extremity."[1] Its complications, which often result in amputation, are a leading cause of the global burden of disability; therefore, it constitutes a major public health problem[2,3] with significant adverse consequences on the healthcare system and health economics.[4] Moreover, people affected by diabetic foot have a poorer quality of life (QoL) than people affected by diabetes but without diabetic foot.[5] In addition, poor QoL is associated with a worse diabetic foot prognosis.[6] Therefore, QoL is a crucial subject in the area of diabetic foot.[7]

Patient-reported outcome measures (PROMs) are standardised measures to quantify the patient's perspective and help understand how the disease, the health system, and healthcare impact patients.[8] They also enable patient-centered care management and are helpful indicators for comparing different health services or providers from the point of view of quality improvement.[8,9] Measurement instruments that measure QoL are a class of PROMs.[9] They yield a global summary of wellbeing and can be generic or condition specific. Using generic QoL measurement instruments, such as the EQ-5D or the Medical Outcomes Study Short Form Short Form 36-Item (SF-36) (and related measures), helps assess individuals and compare groups with and without a health condition.[9,10] Nevertheless, they may underestimate changes in QoL in specific populations and may not capture specific problems related to a particular condition or disease.[9] In this sense, the condition-specific PROMs provide additional and complementary information on the person's QoL.[9,11]

Psychometric systematic reviews are valuable tools for selecting PROMs for research and clinical practice and identifying critical gaps in knowledge on their quality.[12] In this regard, the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) initiative has recently emerged to improve the selection of PROMs by developing tools to select the most suitable PROM for a given purpose.[13] Among these tools are the methodological guidelines for psychometric systematic reviews of PROMs.[14-16] Previous systematic reviews have summarized and analyzed the PROMs of QoL used in the spectrum of diabetes-related foot disease.[17-21] Nevertheless, Hogg et al[17] did not conduct their review following the COSMIN guidelines.

Table 1. COSMIN definitions of measurement properties.[14-16]

Measurement property	Definition
Content validity	The degree to which the content of a PROM is an adequate reflection of the construct to be measured
Criterion validity	The degree to which the scores of a PROM are an adequate reflection of a "gold standard"
Cross-cultural validity	The degree to which the performance of the items on a translated or culturally adapted PROM are an adequate reflection of the performance of the items of the original version of the PROM
Hypotheses testing for construct validity	The degree to which the scores of a PROM are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments, or differences between relevant groups) based on the assumption that the PROM validly measures the construct to be measured
Internal consistency	The degree of the interrelatedness among the items
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured
Reliability	The extent to which scores for patients who have not changed are the same for repeated measurement under several conditions: eg, over time (test-retest), by different persons on the same occasion (inter-rater), or by the same persons (ie, raters or responders) on different occasions (intra-rater)
Structural validity	The degree to which the scores of a PROM are an adequate reflection of the dimensionality of the construct to be measured
Responsiveness	The ability of a PROM to detect change over time in the construct to be measured

COSMIN indicates Consensus-Based Standards for the Selection of Health Measurement Instruments; PROM, patient-reported outcome measure.

Therefore, critical aspects of a review of this nature can be improved, especially the search strategy, the quality assessment of the included studies and the included PROMs, and the data synthesis.[22] Smith et al[20] conducted their review according to COSMIN guidelines; nevertheless, it focuses on assessing PROMs of QoL in people affected by diabetic neuropathy. Finally, the remaining 3 reviews focus on a broad range of PROMs,[18,19,21] not exclusively on those measuring QoL.

To address these issues, we conducted a psychometric systematic review to identify the most suitable PROMs of QoL in people affected by diabetic foot.

Methods

We conducted this review following the COSMIN initiative.[14-16] This review is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 statement[23] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension.[24] The review protocol was registered in the International Prospective Register of Systematic Reviews (PROSPERO) on August 6, 2018 (registration number: CRD42018096213); no changes have been made to the protocol.

Eligibility Criteria and Information Sources

We performed a literature search in MEDLINE (PubMed), CINAHL (EBSCOhost), and PsycINFO (EBSCOhost) databases. We also searched the gray literature on OpenGrey and Grey Literature Report databases. Reports were eligible if they met the following criteria: (1) the PROM had to be a disease-specific and not a generic measure of QoL, (2) the whole or a part of the study population was people affected by diabetic foot,[1] and (3) the aim of the study was the development of a PROM, the evaluation of one or more of its measurement properties, or the evaluation of its interpretability and feasibility. Studies that use only the PROM as an outcome measure (eg, clinical trials) or those used to validate another measurement instrument were excluded. We only included full-text articles because minimal information on a study is often found in abstracts. Measurement instruments development studies were also included even if these studies did not involve people affected by diabetic foot.

Search Strategy

A comprehensive literature search was initially performed to identify studies published from databases inception to February 1, 2022. We combined terms in controlled language and free text. Likewise, we added a highly sensitive filter developed by COSMIN for the MEDLINE (PubMed) search to identify studies on PROMs.[25] We imposed no language restrictions on any of the searches. The reproducible searches for all databases are available at https://zenodo.org/record/4501209#.YphTBS8lOgQ. We manually screened reference lists of included studies. In addition, we contacted with the authors of the included studies to retrieve the maximum possible information about the identified PROMs. We also complemented this initial search with an additional search for the particular PROMs identified using only the PROM name, and we browsed the identified PROMs' websites.

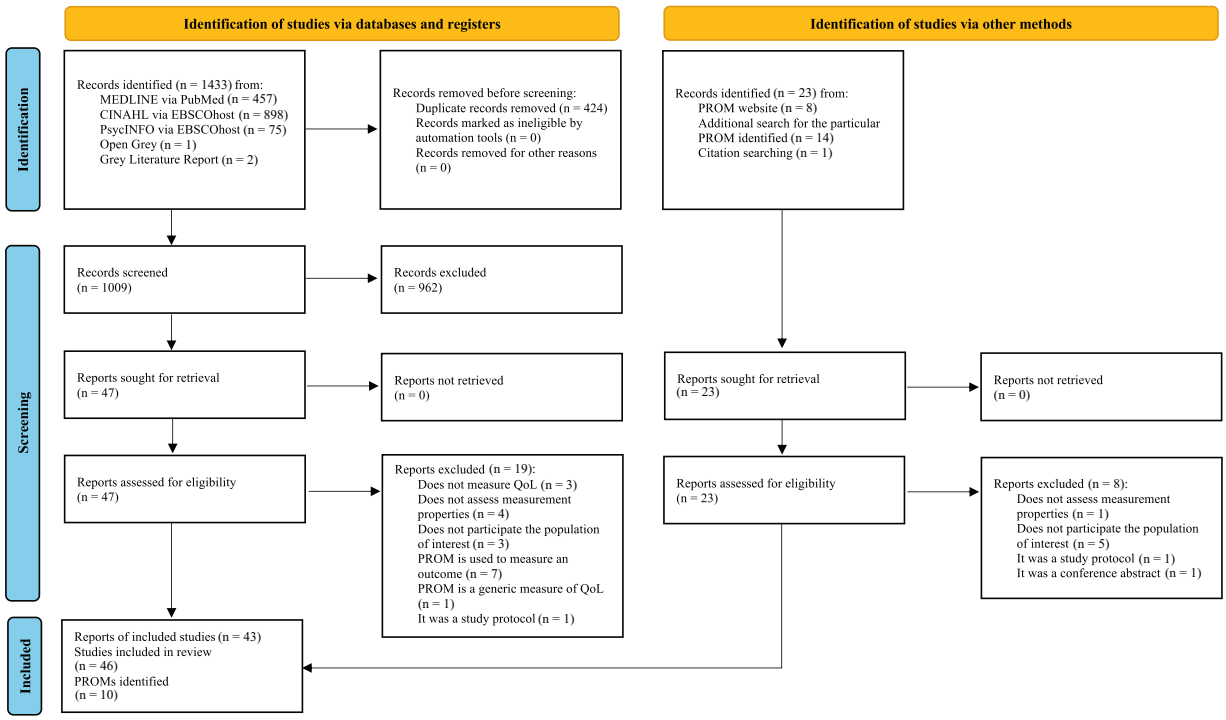
Selection Process

We imported the retrieved references into the Rayyan QCRI web application program.[26] This program facilitated collaboration among the reviewers during the study selection process. Two reviewers manually removed duplicates using Rayyan QCRI's duplicate identification strategy. These 2 reviewers independently assessed the titles and abstracts of the references retrieved, confronting them with the eligibility criteria. If a reference seemed relevant to at least one of the reviewers, the full text of the article was independently reviewed by these 2 reviewers. Conflicts over inclusion between these reviewers were discussed, and a third reviewer was consulted in case of not reaching a consensus.

Data Analysis

Included studies were grouped by PROM to identify the number of studies and PROMs separately. First, we assessed the methodological quality of every single study using the "COSMIN

Figure 1. PRISMA 2020 flow diagram.



PRISMA indicates Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

risk of bias" checklist.[14] We analyzed the following measurement properties: content validity, construct validity (structural validity, hypotheses testing, and cross-cultural validity), criterion validity, reliability (internal consistency, test-retest reliability, and measurement error), and responsiveness (see definitions of each measurement property in Table 1[14-16]). Concerning criterion validity, we agreed, based on the COSMIN guidelines,[14,15] that no gold standard exists for identified PROMs. The only exception is when a shortened instrument is compared with the original long version. In this case, we considered the original long version as the gold standard. When the studies compared the study PROM scores with a widely used instrument such as the SF-36, it was considered a construct validation.[27] Regarding hypothesis testing for construct validity and responsiveness, in accordance with COSMIN guidelines, it was not the *P* values but the direction and magnitude of the observed correlations that were taken into account. Thus, the review team agreed that correlations (changes in) of at least 0.50 between the PROM under study and a comparison instrument measuring the same construct would be interpreted as adequate and correlations from 0.30 to 0.50 between instruments measuring related but dissimilar constructs.[14-16] We determined which measurement properties were to be assessed in each study and rated the methodological quality of each of these studies as "very good," "adequate," "doubtful," or "inadequate." Second, evidence of every single study was rated against criteria for good measurement properties as "sufficient," "insufficient," and "indeterminate."[15] Third, the evidence was summarized per measurement property per PROM. Regarding hypothesis testing for construct validity, the results of all studies by PROM were taken together. We decided that the evidence was sufficient if 75% or more of the hypotheses were confirmed.[14-16] We pooled the test-retest reliability coefficients of the measurement instruments by the meta-essentials tool for correlational data version 1.5.[28] We used random-effects models based on the diversity of the population studied. The extent and impact of study heterogeneity were assessed by the tau^2 and the I^2 statistics, respectively. We summarized the rest of the measurement properties qualitatively. The overall result was rated against the criteria for good measurement properties. Finally, the quality of the evidence was graded by using a "Grading of Recommendations Assessment, Development and Evaluation" approach modified by COSMIN.[15] This approach uses 4 factors to determine the quality of the evidence: (1) risk of bias (methodological quality of the studies), (2) inconsistency of the results of the studies, (3) inaccuracy (small population sample size), and (4) indirect evidence (evidence from different populations not strictly related to the target population, in our case, eg, people affected by chronic wounds in general). We discussed a priori how ratings should be determined, and we piloted the ratings with a few articles from the review to take the scope of the review into account. All ratings were done by all the reviewers independently. We had regular consensus meetings to discuss rating issues and to ensure they were rated consistently. Information on the included studies and the identified PROMs was collected in the data extraction spreadsheets developed by COSMIN (available at https://cosmin.nl/wp-content/uploads/Scoring-form-COSMIN-boxes_april_final.xlsx). Interpretability and feasibility are not considered measurement properties, but they are essential aspects when selecting a PROM. Therefore, we also collected in specific tables these aspects and described them.

Based on the available evidence and its quality grade, we made recommendations for use per each PROM identified. These evidence-based recommendations were classified into 3 categories according to COSMIN guidelines[14-16]: (1) PROMs whose content validity had sufficient evidence and at least a low quality of evidence for a sufficient internal consistency of its scores, (2) PROMs with high-quality evidence for an insufficient

Table 2. Characteristics of the included PROMs.

PROM (reference to the first article)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/scoring	Original language	Available translations
AOFAS-DFQ[36]	People affected by diabetes and Charcot arthropathy	Self-administration	1 week	- 66 questions, many of which were multi-item in nature - It is organized into 5 sections: demographic data, general health (SF-36), foot problem and diabetes, health conditions, and foot care.	Items are not all on a common metric system; some items have scores such as 0/1 and others on and 5-point or 6-point Likert scale.	- Items are used to compute scores in 6 sections: (1) general health, (2) physicality, (3) emotion, (4) worry, (5) foot status, and (6) care. - Each section has many subscales: general health (6 subscales), physicality (6 subscales), emotion (7 subscales), worry (4 subscales), foot status (2 subscales), and care (4 subscales). - Each domain is transformed to create scores that range from 0 to 100. - Higher scores = better QoL	English (USA)	NIA
CWIS[51]	People affected by chronic wounds	Self-administration	1 week	- 26 items divided into 3 subscales: physical symptoms and everyday living (experienced and stressfulness of experience) (12 items), social life (experienced and stressfulness of experience) (7 items), and wellbeing (7 items). - Two added items that measure global QoL and satisfaction with QoL.	- 5-point Likert scale for 26 items - 11-point Likert scale for the 2 added items	- Items for each scale are summated and transformed onto a 0-100 scale. - High scores = better QoL	English (UK)	Chinese Dutch English (USA) French German Italian Portuguese Sinhalese Spanish (Mexico and Spain) Swedish Tao Welsh

continued on next page

Table 2. Continued

PROM (reference to the first article)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/scoring	Original language	Available translations
DFS[29]	People affected by DFU	Patient and caregiver self-administration versions	Different recall periods are available: - The last 4 weeks - Now - No recall period version	64 items: - 58 items are grouped into 11 scales: leisure (5 items), physical health (6 items), daily activities (6 items), emotions (17 items), noncompliance (2 items), family (5 items), friends (5 items), treatment (4 items); satisfaction (1 item), positive attitude (5 items), and financial (2 items). - 6 items addressing employment-related issues	- 5-point Likert scale ranging from 1 "not at all" or "none of the time" to 5 "a great deal" or "extremely" - Some items have an additional 0 "not applicable" option. - For the employment-related items, 2 additional response options are used: dichotomous (yes/no) and a 6-point Likert scale.	- Scores are based on the sum of all items associated with each subscale (raw item scores are reverse coded when necessary). - The scores per dimension range between 0 and 100. - Items addressing employment-related issues are not used to compute scores. - Higher scores = better QoL	English (UK)	Czech Danish Dutch English (USA) French Indonesian Italian Norwegian
DFS-SF[33]	People affected by DFU	Self-administration or interview-administration	Different recall periods are available - Now - The last 4 weeks	29 items grouped into 6 subscales: - Leisure (5 items) - Physical health (5 items) - Negative emotions (6 items) - Dependence/daily life (5 items) - Worried about ulcers/feet (4 items) - Bothered by ulcer care (4 items)	5-point Likert scale ranging from 1 "not at all" or "none of the time" to 5 "a great deal" or "all the time" or "extremely"	- Domain scores are based on the sum of all items associated with each subscale (raw item scores are reverse coded when necessary). - The scores per dimension are transformed on a scale from 0 to 100. - Higher scores = better QoL	- Danish - Dutch - English UK and USA - French - German - Italian	Arabic (Jordan) Bahasa Chinese (Hong Kong) French (Canada) Greek Korean Kannada (India) Polish Portuguese (Brazil) Spanish (Spain and USA) Turkish
HRQLQDFU[44]	People affected by DFU	Self-administration	Now	20 items that included 6 domains: physical health, physical symptoms, daily activity, emotional status, social status, and financial effect	Likert-type 4-point rating scale	- Scores are based on the sum of all items associated with each domain. - Higher score = better QoL	English (India)	NIA

continued on next page

Table 2. Continued

PROM (reference to the first article)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/scoring	Original language	Available translations
NeuroQoL[59]	People affected by diabetic peripheral neuropathy and foot ulcers	Self-administration and interview-administration	Last 4 weeks	35 items - 13 items assess specific somatic experiences in 3 domains. - 14 items assess specific functional, social, and emotional experiences in 3 domains. - A single item assesses the quality of life in each of the 6 domains. - Two final items in the scale assess overall satisfaction or QoL, one item requesting that the patient make a judgment specific to his or her experience with foot problems, and a final item asking for an overall judgment of QoL.	5-point Likert scale: never to all of the time	- For each of 27 specific items, patients are asked to judge the degree to which the somatic experience, restriction of activities, social function, and emotional states have been a bother or important to them. The bother/importance items were scored as 1 = none, 2 = some, and 3 = very. - Weighted scores are calculated by multiplying the scale score by the corresponding bother/importance score. - Higher score = worst QoL	English (USA and UK)	Arabic (Jordan) Portuguese (Brazil)
Norfolk QoL-DN[60]	People affected by diabetic neuropathy	Self-administration	- The last 4 weeks (39 items) - Now (8 items)	47 items divided into 5 subscales: physical (1) functioning/large fiber (15 items), (2) ADLs (5 items), small fiber (4 items), symptoms (8 items), and autonomic (3 items) - 8 items were excluded from the final factorial analysis - 4 items about the duration of symptoms, nature of symptoms, and medications (these items are not included in the scoring of any of the scales)	- Items 1-7, scores for each symptom equal the number of body sites for which a symptom is reported. - Items 12-15, "yes" or "no" Items 16-46, a 5-point Likert-type scale - For item 47, 3 possible responses: 1-2, 3-4, or 5 or more medications	- The subscales are calculated without weighting and reported as the sum of the questionnaire items related to each subscale. - Items 8-11 are for clinical purposes and not included in the scoring - Higher score = worst QoL	English (USA)	German

continued on next page

Table 2. Continued

PROM (reference to the first article)	Target population	Mode of administration	Recall period	(Sub)scale(s) (number of items)	Response options	Range of scores/scoring	Original language	Available translations
Quality of life instrument[54]	People affected by chronic wounds	Interview-administration	Now	20 items included 6 domains: physical activities, feelings, household duties, leisure time activities, social relations, and general activities.	5-point rating scale	- Scores above the mean (17 points) were classified as satisfactory and those below unsatisfactory. - Higher scores = better QoL	English (India)	NIA
Wound-QoL[34]	People affected by chronic wounds	Self-administration and interview-administration	1 week	17 items into 3 subscales: "body" (5 items); "psyche" (5 items), and "everyday life" (6 items) - One item assesses the financial burden (does not belong to either of the subscales).	Each item has 5 possible answers: no at all, a little, moderately, quite a lot. and very much (scores ranged from 0 to 4).	- A global score on overall disease-specific quality of life is computed by averaging all items. It can only be computed if at least 75% of the items have been answered. - Subscales scores can be calculated by averaging the individual items. A subscale can only be computed if no >1 item of the subscale is missing. - Higher scores = worst QoL	German (Germany and Austria)	Arabic (Israel) Chinese (China) Czech Danish Dutch (The Netherlands) English (Canada, UK, and USA) German (Switzerland) Hebrew French Italian Latvian Lithuanian Polish Portuguese (Brazil and Portugal) Russian Slovakian Spanish (Central America and Spain) Swedish Turkish
Wound-QoL revised version[70]	People affected by chronic wounds	Self-administration and interview-administration	1-week	14 items into 3 subscales: "body" (4 items); "psyche" (4 items), and "everyday life" 5 items) - One item assesses the burden (does not belong to either of the subscales).	Each item has 5 possible answers: no at all, a little, moderately, quite a lot, and very much (scores ranged from 0 to 4).	Higher scores = worst QoL	German	Dutch (The Netherlands) English German Hebrew Spanish (Spain) Swedish

ADL indicates activity of daily living; AOFAS-DFQ, American Orthopaedic Foot and Ankle Society Diabetic Foot Questionnaire; CWIS, Cardiff Wound Impact Schedule/Scale; DFS-SF, Diabetic Foot Ulcer Scale-Short Form; DFU, diabetic foot ulcers; HRQLQDFU, Health-Related Quality Of Life Questionnaire in Diabetic Foot; NeuroQoL, Neuropathy- and Foot Ulcer-Specific Quality of Life; NIA, no information available; Norfolk QoL-DN, Norfolk Quality of Life-Diabetic Neuropathy; PROM, patient-reported outcome measure; QoL, quality of life; SF-36, Short Form 36-Item; UK, United Kingdom; USA, United States of America.

measurement property, and (3) PROMs not classified either as A or as B. PROMs classified as A were recommended for use, and those classified as B were unrecommended. PROMs classified as C could be recommended, but a more significant number of studies were needed to assess their quality. Because we did not find any instruments categorized for category A, we based recommendations for use on those with the best evidence for content validity and the best issues of interpretability and feasibility among those categorized in category C.

Results

The literature search and study selection process are detailed in Figure 1. Forty-six studies (43 reports) providing information on the measurement properties of 10 different PROMs were included.[29-71] These identified measurement instruments were the American Orthopaedic Foot and Ankle Society Diabetic Foot Questionnaire (AOFAS-DFQ),[36] the Cardiff Wound Impact Schedule/Scale (CWIS),[30,37,40,41,43,47,51,56,58,64] the Diabetic Foot Ulcer Scale (DFS),[29,52,53,62] the Diabetic Foot Ulcer Scale-Short Form (DFS-SF),[33,42,45,46,48-50,57,65] the Health-Related Quality of Life Questionnaire in Diabetic Foot (HRQLQDFU),[44] the Neuropathy- and Foot Ulcer-Specific Quality of Life (NeuroQoL),[59,63,66] the Norfolk Quality of Life-Diabetic Neuropathy (Norfolk QoL-DN),[60,61] the Quality of Life Instrument,[54] the Wound-QoL,[31,32,34,35,38,39,55,67-71] and the Wound-QoL revised version.[70] The characteristics of these PROMs and the included studies are presented in Table 2[29,33,34,36,44,51,54,59,60,70] and Appendix Table 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1737. In addition, the reports that appeared to meet the eligibility criteria but were excluded and the reasons for exclusion are listed in Appendix Text 1 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1737.

The summary of the assessments of PROM development and the methodological quality of the included studies is presented in Table 3 and Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1737. Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1737 provides the ratings of each study against the criteria for good measurement properties. Finally, the quality of the evidence is summarized in Table 4.

Content Validity

In 5 of the 10 identified PROMs,[34,36,44,54,70] the development methodological quality was rated as inadequate because the design was not performed on a sample representing the target population. Regarding Wound-QoL design,[34] although members of the target population participated in it, the selection of items was based on their quantitative properties and on expert judgment of their relevance and comprehensibility. Moreover, only the HRQLQDFU[44] was pilot tested to evaluate comprehensibility and comprehensiveness in its development stage. Nevertheless, it appears that the items were not tested in their final form or were not retested after substantial adjustments. For the remaining 5 PROMs,[29,33,51,59,60] the development methodological quality was rated as doubtful, given that some methodological issues were not sufficiently clarified, such as the type of method used in the qualitative studies, the interviewers' skills, the use of an interview guide, and details related to the recording and transcription of the interviews conducted.

We identified 25 studies that analyzed content validity issues (see Table 3[29-71]). These studies assessed the comprehensibility of PROMs original versions translated into other languages. The quality of 20 of these studies was rated as doubtful mainly because crucial aspects of the methodology used were not sufficiently clarified. We rated the quality of 5 studies[31,39,53,57,64] as inadequate because either the method used was inappropriate or respondents were not asked about the comprehensibility of all items and response options.

In summary (see Table 4), we graded the content validity sufficient with moderate evidence for 2 PROMs (DFS and DFS-SF), low evidence for 4 (CWIS, NeuroQoL, Norfolk-DN, and Wound-QoL), and very low evidence for 4 (AOFAS-DFQ, HRQLQDFU, QoL Instrument, and Wound-QoL revised version).

Construct Validity

We did not identify any studies that assessed the structural validity of the AOFAS-DFQ, the HRQLQDFU, and the QoL Instrument. We rated the structural validity of 3 instruments (DFS, NeuroQoL, and Norfolk QoL-DN) as indeterminate because the identified studies did not provide sufficient information to assess the criteria for good measurement properties. The remaining 4 PROMs showed insufficient evidence for structural validity. The quality of this evidence was high for DFS-SF, moderate for the Wound-QoL and the Wound-QoL revised version, and low for the CWIS (see Table 4). Hypotheses testing for construct validity was assessed in 7 PROMs (CWIS, DFS, DFS-SF, NeuroQoL, Norfolk QoL-DN, QoL Instrument, and Wound-QoL). The results of the convergent and discriminative validity tests are described in Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1737. Regarding this measurement property, only one of these PROMs, the NeuroQoL, showed sufficient evidence with a moderate quality (see Table 4). Cross-cultural validity was assessed in 2 PROMs (the Wound-QoL and the Wound-QoL revised version) and was considered sufficient for both, although in both cases with a very low quality of evidence.

Criterion Validity

We identified only a single study[70] that assessed the criterion validity of a shortened PROM (the Wound-QoL revised version) compared with the original long version (the Wound-QoL). As shown in Table 4, the results indicate that the evidence is sufficient for this measurement property with moderate quality of evidence.

Reliability

For all identified PROMs, except the QoL Instrument, the internal consistency of their scores was assessed. Nevertheless, we could not interpret the criterion for good measurement properties because none of these PROMs showed at least low evidence for sufficient structural validity (see Table 4). Test-retest reliability of scores was assessed in 6 identified PROMs (AOFAS-DFQ, CWIS, DFS, DFS-SF, NeuroQoL, and Wound-QoL). We were only able to pool the results of the CWIS and Wound-QoL studies. Regarding DFS-SF, although this measurement property was assessed in 4 studies, only one study provided sufficient information to pool the data (see Appendix Table 3 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1737). For the remaining 3 PROMs, we found only one study for the AOFAS-DFQ and the NeuroQoL, and for the DFS, not enough information was provided to assess the criteria for good measurement properties. Only the CWIS, the NeuroQoL, and the Wound-QoL showed sufficient evidence for this measurement property, although of very low quality for the first 2 and of low quality for the third (see Table 4). Measurement error was only assessed in 2 identified PROMs. In the case of AOFAS-DFQ, we could not interpret it due to the lack of information about minimal important change. Nevertheless, the

Table 3. Quality of the PROM development and the studies on measurement properties.

PROM/studies	Instrument development	Content validity					Structural validity
		Asking patients			Asking experts		
		Relevance	Comprehensiveness	Comprehensibility	Relevance	Comprehensiveness	
AOFAS-DFQ							
Dhawan et al[36]	I						
CWIS							
Price and Harding[51]	D						I
Acquadro et al[30]				D			
Goodridge et al[40]							
Jaksa and Mahoney[43]							
Fagerdahl et al[37]				D			
Huang et al[41]				D			D
Sriyani et al[56]				D			
Lozano-Platonoff et al[47]				D			
van Doorn et al[58]				D			
Granado-Casas et al[64]				I			A
DFS							
Abetz et al[29]	D						A
Ribu et al[52]							
Vymětalová and Zeleníková[62]				D			
Sari et al[53]				I			
DFS-SF							
Bann et al[33] (Study 1)[33]	D						A
Bann et al[33] (Study 2)[33] Bann et al[33] (Study 3)[33]							V
Hui et al[42]				D			
Kontodimopoulos et al[45]				D			
Macioch et al[48]				D			
Lee[46]				D			V
Martinez-Gonzalez et al[49]				D			A
Oliveira Kaizer et al[50]				D			V
Toygar et al[57]				I			V
Putri et al[65]							
HRQLQDFU							
Kateel et al[44]	I						
NeuroQoL							
Vileikyte et al[59]	D						A
Xavier et al[63]				D			
Ababneh et al[66]				D			
Norfolk QoL-DN							
Vinik et al[60]	D						A
Vinik et al[61]				D			A
Quality of life instrument							
Shukla et al[54]	I						
Wound-QoL							
Blome et al[34]	I						A
Augustin et al[32]							
Deufert and Graml[35]							
Sommer et al[55]							
Fagerdahl and Bergström[38]				D			
Gamus et al[39]				I			I
Amesz et al[31]				I			
Sommer et al[67]				D			V
Conde Montero et al[68]				D			A
Knudsen et al[69]			D	D			
Stülpnagel et al[70] (Study 1)[70]							V
Topp et al[71]							
Wound-QoL revised version							
Stülpnagel et al[70] (Study 2)[70]	I						V

Note. Empty cells indicate that study (of part of it) was not performed
A indicates adequate; AOFAS-DFQ, American Orthopaedic Foot and Ankle Society Diabetic Foot Questionnaire; CWIS, Cardiff wound impact schedule/scale; D, doubtful; DFS, diabetic foot ulcer scale; DFS-SF, diabetic foot ulcer scale-short form; HRQLQDFU, health-related quality of life questionnaire in diabetic foot; I, inadequate; NeuroQoL, neuropathy- and foot ulcer-specific quality of life; Norfolk QoL-DN, Norfolk quality of life-diabetic neuropathy; PROM, patient-reported outcomes measure; V, very good.

Table 3. Continued

Internal consistency	Cross-cultural validity	Reliability	Measurement error	Criterion validity	Construct validity		Responsiveness			
					Convergent validity	Known groups validity	Comparison with gold standard	Comparison with other instruments	Comparison between subgroups	Comparison before and after intervention
V		I	I					I		
V		I			V	D				
V										
		I			V	D				
V		I			V	D				I
V					V					
V		D			V	D				
V					V	D				
V		I			V	D				
V		D			V				D	
V		I				V				D
						V				
					A	D				
V V		I I			V V					D D D
V					V	D				
V					V	D				
V					V					
V					V	D				
V		A			V				D	
V					V					
V		D								
V										
I										
V					V	D				
V					V	D				
V		D								
V		D			D	A				
					V	D				
						D				
V					V			V		
I					V			V		
V		A								
V		I			V			I		
V					D					
V					V	V		V		
V					D					
V					D			D		
V		D	D		V		V			
V	D									
V							V			
V	D			V						

Wound-QoL showed sufficient evidence, albeit very low quality (see Table 4).

Responsiveness

This measurement property was assessed in 5 PROMs. Only the DFS-SF showed sufficient evidence with low quality for this measurement property (see Table 4). The CWIS and the Wound-QoL showed insufficient evidence given that <75% of the hypotheses were confirmed.

Categorization of PROMs According to Suitability Recommendations

We did not identify any instruments categorized as A. The DFS and the DFS-SF were categorized as B. The remaining 8 PROMs were categorized as C, 4 of which (CWIS, NeuroQoL, Norfolk QoL-DN, and Wound-QoL) had better evidence for content validity than the other 4 (see Table 4). Of these 4 PROMs, the Wound-QoL was the most feasible (the lowest number of items and the shortest completion time) (see Appendix Table 4 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2022.04.1737).

Discussion

In this psychometric systematic review, we identified 10 available QoL PROMs for people affected by diabetic foot. The DFS and the DFS-SF were the PROMs with a higher grade of evidence for content validity. None of the PROMs identified provided sufficient evidence for structural validity, so we could not interpret the available evidence on internal consistency of their scores. Only the NeuroQoL provided sufficient evidence for the hypotheses testing for construct validity. Cross-cultural validity was examined in the Wound-QoL and the Wound-QoL revised version. The CWIS, the NeuroQoL, and the Wound-QoL showed sufficient evidence for test-retest reliability of their scores. Measurement error of the scores was assessed in 2 PROMs, but only the Wound-QoL could be interpreted. The DFS-SF showed sufficient evidence for the responsiveness of its scores. Therefore, available PROMs had limited evidence for their measurement properties.

In comparison with an earlier review,[17] we found a larger number of studies of some PROMs (CWIS and DFS-SF) and identified new ones such as the Wound-QoL and the Wound-QoL revised version. According to the COSMIN guidelines, only PROMs with sufficient content validity and at least a low quality of evidence for sufficient internal consistency can be considered the most suitable PROMs for use.[14-16] Structural validity is the starting point for determining the quality of evidence for internal consistency.[14-16] Although all of the identified PROMs showed sufficient content validity, none of them provided sufficient structural validity. Two PROMs showed high-quality evidence for insufficient measurement properties, the DFS for hypothesis testing and the DFS-SF for structural validity and hypothesis testing. Consequently, their use cannot be recommended until more evidence is developed. The CWIS, the NeuroQoL, the Norfolk QoL-DN, and the Wound-QoL could be potentially recommended for use until further evidence is provided. Of these 4 instruments, the Wound-QoL is easier to apply and would therefore be more clinical utility than the others.

Comprehensive database searches, the use of a rigorous and up-to-date psychometric review methodology,[14-16] and the systematic assessment of QoL condition-specific PROMs in a large number of studies are the key strengths of this review. Nevertheless, methodological quality scoring, interpretation of results, and evidence grading remain subjective processes. Therefore, the process of independent review and consensus ratings helped to resolve discrepancies and reduce variability in interpretation. In addition, psychometric reviews are quite complex because they involve multiple reviews, one review for each measurement property. Accordingly, the review team included reviewers with knowledge of the construct of interest and experience with the target population and with the field of psychometrics and qualitative research. Finally, public and patient involvement in research, both in the field of wound care and in general, is of increasing interest to researchers and clinicians.[72,73] Hence, one possible approach to consider would be to include people affected by diabetic foot as research partners in future psychometric reviews for rating the content of the identified PROMs.

Measuring the QoL of people affected by diabetic foot requires robust PROMs that allow researchers and clinicians to come as close as possible to the reality experienced by the person. The COSMIN initiative provides useful tools[14-16,74] for the development of studies aimed at providing more evidence on available PROMs and for the design and evaluation of new PROMs. Content validity is the most important measurement property of a PROM.[16] Input from the target population of a PROM is essential to assess the 3 aspects of this measurement property: relevance, comprehensibility, and comprehensiveness.[16] To assess these aspects, it is necessary to use both quantitative and qualitative methodologies. We found this to be the weakest point of the development of identified PROMs, especially the use of rigorous qualitative methodologies for identifying and selecting instrument items. In contrast, a PROM's content validity may be different when used in different populations or different contexts. Thus, each new use may require new supporting evidence. Nevertheless, the methodological quality of the content validity studies was mostly doubtful and in some cases inadequate. Therefore, This is clearly a research area that needs to be improved. Structural validity and cross-cultural validity are aspects of construct validity. Information on model fit indices was not available for 6 identified PROMs. We found only 2 studies that analyzed the cross-cultural validity of the identified PROMs despite the numerous adaptations and translations that have been carried out. This type of validity is essential to determine the equivalence of scores between the original population and the new target population.[15] Most studies used *P* values to test hypotheses for construct validity or responsiveness rather than to assess whether the magnitude of correlations or observed differences were similar or greater than expected. The expected differences or changes need to be defined in advance, and the interpretation of the tests needs to be based on their observed magnitude and direction and not only on statistical criteria.[14-16] Test-retest reliability is a prerequisite for determining the potential use of a PROM to assess trends or changes over time. We identified 17 studies that evaluated this property. Nevertheless, only 2 of these studies were rated with adequate methodological quality. The reasons for this weakness in the quality of the studies were no evidence provided that respondents were stable, use of inappropriate time interval, and different test conditions in both assessments. Therefore, future studies need to consider all these limitations to provide better quality evidence.

Table 4. Summary of findings for each patient-reported outcome measure according to the recommendation for use.

PROM	Measurement property	Summary results	Overall rating	Quality of evidence
Category A: PROMs whose content validity had sufficient evidence, and at least a low quality of evidence for a sufficient internal consistency of its scores *(recommended for use)*				
No PROMs categorized as "A" were found in this review				
Category B: PROMs with high-quality evidence for an insufficient measurement property *(unrecommended for use)*				
DFS	Content validity	NA	Sufficient	Moderate: PROM development study doubtful quality, and there is at least one content validity study of doubtful quality.
	Structural validity	Not all information for "+" reported	Indeterminate	NA
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met	Indeterminate	NA
	Reliability	ICC or weighted Kappa not reported	Indeterminate	NA
	Hypothesis testing	29 out of 39 (74%) hypotheses confirmed	Insufficient	High: there are multiple studies of very good quality available.
	Responsiveness	No hypothesis defined	Indeterminate	No hypothesis defined
DFS-SF	Content validity	NA	Sufficient	Moderate: PROM development study doubtful quality, and there is at least one content validity study of doubtful quality
	Structural validity	6 factors: CFI ranged from 0.84 to 0.94 and RMSEA ranged from 0.06 to 0.10	Insufficient	High: there are multiple studies of very good quality available.
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met (?)	Indeterminate	NA
	Reliability	ICC range 0.51-0.92; inconsistent	Insufficient	Low: there is one study of adequate validity, and results are inconsistent.
	Hypothesis testing	17 of 27 (63%) hypotheses confirmed	Insufficient	High: there are multiple studies of very good quality available.
	Responsiveness	1 of 1 (100%) hypothesis confirmed	Sufficient	Low: there is one study of doubtful quality available.
Category C: PROMs categorized not in A or B *(recommended for use until further evidence is provided)*				
AOFAS-DFQ	Content validity	NA	Sufficient	Very low: PROM development study inadequate quality, no content validity studies, and only part of the study population consisted of patients with the disease of interest.
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met	Indeterminate	NA
	Reliability	ICC range from 0.35 to 0.85; total sample size = 57	Insufficient	Very low: there is only one study of inadequate quality available, only part of the study population consisted of patients with the disease of interest, and the sample size was lower than 100.
	Measurement error	MIC not defined	Indeterminate	NA
	Responsiveness	No hypothesis defined	Indeterminate	NA

continued on next page

Table 4. Continued

PROM	Measurement property	Summary results	Overall rating	Quality of evidence
CWIS	Content validity	NA	Sufficient	Low: PROM development study doubtful quality, there is at least one content validity study of doubtful quality, and only part of the study population consisted of patients with the disease of interest.
	Structural validity	3 factors: CFI = 0.69 and RMSEA = 0.09	Insufficient	Low: there is only one study of adequate quality, and only part of the study population consisted of patients with the disease of interest.
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met (?)	Indeterminate	NA
	Reliability	Everyday living subscale, ICC = 0.81 (95% CI 0.66-0.90), I^2 = 54.3%; social life subscale, ICC = 0.74 (95% CI 0.62-0.83), I^2 = 33.8%; and wellbeing subscale, ICC = 0.69 (95% CI 0.48-0.82), I^2 = 54.4%; total sample size = 283; inconsistent	Sufficient	Very low: there are 2 studies of doubtful quality available, results are moderately inconsistent, and only part of the study population consisted of patients with the disease of interest.
	Hypothesis testing	29 of 67 (43%) hypotheses confirmed	Insufficient	Moderate: there are multiple studies of very good quality available, and only part of the study population consisted of patients with the disease of interest.
	Responsiveness	3 of 5 (60%) hypothesis confirmed	Insufficient	Low: there is only one study of doubtful quality available.
HRQLQDFU	Content validity	NA	Sufficient	Very low: PROM development study inadequate quality and no content validity studies
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met	Indeterminate	NA
NeuroQoL	Content validity	NA	Sufficient	Low: PROM development study doubtful quality, there is at least one content validity study of doubtful quality, and only part of the study population consisted of patients with the disease of interest.
	Structural validity	Not all information for "+" reported	Indeterminate	NA
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met	Indeterminate	NA
	Reliability	ICC range 0.76-0.90; total sample size = 15	Sufficient	Very low: there is only one study of doubtful quality available, and sample size was lower than 50.
	Hypothesis testing	7 of 9 (78%) hypotheses confirmed	Sufficient	Moderate: there is one study of very good quality available, and only part of the study population consisted of patients with the disease of interest.

continued on next page

Table 4. Continued

PROM	Measurement property	Summary results	Overall rating	Quality of evidence
Norfolk QoL-DN	Content validity	NA	Sufficient	Low: PROM development study doubtful quality, there is at least one content validity study of doubtful quality, and only part of the study population consisted of patients with the disease of interest.
	Structural validity	Not all information for "+" reported	Indeterminate	NA
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met	Indeterminate	NA
	Hypothesis testing	9 of 14 (64%) hypotheses confirmed	Insufficient	Moderate: there is one study of very good quality available, and only part of the study population consisted of patients with the disease of interest.
Quality of life instrument	Content validity	NA	Sufficient	Very low: PROM development study inadequate quality, no content validity studies, and only part of the study population consisted of patients with the disease of interest.
	Hypothesis testing	4 of 6 (67%) hypotheses confirmed	Insufficient	Very low: there is only one study of doubtful quality available, and only part of the study population consisted of patients with the disease of interest.
Wound-QoL	Content validity	NA	Sufficient	Low: PROM development study inadequate quality, there is at least one content validity study of doubtful quality, and only part of the study population consisted of patients with the disease of interest.
	Structural validity	3 factors: CFI ranged from 0.90 to 0.91 and RMSEA = 0.09	Insufficient	Moderate: there are 2 studies of very good quality available, and only part of the study population consisted of patients with the disease of interest.
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met	Indeterminate	NA
	Cross-cultural validity	mCFA metric invariance across different countries: CFI = 0.01 and RMSEA = 0.01	Sufficient	Very low: there is only one study of doubtful quality available, and only part of the study population consisted of patients with the disease of interest.
	Reliability	Total score, ICC = 0.87 (95% CI 0.84-0.90), I^2 = 0.0%; body subscale, ICC = 0.80 (95% CI 0.63-0.89), I^2 = 11.7%; psyche subscale, ICC = 0.81 (95% CI 0.73-0.87), I^2 = 0.0%; and everyday life subscale, ICC = 0.82 (95% CI 0.80-0.84), I^2 = 0.0%; total sample size = 185; consistent	Sufficient	Low: there is only one study of adequate quality available, and only part of the study population consisted of patients with the disease of interest.

continued on next page

Table 4. Continued

PROM	Measurement property	Summary results	Overall rating	Quality of evidence
	Measurement error	SDC < MIC in all scores except in psyche subscale	Sufficient	Very low: there is only one study of doubtful quality available, and only part of the study population consisted of patients with the disease of interest.
	Hypothesis testing	39 of 72 (54%) hypotheses confirmed	Insufficient	Moderate: there are multiple studies of very good quality available, and only part of the study population consisted of patients with the disease of interest.
	Responsiveness	12 of 25 (48%) hypotheses confirmed	Insufficient	Moderate: there are multiple studies of very good quality available, and only part of the study population consisted of patients with the disease of interest.
Wound-QoL revised version	Content validity	NA	Sufficient	Very low: PROM development study inadequate quality, no content validity studies and only part of the study population consisted of patients with the disease of interest.
	Structural validity	3 factors: CFI = 0.94 and RMSEA = 0.08	Insufficient	Moderate: there is one study of very good quality available, and only part of the study population consisted of patients with the disease of interest.
	Internal consistency	Criteria for "at least low evidence for sufficient structural validity" not met	Indeterminate	NA
	Cross-cultural validity	mCFA metric invariance across different countries, continent origin, age, sex, and wound type. All analyses showed CFI < 0.02 and RMSEA < 0.015	Sufficient	Very low: there is only one study of doubtful quality available, and only part of the study population consisted of patients with the disease of interest.
	Criterion validity	4 of 4 (100%) hypotheses confirmed	Sufficient	Moderate: there is one study of very good quality available, and only part of the study population consisted of patients with the disease of interest.

AOFAS-DFQ indicates American Orthopaedic Foot and Ankle Society Diabetic Foot Questionnaire; CFI, comparative fit index; CI, confidence interval; CWIS, Cardiff Wound Impact Schedule/Scale; DFS, diabetic foot ulcer scale; DFS-SF, Diabetic Foot Ulcer Scale-Short Form; HRQLQDFU, Health-Related Quality of Life Questionnaire in Diabetic Foot; ICC, intraclass correlation coefficient; mCFA, multilevel confirmatory factor analysis; MIC, minimal important change; NA, not applicable; NeuroQoL, Neuropathy- and Foot Ulcer-Specific Quality of Life; Norfolk QoL-DN, Norfolk Quality of Life-Diabetic Neuropathy; PROM, patient-reported outcomes measure; RMSEA, root mean square error of approximation; SDC, smallest detectable change.

Conclusions

This review found no fully suitable condition-specific PROM to measure QoL in people affected by diabetic foot. Pending further evidence, 4 PROMs could be provisionally recommended for use. Of these PROMs, Wound-QoL is the most feasible. These findings suggest that, in general, there are important evidence gaps for the measurement properties of QoL PROMs in people affected by diabetic foot. Clinicians, researchers, healthcare decision makers, and policy makers should be aware of these limitations before using these PROMs for decision making. The present review establishes a framework for future assessments of available PROMs or for the design of new ones.

Supplemental Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2022.04.1737.

Article and Author Information

Accepted for Publication: April 19, 2022

Published Online: June 1, 2022

doi: https://doi.org/10.1016/j.jval.2022.04.1737

Author Affiliations: Nursing Department, Faculty of Nursing, Universitat de Girona, Girona, Spain (Romero-Collado, Zabaleta-del-Olmo); Sant Pau Nursing School University, Barcelona, Spain (Hernández-Martínez-Esparza, Santesmases-Masana); Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIAPJGol), Barcelona, Spain (Zabaleta-del-Olmo); Gerència Territorial de Barcelona, Institut Català de la Salut, Barcelona, Spain (Zabaleta-del-Olmo, Urpí-Fernández).

Correspondence: Edurne Zabaleta-del-Olmo, PhD, Gran Via Corts Catalanes, 587 àtic, 08007 Barcelona, Spain. Email: ezabaleta@idiapjgol.org

Author Contributions: *Concept and design:* Romero-Collado, Zabaleta-del-Olmo
Acquisition of data: Romero-Collado, Hernández-Martínez-Esparza, Zabaleta-del-Olmo, Urpí-Fernández, Santesmases-Masana
Analysis and interpretation of data: Romero-Collado, Hernández-Martínez-Esparza, Zabaleta-del-Olmo, Urpí-Fernández, Santesmases-Masana
Drafting of the manuscript: Zabaleta-del-Olmo
Critical revision of the paper for important intellectual content: Romero-Collado, Hernández-Martínez-Esparza, Zabaleta-del-Olmo, Urpí-Fernández, Santesmases-Masana
Statistical analysis: Zabaleta-del-Olmo
Administrative, technical, or logistic support: Romero-Collado, Hernández-Martínez-Esparza, Zabaleta-del-Olmo, Urpí-Fernández, Santesmases-Masana
Supervision: Romero-Collado, Hernández-Martínez-Esparza, Zabaleta-del-Olmo, Urpí-Fernández, Santesmases-Masana

Conflict of Interest Disclosures: The authors reported no conflicts of interest.

Funding/Support: The authors received no financial support for this research.

Acknowledgment: The authors appreciate the review of the English text by P. Bialoskorski, MA. Open Access funding provided thanks to the CRUE-CSIC agreement with Elsevier.

REFERENCES

1. van Netten JJ, Bus SA, Apelqvist J, et al. Definitions and criteria for diabetic foot disease. *Diabetes Metab Res Rev*. 2020;36(suppl 1):e3268.
2. Lazzarini PA, Pacella RE, Armstrong DG, van Netten JJ. Diabetes-related lower-extremity complications are a leading cause of the global burden of disability. *Diabet Med*. 2018;35(9):1297–1299.
3. Edmonds M, Manu C, Vas P. The current burden of diabetic foot disease. *J Clin Orthop Trauma*. 2021;17:88–93.
4. Raghav A, Khan ZA, Labala RK, Ahmad J, Noor S, Mishra BK. Financial burden of diabetic foot ulcers to world: a progressive topic to discuss always. *Ther Adv Endocrinol Metab*. 2018;9(1):29–31.
5. Khunkaew S, Fernandez R, Sim J. Health-related quality of life among adults living with diabetic foot ulcers: a meta-analysis. *Qual Life Res*. 2019;28(6):1413–1427.
6. Ribu L, Birkeland K, Hanestad BR, Moum T, Rustoen T. A longitudinal study of patients with diabetes and foot ulcers and their health-related quality of life: wound healing and quality-of-life changes. *J Diabetes Complications*. 2008;22(6):400–407.
7. Jeffcoate WJ, Bus SA, Game FL, et al. Reporting standards of studies and papers on the prevention and management of foot ulcers in diabetes: required details and markers of good quality. *Lancet Diabetes Endocrinol*. 2016;4(9):781–788.
8. Benson T. Measure what we want: a taxonomy of short generic person-reported outcome and experience measures (PROMs and PREMs). *BMJ Open Qual*. 2020;9(1):e000789.
9. Cella D, Hahn E, Jensen S, et al. *Patient-Reported Outcomes in Performance Measurement*. Research Triangle Park, NC: RTI Press; 2015.
10. Cella D, Nowinski CJ. Measuring quality of life in chronic illness: the functional assessment of chronic illness therapy measurement system. *Arch Phys Med Rehabil*. 2002;83(12suppl 2):S10–S17.
11. Owolabi MO. Which is more valid for stroke patients: generic or stroke-specific quality of life measures? *Neuroepidemiology*. 2010;34(1):8–12.
12. Munn Z, Stern C, Aromataris E, Lockwood C, Jordan Z. What kind of systematic review should I conduct? A proposed typology and guidance for systematic reviewers in the medical and health sciences. *BMC Med Res Methodol*. 2018;18(1):5.
13. COnsensus-based Standards for the selection of health Measurement INstruments. COSMIN. https://www.cosmin.nl. Accessed June 3, 2021.
14. Mokkink LB, de Vet HCW, Prinsen CAC, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1171–1179.
15. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147–1157.
16. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27(5):1159–1170.
17. Hogg FRA, Peach G, Price P, Thompson MM, Hinchliffe RJ. Measures of health-related quality of life in diabetes-related foot disease: a systematic review. *Diabetologia*. 2012;55(3):552–565.
18. Jia Y, Huang H, Gagnier JJ. A systematic review of measurement properties of patient-reported outcome measures for use in patients with foot or ankle diseases. *Qual Life Res*. 2017;26(8):1969–2010.
19. Ortega-Avila A, Cervera-Garvi P, Ramos-Petersen L, Chicharro-Luna E, Gijon-Nogueron G. Patient-reported outcome measures for patients with diabetes mellitus associated with foot and ankle pathologies: a systematic review. *J Clin Med*. 2019;8(2), 146-146.
20. Smith SC, Lamping DL, Maclaine GDH. Measuring health-related quality of life in diabetic peripheral neuropathy: a systematic review. *Diabetes Res Clin Pract*. 2012;96(3):261–270.
21. Pérez-Panero AJ, Ruiz-Muñoz M, Fernández-Torres R, Formosa C, Gatt A, Gónzalez-Sánchez M. Diabetic foot disease: a systematic literature review of patient-reported outcome measures. *Qual Life Res*. 2021;30(12):3395–3405.
22. Terwee CB, Prinsen CAC, Ricci Garotti MG, Suman A, de Vet HCW, Mokkink LB. The quality of systematic reviews of health-related outcome measurement instruments. *Qual Life Res*. 2016;25(4):767–779.
23. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
24. Rethlefsen ML, Kirtley S, Waffenschmidt S, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst Rev*. 2021;10(1):39.
25. Terwee CB, Jansma EP, Riphagen II , de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115–1123.
26. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan—a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1):210.
27. Mokkink LB, Terwee CB, Knol DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10(1):22.
28. Suurmond R, van Rhee H, Hak T. Introduction, comparison, and validation of meta-essentials: a free and simple tool for meta-analysis. *Res Synth Methods*. 2017;8(4):537–553.
29. Abetz L, Sutton M, Brady L, McNulty P, Gagnon DD. The Diabetic Foot Ulcer Scale (DFS): a quality of life instrument for use in clinical trials. *Pract Diabetes Int*. 2002;19(6):167–175.
30. Acquadro C, Price P, Wollina U. Linguistic validation of the Cardiff Wound Impact Schedule into French, German and US English. *J Wound Care*. 2005;14(1):14–17.
31. Amesz SF, Klein TM, Meulendijks AM, et al. A translation and preliminary validation of the Dutch Wound-QoL questionnaire. *BMC Dermatol*. 2020;20(1):5.
32. Augustin M, Conde Montero E, Zander N, et al. Validity and feasibility of the wound-QoL questionnaire on health-related quality of life in chronic wounds. *Wound Repair Regen*. 2017;25(5):852–857.
33. Bann CM, Fehnel SE, Gagnon DD. Development and validation of the diabetic foot ulcer scale-short form (DFS-SF). *Pharmacoeconomics*. 2003;21(17):1277–1290.
34. Blome C, Baade K, Sebastian Debus E, Price P, Augustin M. The "Wound-QoL": a short questionnaire measuring quality of life in patients with chronic wounds based on three established disease-specific instruments. *Wound Repair Regen*. 2014;22(4):504–514.
35. Deufert D, Graml R. Disease-specific, health-related quality of life (HRQoL) of people with chronic wounds—a descriptive cross-sectional study using the Wound-QoL. *Wound Med*. 2017;16:29–33.
36. Dhawan V, Spratt KF, Pinzur MS, Baumhauer J, Rudicel S, Saltzman CL. Reliability of AOFAS diabetic foot questionnaire in Charcot arthropathy: stability, internal consistency, and measurable difference. *Foot Ankle Int*. 2005;26(9):717–731.
37. Fagerdahl A-M, Boström L, Ulfvarson J, Bergström G, Ottosson C. Translation and validation of the wound-specific quality of life instrument Cardiff Wound Impact Schedule in a Swedish population. *Scand J Caring Sci*. 2014;28(2):398–404.
38. Fagerdahl A-M, Bergström G. Translation and validation of a wound-specific, quality-of-life instrument (the wound-QoL) in a Swedish population. *Ostomy Wound Manage*. 2018;64(5):40–46.
39. Gamus A, Kaufman H, Keren E, Brandin G, Peles D, Chodick G. Validation of "wound QoL." Hebrew version disease-specific questionnaire for patients with lower extremity ulcerations. *Int Wound J*. 2018;15(4):600–604.
40. Goodridge D, Trepman E, Sloan J, et al. Quality of life of adults with unhealed and healed diabetic foot ulcers. *Foot Ankle Int*. 2006;27(4):274–280.
41. Huang Y, Wu M, Xing P, et al. Translation and validation of the Chinese Cardiff wound impact schedule. *Int J Low Extrem Wounds*. 2014;13(1):5–11.
42. Hui LF, Yee-Tak FD, Yam M, Yuk Ip W. Translation and validation of the Chinese diabetic foot ulcer scale – short form. *Patient Centered Outcomes Res*. 2008;1(2):137–145.

43. Jaksa PJ, Mahoney JL. Quality of life in patients with diabetic foot ulcers: validation of the Cardiff Wound Impact Schedule in a Canadian population. *Int Wound J*. 2010;7(6):502–507.
44. Kateel R, Augustine AJ, Ullal S, Prabhu S, Bhat R, Adhikari P. Development and validation of health related quality of life questionnaire (Indian scenario) in diabetic foot ulcer patients. *Diabetes Metab Syndr*. 2017;11(suppl 2): S651–S653.
45. Kontodimopoulos N, Veniou A, Tentolouris N, Niakas D. Validity and reliability of the Greek version of the diabetic foot ulcer scale - short form (DFS-SF). *Hormones (Athens)*. 2016;15(3):394–403.
46. Lee YN. Translation and validation of the Korean version of the diabetic foot ulcer scale-short form. *Int Wound J*. 2019;16(suppl 1):3–12.
47. Lozano-Platonoff A, Contreras-Ruiz J, Dominguez-Cherit J, Cardenas-Sanchez A, Alvarez-Rivero V, Martínez-Regalado JA. Translation, cross-cultural adaptation and validation of the "Cardiff wound impact schedule," a wound-specific quality of life instrument, to the native Spanish of Mexican patients. *Int Wound J*. 2020;17(1):167–173.
48. Macioch T, Sobol E, Krakowiecki A, Mrozikiewicz-Rakowska B, Kasprowicz M, Hermanowski T. Health related quality of life in patients with diabetic foot ulceration — translation and polish adaptation of diabetic foot ulcer scale short form. *Health Qual Life Outcomes*. 2017;15(1):15.
49. Martinez-Gonzalez D, Dòria M, Martínez-Alonso M, et al. Adaptation and validation of the diabetic foot ulcer scale-short form in Spanish subjects. *J Clin Med*. 2020;9(8):2497.
50. Oliveira Kaizer UA, Alexandre NMC, Rodrigues RCM, Cornélio ME, Melo Lima MH, São-João TM. Measurement properties and factor analysis of the diabetic foot ulcer scale-short form (DFS-SF). *Int Wound J*. 2020;17(3):670–682.
51. Price P, Harding K. Cardiff Wound Impact Schedule: the development of a condition-specific questionnaire to assess health-related quality of life in patients with chronic wounds of the lower limb. *Int Wound J*. 2004;1(1):10–17.
52. Ribu L, Rustøen T, Birkeland K, Hanestad BR, Paul SM, Miaskowski C. The prevalence and occurrence of diabetic foot ulcer pain and its impact on health-related quality of life. *J Pain*. 2006;7(4):290–299.
53. Sari Y, Purnawan I, Taufik A, Sumeru A. Quality of life and associated factors in Indonesian diabetic patients with foot ulcers. *Nurse Media J Nurs*. 2018;8(1):13–24.
54. Shukla VK, Shukla D, Tripathi AK, Agrawal S, Tiwary SK, Prakash V. Results of a one-day, descriptive study of quality of life in patients with chronic wounds. *Ostomy Wound Manag*. 2008;54(5):43–49.
55. Sommer R, Augustin M, Hampel-Kalthoff C, Blome C. The Wound-QoL questionnaire on quality of life in chronic wounds is highly reliable. *Wound Repair Regen*. 2017;25(4):730–732.
56. Sriyani KA, Gunawardena N, Wasalathanthri S, Hettiarachchi P. Validation of Sinhala version of Cardiff wound impact schedule in patients with diabetic leg and foot ulcers. *Asian Nurs Res*. 2016;10(3):240–245.
57. Toygar İ, Hançerlioğlu S, Gül S, Utku T, Şimşir IY, Çetinkalp Ş. Turkish adaptation of diabetic foot ulcer scale–short form. *Int J Low Extrem Wounds*. 2020;19(3):269–274.
58. van Doorn LP, Sijberden JP, Brouwers JJWM, Goossens LD, Hamming JF. Validation of the Dutch translation of the Cardiff wound impact schedule for evaluation of the health-related quality of life of patients with chronic wounds. *Int Wound J*. 2020;17(5):1225–1230.
59. Vileikyte L, Peyrot M, Bundy C, et al. The development and validation of a neuropathy- and foot ulcer-specific quality of life instrument. *Diabetes Care*. 2003;26(9):2549–2555.
60. Vinik EJ, Hayes RP, Oglesby A, et al. The development and validation of the Norfolk QOL-DN, a new measure of patients' perception of the effects of diabetes and diabetic neuropathy. *Diabetes Technol Ther*. 2005;7(3):497–508.
61. Vinik EJ, Paulson JF, Ford-Molvik SL, Vinik AI. German-translated Norfolk quality of life (QOL-DN) identifies the same factors as the English version of the tool and discriminates different levels of neuropathy severity. *J Diabetes Sci Technol*. 2008;2(6):1075–1086.
62. Vymětalová R, Zeleníková R. Painful ulceration and quality of life of patients with the diabetic foot syndrome. *Cent Eur J Nurs Midwif*. 2016;7(4):512–517.
63. da Franca Xavier AT, Foss MC, Marques Junior W, dos Santos CB, Onofre PTBN, Pace AE. Cultural adaptation and validation of the neuropathy - and foot ulcer - specific quality of life instrument (NeuroQol) for Brazilian Portuguese - phase 1. *Rev Lat Am Enfermagem*. 2011;19(6):1352–1361.
64. Granado-Casas M, Martinez-Gonzalez D, Martínez-Alonso M, et al. Psychometric validation of the Cardiff wound impact schedule questionnaire in a Spanish population with diabetic foot ulcer. *J Clin Med*. 2021;10(17):4023.
65. Putri NMME, Yasmara D, Yen MF, Pan SC, Fang SY. Body image as a mediator between gender and quality of life among patients with diabetic foot ulcers in Indonesia. *J Transcult Nurs*. 2021;32(6):655–663.
66. Ababneh A, Finlayson K, Lazzarini P, Edwards H. Cross-cultural adaptation and reliability testing of Arabic versions of several diabetic foot psychosocial scales [published online May 7, 2021]. Int J Low Extrem Wounds. https://doi.org/10.1177/15347346211016699.
67. Sommer R, von Stülpnagel CC, Fife CE, et al. Development and psychometric evaluation of the U.S. English Wound-QoL questionnaire to assess health-related quality of life in people with chronic wounds. *Wound Repair Regen*. 2020;28(5):609–616.
68. Conde Montero E, Sommer R, Augustin M, et al. Validación de la versión española del cuestionario Wound-QoL. *Actas Dermosifiliogr (Engl Ed)*. 2021;112(1):44–51.
69. Knudsen JT, Johansen CW, Hansen AØ, Eshoj HR. The Danish wound-quality of life (Wound-QoL) questionnaire: translation and psychometric properties. *Wound Repair Regen*. 2021;29(6):973–984.
70. Stülpnagel CC, Silva N, Augustin M, et al. Assessing the quality of life of people with chronic wounds by using the cross-culturally valid and revised Wound-QoL questionnaire. *Wound Repair Regen*. 2021;29(3):452–459.
71. Topp J, Blome C, Augustin M, et al. Determining the minimal important difference for the wound-QoL questionnaire. *Patient Prefer Adherence*. 2021;15:1571–1578.
72. Brett J, Staniszewska S, Mockford C, et al. Mapping the impact of patient and public involvement on health and social care research: a systematic review. *Health Expect*. 2014;17(5):637–650.
73. O'Regan M, Gethin G, O'Loughlin A, et al. Public & patient involvement to guide research in wound care in an Irish context. A round table report. *J Tissue Viability*. 2020;29(1):7–11.
74. Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN study design checklist for patient-reported outcome measurement instruments. COSMIN. https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf. Accessed September 15, 2021.