Qual Life Res (2017) 26:1969–2010
DOI 10.1007/s11136-017-1542-4

A systematic review of measurement properties of patient-reported outcome measures for use in patients with foot or ankle diseases

Yuanxi Jia[1] · Hsiaomin Huang[2] · Joel J. Gagnier[2,3]

Accepted: 3 March 2017 / Published online: 17 March 2017

Abstract
Purpose To identify currently available patient-reported outcome measures (PROMs) used in patients with foot or ankle diseases; and to critically appraise, compare and synthesize the psychometric evidence for the identified PROMs.
Methods Literature searches were performed in Medline and EMBASE from their inception to January 25th, 2016. Methodological quality was evaluated using the COSMIN checklist. The final rating of the methodological quality of each study for each property was the lowest rating among the items within that property. The psychometric evidence of the properties investigated in the included articles was assessed using the quality criteria established by Terwee et al. The methodological quality ratings and psychometric evidence assessments were synthesized using the method first proposed by Schellingerhout et al.
Results In total, 3077 articles were identified by the literature search, from which 115 studies investigating 50 PRO instruments were included in the review process. The Foot Function Index (FFI) was the most explored instrument, while the Manchester-Oxford Foot Questionnaire (MOXFQ) demonstrated the best properties.
Conclusion Most PROMs on foot and ankle diseases have limited evidence for their psychometric properties. The MOXFQ, with the highest overall ratings, could be a useful PROM for evaluating patients with foot or ankle diseases, based on current available evidence. More research is needed to improve the quality of the standards used to assess PROMs and the studies making these assessments.

Keywords Outcome measures · Systematic review · Foot disease · Ankle disease · Methodology · Psychometrics

Introduction

Diseases of the foot and ankle restrict activities of daily living and cause significant immobility and disability [1–3]. They are frequently secondary to traumatic or nontraumatic problems such as metatarsalgia, hallux valgus, abnormal position of toes, ankle sprain, and arthritis, and may also be attributed to improper footwear or abnormal biomechanics [4–7]. Approximately, 24% of women and 20% of men aged 18–80 years suffer from foot or ankle diseases [8].

Many outcome measures have been developed to evaluate various clinical interventions, many of which are patient-reported outcome measures (PROMs) [9]. Unfortunately, not all of these instruments are of optimal quality [10–12], and the heterogeneity in outcome measures used across the clinical research literature in this area makes it difficult to perform systematic reviews of the data. That is, data from separate instruments cannot be combined using meta-analytic methods.

A high-quality PROM should be reliable, valid, and responsive [13]. Several methods are available to evaluate the quality of the psychometric evidence for PROMs [14, 15]. For example, the COSMIN checklist is a

✉ Joel J. Gagnier
jgagnier@med.umich.edu

[1] Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA

[2] Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, MI, USA

[3] Department of Epidemiology, School of Public Health, University of Michigan, 24 Frank Lloyd Wright Drive, Lobby A, Ann Arbor, MI, USA

well-accepted appraisal tool for the assessment of the methodological quality of studies examining the psychometric properties of health-related quality-of-life instruments [14–16]. Also, Terwee et al. [16] developed criteria for assessing the level of psychometric evidence of a PROM. Since their inception in 2010, these criteria sets have been used to evaluate a variety of PROMs [13, 17–20].

There is currently no comprehensive study evaluating PROMs for patients with foot or ankle diseases using accepted criteria. The objectives of this study were to (1) identify currently available PROMs for foot or ankle disease patients; (2) critically appraise, compare and synthesize the psychometric evidence for the identified PROMs.

Methods

Search strategy

Literature searches were performed in Medline and EMBASE from their inception up to January 25, 2016. The search strategy (Appendix A) included terms such as foot, ankle, toe, and heal and all psychometric properties derived from Terwee et al. [21].

Selection criteria

Inclusion criteria for articles were as follows: original investigations detailing a nongeneric PROM's evaluation on its psychometric properties, population with disease/pathology of the foot or ankle, and publication in English language. Reference lists were separately and independently screened by two people, who met to review their assessments, and a third party was consulted to resolve disagreements. Duplicates were excluded and additional articles were culled from reference lists of included articles and relevant review papers.

Assessment of methodological quality

Methodological quality was evaluated using the COSMIN checklist [22]. Figure 1 describes the COSMIN checklist scoring procedures [23]. The methodological quality of articles evaluating the following measurement properties was assessed: internal consistency, reliability, measurement error, content validity, construct validity (including structural validity and hypothesis testing), criterion validity, and responsiveness. We used the updated scoring system that was developed in 2012 for the COSMIN checklist [16], which contains four possible response options (excellent, good, fair, poor). The final rating of the methodological

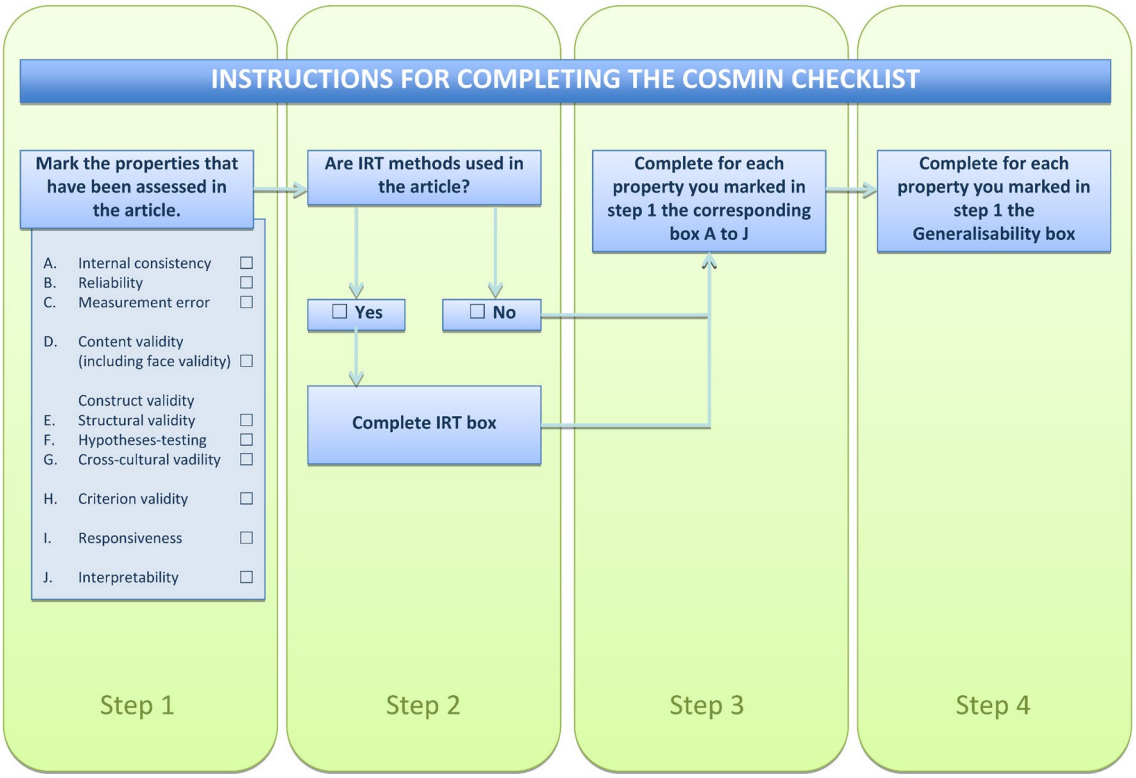

Fig. 1 Instructions for completing the COSMIN checklist

quality of each study for each property was given by the lowest rating among the items within that property. For example, if a property (e.g., internal consistency) gets an excellent on one item but a poor on another, then the overall methodological quality for that property is poor. Two reviewers assessed the methodological quality of the articles separately and independently using this updated COSMIN scoring system. Disagreement was resolved by discussion, or, if consensus was not reached after discussion, a third party was consulted to resolve the disagreement. Detailed COSMIN scoring methods are available online at http://www.cosmin.nl/the-cosmin-checklist_8_5.html.

Assessment of psychometric properties

The psychometric evidence of the measurement properties investigated in the included articles was assessed using the quality criteria established by Terwee et al. [15] (Table 1). The rating scale includes quality criteria for internal consistency, reliability, measurement error, content validity, criterion validity, structural validity, hypothesis testing (including convergent validity, discriminant validity and discriminant validity), and responsiveness. Each criterion was rated as positive (+), indeterminate/unknown (?), or negative (−). If no information for the property is available in the literature, a rating of zero (0) was given to the property. Interpretability, according to the COSMIN manual, is a characteristic of an instrument that relates to the ability to assign qualitative meaning to an instrument's quantitative scores or changes in scores. To assess interpretability, we evaluated the existence of ceiling/floor effects and extracted the minimally important change (MIC) and minimal detectable change (MDC).

Synthesis method

The methodological quality ratings and psychometric evidence assessments were synthesized using the method first proposed by Schellingerhout et al. [24]. This method has previously been used in systematic reviews of PROMs [25]. The overall synthesis score combines the consistency of the psychometric evidence with the methodological quality of the included studies and the level of evidence proposed by the Cochrane Back Review Group [26]. The overall results are then categorized as positive (+), unknown/indeterminate (?), negative (−), or no evidence (0) accompanied with the overall level of evidence ranging from unknown to strong (see Table 2). A rating of conflicting results (−/+) is given when the number of positive ratings equals the number of negative ratings. Using this method, when combined across studies, the levels of evidence are: strong (representing consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality), moderate (representing consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality), limited (representing one study of fair methodological quality), conflicting (representing conflicting findings), and unknown (representing the existence of only studies of poor methodological quality).

Results

The literature search yielded 3077 articles. Excluding nonpertinent papers, 115 studies investigating 50 PRO instruments were included in this review with several studies assessing multiple instruments [28–141]. The included PRO instruments are summarized in Table 3. The synthesized evidence for methodological quality and the psychometric properties are summarized for all instruments in Table 4, and detailed assessments for each instrument are provided in the tables in Appendix B and in written text in Appendix C.

Psychometric properties

According to the available evidence we found, the MOXFQ [45] had the best overall psychometric properties, with positive evidence on the internal consistency, reliability, measurement error, structural validity, convergent validity, discriminant validity, discriminative validity, and responsiveness. There was positive evidence for the FAAM [62] on four properties (reliability, measurement error, structural validity, and discriminant validity), but also negative evidence on two properties (internal consistency and convergent validity). We found positive evidence on three properties (internal consistency, structural validity, and convergent validity) and negative evidence on two properties (reliability and responsiveness) for the MFPDI [35]. Also, for the FAOS [142], there was positive evidence on three properties (structural validity, convergent validity, and discriminant validity), negative evidence on one property (internal consistency) and conflicting evidence on one property (reliability). Evidence for only one property was found for eight instruments, and no evidence was found for 27 instruments.

Methodological quality

Overall

The most important aspects of each property (except criterion validity) that jeopardized methodological quality were as follows:

Table 1 Quality criteria for measurement properties

Property	Description [15, 22]	Rating	Quality criteria
Reliability			
Internal consistency	The extent to which items in a questionnaire (sub)scale are correlated (homogeneous), thus measuring the same concept	+	(Sub)scale unidimensional AND Cronbach's alpha(s) ≥ 0.70
		?	Dimensionality not known OR Cronbach's alpha not determined
		−	(Sub)scale not unidimensional OR Cronbach's alpha(s) < 0.70
Reliability	The extent to which patients can be distinguished from each other, despite measurement error (relative measurement error)	+	ICC/weighted Kappa ≥ 0.70 OR Pearson's $r \geq 0.80$
		?	Neither ICC/weighted Kappa, nor Pearson's r determined
		−	ICC/weighted Kappa < 0.70 OR Pearson's $r < 0.80$
Measurement error	The extent to which the scores on repeated measures are close to each other (absolute measurement error)	+	MIC > SDC OR MIC outside the LOA
		?	MIC not defined
		−	MIC $\leq$ SDC OR MIC equals or inside LOA
Validity			
Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	+	The target population considers all items in the questionnaire to be relevant AND considers the questionnaire to be complete
		?	No target population involvement
		−	The target population considers items in the questionnaire to be irrelevant OR considers the questionnaire to be incomplete
Structural validity	The extent to which the scores of an instrument are an adequate reflection of the dimensionality of the construct to be measured	+	Factors should explain at least 50% of the variance
		?	Explained variance not mentioned
		−	Factors explain <50% of the variance
Hypothesis testing	The extent to which scores on a particular instrument relate to other measures in a manner that is consistent with theoretically derived hypotheses concerning the concepts that are being measured	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses) AND correlation with related constructs is higher than with unrelated constructs
		?	Solely correlations determined with unrelated constructs
		−	Correlation with an instrument measuring the same construct < 0.50 OR <75% of the results are in accordance with the hypotheses OR correlation with related constructs is lower than with unrelated constructs
Criterion validity	The extent to which scores on a particular instrument relate to a gold standard	+	Convincing arguments that gold standard is "gold" AND correlation with gold standard ≥ 0.7
		?	No convincing arguments that gold standard is "gold" OR doubtful design or method
		−	Correlation with gold standard < 0.7 despite adequate design and method
Responsiveness			

Table 1 (continued)

Property	Description [15, 22]	Rating	Quality criteria
Responsiveness	Responsiveness is a measure of longitudinal validity. In analogy to construct validity, longitudinal validity should be assessed by testing predefined hypotheses, e.g., about expected correlations between changes in measures, or expected differences in changes between ''known'' groups	+	(Correlation with an instrument measuring the same construct ≥ 0.50 OR at least 75% of the results are in accordance with the hypotheses OR AUC ≥ 0.70) AND correlation with related constructs is higher than with unrelated constructs
		?	Solely correlations determined with unrelated constructs
		−	Correlation with an instrument measuring the same construct <0.50 OR $<75\%$ of the results are in accordance with the hypotheses OR AUC <0.70 OR correlation with related constructs is lower than with unrelated constructs

+ positive rating, ? indeterminate rating, − negative rating

AUC area under curve, *ICC* intraclass correlation coefficient, *LOA* limits of agreement, *MIC* minimal important change, *SDC* smallest detectable change

Internal consistency: The unidimensionality of the domains was not explored.

Reliability: The interval was not acceptable.

Measurement error: The interval was not acceptable.

Content validity: Only relevance was explored.

Structural validity: How to handle missing values was not mentioned.

Hypothesis validity: A priori was not made.

Reliability: No comparator was used.

Appendix B contains details of the methodological quality (i.e., COSMIN assessment) of each instrument for each included study.

Time interval

For reliability and measurement error, the interval between the two administrations of an instrument varied a lot, while a 2-week interval is recommended by the COSMIN checklist [23]. Among the 84 assessments of reliability, fewer than 30 assessments explicitly used the 2-week interval. Among the 25 assessments of measurement error, 18 failed to use the 2-week interval.

Criterion validity or construct validity

In most situations, we could not find a gold standard for a PROM, except for the use of the original long version of a PROM as the gold standard for a short version [23]. According to the COSMIN checklist, methodology should be rated as poor if other PROMs are used as the gold standard. But it was also reasonable to consider the 'criterion validity' to be the true construct validity only because the authors used different taxonomy. Because the subjective judgments required by the checklist as the terms for measurement properties used in an article may not be similar to the terms used in COSMIN, our study treated all assessments on criterion validity as construct validity to save research resources.

Cross-cultural adaptation

Cross-cultural adaptation is an important property to be measured when an instrument is used in settings unlike that in which it originated. It is defined as the degree to which the performance of the items in a translated or culturally adapted instrument is an adequate reflection of the performance of the items in the original version of the instrument [23]. In our review, 34 studies attempted to translate the original version into another language, including 14 instruments (VAS-FA, Q-DFD, OxAFQ, NeuroQol, MOXFQ, MFPDI, HFS-14, FISRA, FFI, FAOS, AAOS-FAQ, FAAM, DFUS-SF, and CAIT). Three studies did not mention whether the translation process was independent

Table 2 Levels of evidence for the overall quality of the measurement property [27]

Level	Rating	Criteria
Strong	+++ or −−−	Consistent findings in multiple studies of good methodological quality OR in one study of excellent methodological quality
Moderate	++ or −−	Consistent findings in multiple studies of fair methodological quality OR in one study of good methodological quality
Limited	+ or −	One study of fair methodological quality
Conflicting	+/−	Conflicting findings
Unknown	?	Only studies of poor methodological quality

[..] Reference number, + positive result, − negative result

between the two experts [65, 72, 106]. One study failed to explain how the discrepancies were handled [36]. No committee was involved in ten studies [36, 53, 65, 67, 68, 72, 79, 80, 88, 122]. Ten studies did not pre-test the translated version in similar patients to make necessary revisions [43, 53, 65, 67, 72, 79, 88, 89, 126, 138]. The overall quality was poor; only one study was conducted among two groups of patients for the original language and targeted language with factor analysis [40].

Discussion

Our systematic review identified 50 instruments for patients with foot or ankle diseases. These instruments vary greatly in the exact condition and domains assessed (see Table 3). The FFI was most explored, while the MOXFQ proved to possess the best properties according to the available evidence so far. But, one must be careful to choose an instrument specific to their needs, whether scientific or clinical.

To note, a lack of evidence, or a preponderance of negative evidence, does not imply that an instrument is unsuitable; further studies are needed to explore the evidence before a confirmatory decision can be made. More effort is needed to determine how best to interpret and synthesize all the evidence, and determine the conditions needed to make recommendations that are well supported by the evidence. Furthermore, the evidence derived from a study may depend on the characteristics of the patients involved [162]. For example, one instrument was not likely to show good responsiveness when the patients improved only slightly following a certain treatment. With increasing patient variation, comparisons across studies become difficult to make, so more rigorous methods are needed to overcome the potential bias associated with study subjects.

Also, one must be careful when assessing reliability. Specifically, the time interval should be long enough to prevent bias, and short enough to ensure that patients have not been changed on the construct to be measured. But an appropriate time interval depends on the construct to be measured and the target population. The 2-week interval may not be appropriate in some circumstances—for example, diseases that may change rapidly and inclusion of elderly subjects who may have poor memory.

Strengths and limitations

This study has several strengths. It is the first study to systematically evaluate the methodological quality and psychometric evidence of PROMs for patients with foot or ankle diseases and the first to synthesize the evidence to give an overall assessment of the PROMs. Furthermore, the study used the COSMIN checklist, Terwee et al.'s criteria, and the synthesis method that had been validated and used frequently in other research [13, 17–20, 25, 163, 164]. Also, with the aid of a library and information scientist, multiple databases and the reference lists of included articles were searched; thus, we are confident that all available studies that met the inclusion criteria were collected. It is possible that the search failed to capture studies that have evaluated these instruments in other languages. To date, there is no known evidence for a language bias in systematic reviews of psychometric evidence when non-English language studies are included. Also, it appears that systematic reviews and meta-analyses of clinical trials in conventional medicine do not have biased estimates of treatment effect when they exclude non-English language studies [165, 166]. Another potential limitation was that in many cases the literature was incompletely reported, making our assessments difficult, if not impossible in some cases. While this was not a drawback of the methods of our systematic review, it did affect our ability to clearly evaluate these instruments. Finally. As stated above, the instruments reviewed here vary widely in the number of items and the types of domains being assessed. Thus, while the MOXFQ has the best evidence for its properties, one should refer to Table 3 to determine what instrument, and associated domains, best fit their need. In some cases, the instrument with the best psychometric evidence may not fit a specific research question or patient population. But, we recommend that PROMs with evidence that they lack a specific property (negative evidence) should be viewed with caution.

Future research

In this study, very limited evidence was available due to the questionable quality of the targeted studies. The

Table 3 Details of each instrument and number of included studies for each instrument

Name of the instrument	Year of publication	No. of items and domains	Domains	Response options
The Manchester-Oxford Foot Questionnaire (MOXFQ) [45]	2000	17 items in 3 subdomains	Functional limitation, Pain, Personal Appearance	Three levels
The Manchester Foot Pain and Disability Index (MFPDI) [35]	2006	16 items in 3 subdomains	Pain, Walking/standing, Social interaction	Five points Likert scale
The Foot Function Index (FFI) [47]	1991	23 items in 3 subdomains	Pain, Disability, Activity Limitations	VAS
The Foot and Ankle Ability Measure (FAAM) [62]	2005	29 items in 2 domains	Activities of Daily Living, Sports	Five points Likert scale
The Foot and Ankle Outcome Score (FAOS) [142]	1998	42 items in 5 domains	Pain, Other Symptoms Like Stiffness, Swelling, and Range of Motion; Activities of Daily Living; Sport and Recreational Activities; Foot and Ankle-Related Quality of Life	Five points Likert scale
The Foot Health Status Questionnaire (FHSQ) [85]	1998	13 items in 4 domains	Pain, Function, Footwear, General Foot Health	Five points Likert scale
The Cumberland Ankle Instability Tool [86]	2006	9 items in 1 domain	Unidimensional	Item dependent
The Short Musculoskeletal Function Assessment Questionnaire-Dysfunction Index (SMFAQ-DI) [143]	1999	34 items in 4 domains	Daily activities, Emotional status, Arm/hand function, and Mobility	Five points liker scale
The Short Musculoskeletal Function Assessment Questionnaire-Bother Index (SMFAQ-BI) [143]	1999	12 items in 1 domain	Unidimensional	Five points Likert scale
The Ankle Osteoarthritis Scale (AOS) [92]	1998	18 item in 2 domains	Pain, Disability	VAS
The Western Ontario and McMaster Universities Osteoarthritis Index (WOMUOI) [144]	1988	24 items in 3 domains	Pain, Stiffness, Physical Dysfunction	Five points Likert scale
The Juvenile Arthritis Foot Disability Index (JAFD) [94]	2004	27 items in 3 domains	Impairment, Activity Limitation, Participation Restriction	Five points Likert scale
The Oxford Ankle Foot Questionnaire for Children (OxAFQ) [96]	2008	15 items in 3 domains	Physical, School and Play, Emotional	Five points Likert scale
The Self-Administered Foot Evaluation Questionnaire (SAFE-Q) [139, 145]	2011	34 items in 5 domains	Pain and Pain-Related; Physical Functioning and Daily Living; Social Functioning; Shoe-Related; 5: General Health and Well-Being	Item dependent
The Self-reported Foot and Ankle Score (SEFAS) [146]	2007	12 items	Dimensions not defined	Five Likert scale
ROwan Foot Pain Assessment Questionnaire (ROFPAQ) [105]	2001	39 items in 3 domains	Sensory, Affective, Cognitive	Five Likert scale
Foot Impact Scale for Rheumatoid Arthritis (FISRA) [147]	2005	51 items in 2 domains	Impairment/Footwear; Activities/Participation	Binary

Table 3 (continued)

Name of the instrument	Year of publication	No. of items and domains	Domains	Response options
Diabetic Foot Ulcer Scale (DFUS) [108]	2002	58 items in 11 domains	Leisure, Physical Health, Daily Activities Emotions, Noncompliance, Family, Friends, Positive Attitude, Treatment, Satisfaction, Financial	Five Likert scale
Diabetic Foot Ulcer Scale-Short Form (DFUS-SF) [109]	2003	29 items in 6 domains	Leisure, Dependence/Daily Life, Negative Emotions, Physical Health, Worried about Ulcers/feet, Bothered by Ulcer Care	Five Likert scale
Diabetes Foot Self-Care Behavior Scale (DFSCBS) [111]	2013	7 items in 1 domain	Unidimensional	Four points Likert scale
The American Orthopaedic Foot and Ankle Society Clinical Rating Scale-Ankle/Hindfoot (AOFAS-AH) [148]	1994	9 items in 3 domains	Pain, Function, Alignment	Item dependent
The American Orthopaedic Foot and Ankle Society Clinical Rating Scale-Midfoot (AOFAS-M) [148]	1994	9 items in 3 domains	Pain, Function, Alignment	Item dependent
The American Orthopaedic Foot and Ankle Society Clinical Rating Scale-Hallux Metatarsophalangeal-interphalangeal Joints (AOFAS-HJ) [148]	1994	8 items in 3 domains	Pain, Function, Alignment	Item dependent
The American Orthopaedic Foot and Ankle Society Clinical Rating Scale-Lesser metatarsophalangeal-interphalangeal Joints (AOFAS-LJ) [148]	1994	8 items in 3 domains	Pain, Function, Alignment	Item dependent
The Foot and Ankle Disability Index (FADI) [149]	1999	26 items. Dimensionality was not defined		Five points Likert scale
The Foot and Ankle Disability Index Sport (FADI Sport) [149]	1999	8 items in 1 domain	Unidimensional	Five points Likert scale
The American Orthopaedic Foot and Ankle Society Diabetic Foot Questionnaire (AOFAS-DFQ) [150]	1994	66 items in 6 domains	General Health, Physicality, Emotion, Worry, Foot Status, and Care	Three points Likert scale
The Sports Athlete Foot and Ankle Score (SAFAS) [120]	2013	42 items in 4 domains	Symptoms, Pain, Daily Living, Sports	Five points Likert scale
The Hand-foot Syndrome Specific Quality of Life Questionnaire (HFS-14) [151]	2011	14 items in 3 domains	Hands, Feet, Social	Three points Likert scale
The Neuropathy and Foot Ulcer Specific Quality of Life instrument (NeuroQol) [123]	2003	35 items in 6 domains	Pain, Loss/Reduction of Sensitivity, Diffuse Sensory-motor Symptoms, Limitations In Dally Activities, Interpersonal Problems, Emotional Distress	Three points Likert scale

Table 3 (continued)

Name of the instrument	Year of publication	No. of items and domains	Domains	Response options
The Japanese Society of Surgery of the Foot standard rating system (JSSF standard rating system)	2005	5 scales, Ankle-hindfoot scale: 3 domains 7 items, Midfoot scale, 3 domains 5 items, Hallux metatarsophalangeal-interphalangeal scale, 3 domains 7 items, Lesser metatarsophalangeal-interphalangeal scale, 3 domains 7 items, Rheumatoid arthritis foot and ankle scale, 4 domains 12 items		Item dependent
The Visual Analogue Scale-Foot and Ankle (VAS-FA) [152]	2006	20 items in 3 domains	Pain, Function, Other complaints	Visual Analogue Scales
The Cardiff Wound Impact Schedule (CWIS) [153]	2004	26 items in 3 domains	Social Life, Well-Being, Physical Symptoms, and Daily Living	Five points Likert scale
The Musculoskeletal Functional Assessment (MFA) [154]	1996	101 items in 10 domains	Self-care, Sleep and Rest, Hand and Fine Motor Skills, Mobility, Housework, Employment, Leisure, Family Relationships, Cognition and Emotional Adjustment, Coping, and Adaptation	Dichotomous
The Questionnaire for Usability Evaluation of Orthopaedic Shoes Pre-test (QUEOS Pre-test) [128]	2004	67 items in 12 domains	Pain during daily activities, stability during daily activities, callus, wounds, pinch, slip, weight of shoes, cold feet, perspiration, putting on/taking off shoes, maintenance, and cosmetic appearance	Dichotomous/Visual Analogue Scales
The Questionnaire for Usability Evaluation of Orthopaedic Shoes Post-test (QUEOS Post-test) [128]	2004	67 items in 12 domains	Pain during daily activities, stability during daily activities, callus, wounds, pinch, slip, weight of shoes, cold feet, perspiration, putting on/taking off shoes, maintenance, and cosmetic appearance	Dichotomous/Visual Analogue Scales
The Oswestry Disability Questionnaire (ODQ) [155]	1980	10 items in 1 domain		Six points Likert scale
American College of Foot and Ankle Surgeons Scoring Scales (ACFAS scoring scale) [156]	2011	Module 1: First Metatarsophalangeal Joint (MPJ) and First Ray (11 items) Module 2: Forefoot (Excluding First Ray) (12 items) Module 3: Rearfoot (Including Flatfoot) (16 items) Module 4: Ankle (22 items)		Scores
The Olerud-Molander Ankle Score (OMAS) [157]	1984	9 items. Dimensionality was not identified		Scores
The Patient Reported Outcome Measurement Information System Physical Function Computerized Adaptive Tests (PROMIS PF CAT) [158]	2007	Computer Adaptive Test		
The Hand-Foot and Mucositis Symptom and Impact Questionnaire (HAMSIQ) [159]	2013	14 items in 4 domains	Degree of Mouth/Throat, Hand and Foot Soreness, Activity Limitation Mouth/Throat Limitations, Foot Limitations	Four points Likert scale
The Hand-Foot Skin Reaction and Quality of Life Questionnaire (HF-Qol) [134]	2015	18 items in 4 domains*	Gross Physical Function, Hand-Related Functioning, Social Activities, and Psychological Impact of Skin Toxicity	Five points Likert scale

Table 3 (continued)

Name of the instrument	Year of publication	No. of items and domains	Domains	Response options
The Combined Foot Care Confidence Scale/ Foot-Care Behavior Instrument (FCCS-FCB) [160]	2015	29 items in two scales, 2 domains and 6 domains, respectively	Self-efficacy: Foot Self-care, Clinical Aspects of Foot Care. Behavior: Risk Behaviors for Foot care, Footwear, Foot-care Hygiene, the Use of Proper Footwear, the Purchase of Proper Footwear, Foot Care	Item-dependent
The Salford Rheumatoid Arthritis Foot Evaluation Part A (SAFE-Part A) [136]	2012	19 items in 3 domains	Impairment, Disability, Footwear	Scores
The diabetic Foot Self-care Questionnaire of the University of Malaga (DFSQ-UMA) [137]	2015	20 items in 3 domains	Personal Self-care, Podiatric Care, Footwear and Socks	Five points Likert scale
The Questionnaire for Diabetes-Related Foot Disease (Q-DFD) [161]	2009	12 items. Dimensionality was not identified		Item-dependent
The Outcome Instrument for the Foot and Ankle Version 2 (OIFA-2) [139]	2011	34 items in 5 domains and 9 sport items	Foot Pain and Foot Pain-related, 2 Physical Functioning and Daily Living; 3 Social Functioning; 4 Shoe-related, 5 General Health and Well-being	Item-dependent
The Chronic Ankle Instability Scale (CAIS) [140]	2008	14 items in 4 domains	Impairments subscale (6 items) Disabilities subscale (5 items) Participation subscale (2 items) Emotions subscale (2 items)	Five points Likert scale
The Telephone Questionnaire [141]	2015	32 items in 3 domain	Function, daily life, psychology	Item-dependent

*Only the patient-reported part

Table 4 Overall rating per property per questionnaire

Instruments	No. of studies	Internal consistency	Reliability	Measurement error	Content Validity	Structural Validity	Hypothesis validity			Responsiveness
							Convergent	Discriminant	Discriminative	
The Manchester-Oxford Foot Questionnaire (MOXFQ)	10	++	++	++	0	++	+++	++	0	++
The Manchester Foot Pain and Disability Index (MFPDI)	8	++	−	?	0	++	+	?	?	−
The Foot Function Index (FFI)	16	−	?	?	0	+	++	−	0	?
The Foot and Ankle Ability Measure (FAAM)	9	−−−	++	++	0	+++	−−	++	?	?
The Foot and Ankle Outcome Score (FAOS)	14	−−−	+/−	?	?	++	++	++	0	?
The Foot Health Status Questionnaire (FHSQ)	3	?	?	0	0	++	?	?	0	?
The Cumberland Ankle Instability Tool	5	−−	?	0	0	+++	?	0	0	?
The Short Musculoskeletal Function Assessment Questionnaire-Dysfunction Index (SMFAQ-DI)	3	−−−	?	?	0	+++	?	0	?	0
The Short Musculoskeletal Function Assessment Questionnaire-Bother Index (SMFAQ-BI)	2	+++	?	?	0	+++	?	0	?	0
Ankle Osteoarthritis Scale (AOS)	4	?	?	?	0	0	?	0	0	?
The Western Ontario and McMaster Universities Osteoarthritis Index (WOMUOI)	1	?	?	?	0	0	?	0	0	0
The Juvenile Arthritis Foot Disability Index (JAFD)	2	?	?	0	?	0	?	0	?	?
The Oxford Ankle Foot Questionnaire for Children (OxAFQ)	4	?	?	?	0	0	−−−	++	?	?
The Self-Administered Foot Evaluation Questionnaire (SAFE-Q)	3	0	+	0	0	0	?	0	?	?
The Self-reported Foot and Ankle Score (SEFAS)	3	?	?	?	?	0	++	++	0	?
ROwan Foot Pain Assessment Questionnaire (ROFPAQ)	1	?	?	0	?	0	?	?	0	0
Foot Impact Scale for Rheumatoid Arthritis (FISRA)	2	−	0	0	0	0	0	0	0	0
Diabetic Foot Ulcer Scale (DFUS)	1	?	?	0	?	0	?	0	?	?
Diabetic Foot Ulcer Scale-Short Form (DFUS-SF)	2	++	−	0	0	?	+/−	0	0	?
Diabetes Foot Self-Care Behavior Scale (DFSCBS)	1	+	+	0	?	−	?	?	0	0
The American Orthopaedic Foot and Ankle Society Clinical Rating Scale-Ankle/Hindfoot (AOFAS-AH)	7	?	?	?	0	0	?	0	0	?
The American Orthopaedic Foot and Ankle Society Clinical Rating Scale-Midfoot (AOFAS-M)	3	0	?	?	0	0	?	0	0	?
The American Orthopaedic Foot and Ankle Society Clinical Rating Scale-Hallux Metatarsophalangeal-interphalangeal Joints (AOFAS-HJ)	4	?	?	?	0	0	?	0	0	?
The American Orthopaedic Foot and Ankle Society Clinical Rating Scale-Lesser metatarsophalangeal-interphalangeal Joints (AOFAS-LJ)	4	?	?	?	0	0	?	0	0	?
The Foot and Ankle Disability Index (FADI)	2	0	?	0	0	0	0	0	0	?
The Foot and Ankle Disability Index Sport (FADI Sport)	2	0	?	0	0	0	0	0	0	?

Table 4 (continued)

Instruments	No. of studies	Internal consistency	Reliability	Measurement error	Content Validity	Structural Validity	Hypothesis validity			Responsiveness
							Convergent	Discriminant	Discriminative	
The American Orthopaedic Foot and Ankle Society Diabetic Foot Questionnaire (AOFAS-DFQ)	1	?	?	?	0	0	0	0	0	0
The American Orthopaedic Foot and Ankle Society Diabetic Foot Questionnaire (AOFAS-DFQ)	3	?	?	0	0	0	?	0	0	?
The Sports Athlete Foot and Ankle Score (SAFAS)	1	?	0	0	?	0	+	−	0	0
The Hand-foot Syndrome Specific Quality of Life Questionnaire (HFS-14)	2	−	?	0	0	0	?	?	?	0
The Neuropathy and Foot Ulcer Specific Quality of Life instrument (NeuroQol)	2	−	0	0	0	?	-	0	?	0
The Japanese Society of Surgery of the Foot standard rating system (JSSF standard rating system)	1	0	?	0	0	0	0	0	0	0
Visual Analogue Scale-Foot and Ankle (VAS-FA)	2	?	?	0	0	0	?	0	0	0
The Cardiff Wound Impact Schedule (CWIS)	1	0	?	0	0	0	?	0	0	0
The Musculoskeletal Functional Assessment (MFA)	1	0	0	0	0	0	0	0	0	?
The Questionnaire for Usability Evaluation of Orthopaedic Shoes Pre-test (QUEOS Pre-test)	1	?	?	0	0	0	0	0	0	0
The Questionnaire for Usability Evaluation of Orthopaedic Shoes Post-test (QUEOS Post-test)	1	?	?	0	0	0	0	0	0	0
The Oswestry Disability Questionnaire (ODQ)	1	0	0	0	0	0	0	0	0	?
American College of Foot and Ankle Surgeons Scoring Scales (ACFAS scoring scale)	1	0	?	0	?	0	0	0	0	?
The Olerud-Molander Ankle Score (OMAS)	1	?	?	?	0	0	?	0	0	0
The Patient Reported Outcome Measurement Information System Physical Function Computerized Adaptive Tests (PROMIS PF CAT)	1	+++	0	?	0	0	?	0	?	0
The Hand-Foot and Mucositis Symptom and Impact Questionnaire (HAMSIQ)	1	?	0	0	0	0	−	0	?	?
The Hand-Foot Skin Reaction and Quality of Life Questionnaire (HF-Qol)	1	?	?	0	0	?	?	0	?	?
The Combined Foot Care Confidence Scale/ Foot-Care Behavior Instrument (FCCS-FCB)	1	?	0	0	0	−	0	0	0	0
The Salford Rheumatoid Arthritis Foot Evaluation Part A (SAFE-Part A)	1	0	?	0	0	0	?	0	0	0
The Diabetic Foot Self-care Questionnaire of the University of Malaga (DFSQ-UMA)	1	?	?	0	0	+	?	0	0	0
The Questionnaire for Diabetes-Related Foot Disease (Q-DFD)	1	0	?	0	0	0	?	0	0	0

Table 4 (continued)

Instruments	No. of studies	Internal consistency	Reliability	Measurement error	Content Validity	Structural Validity	Hypothesis validity			Responsiveness
							Convergent	Discriminant	Discriminative	
The Outcome Instrument for the Foot and Ankle Version 2 (OIFA-2)	1	–	0	0	0	+	?	0	0	0
The Chronic Ankle Instability Scale (CAIS)	1	?	?	?	?	0	?	0	0	?
The Telephone Questionnaire	1	?	?	0	0	0	?	0	0	0

clinical decision to choose an instrument requires substantial rigorous study to yield stable and reliable evidence. We recommend more studies with better quality focusing on the development and assessment of PROMs in the future.

The consensus-based COSMIN checklist was the first attempt to standardize the assessment process for PROMs [23]. Such a guideline made it possible to compare PROMs across studies. We recommend future studies be designed according to the COSMIN checklist.

The present study combined the COSMIN methodological quality criteria, Terwee's psychometric criteria and Schellingerhout's rules synthesize the evidence. But, these 3 criteria, while commonly used, were not developed together and thus have some discrepancies. For example, while the COSMIN checklist considers three aspects of hypothesis validity (convergent, discriminant, and discriminative validity), Terwee's rules mention only overall hypothesis validity. We encourage future methodological guidance be created for performing systematic reviews of the psychometric properties of PROMs.

Furthermore, a core outcome set (COS) could be established for PROMs for use in patients with foot and ankle conditions. A COS is defined as an agreed upon minimum set of outcomes or outcome measures to use and report clinical research in specific clinical studies [167]. A COS allows users of clinical research that includes PROMs to make meaningful comparisons between studies on similar questions, and furthermore, avoids waste and untrustworthy results by using outcome measures in clinical research with established known properties. Given the array of PROMs used in patients with foot and ankle conditions, and their variable quality, some work on establishing COSs is needed.

Conclusions

Most PROMs on foot and ankle diseases have limited evidence for their psychometric properties. Although not all of the properties were studied, the MOXFQ, with the highest overall ratings, could be a useful PROM for evaluating patients with foot or ankle diseases, on the basis of the available evidence. However, one should proceed with caution in choosing an instrument because no one PROM included here had positive evidence on all ten properties. Also, one must be careful to choose and instrument that assesses the domains they are interested in. Finally, more effort should be made to improve the quality of the standards and studies that assess PROMs.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

Informed consent This systematic review did not require informed consent.

Appendix A: Searching strategy

Medline

ENG[LA] and

(scale[tiab] or scales[tiab] or instrument[tiab] or instruments[tiab] or questionnaire[tiab] or questionnaires[tiab] or form[tiab] or forms[tiab] or score[tiab] or scores[tiab] or measurement[tiab] or measurements[tiab]) AND

(foot[MeSH] or foot[Title] or feet[Title]or ankle[Title] or ankles[Title] or toe[Title] or toes[Title] or heel[Title] or heels[Title]) AND

(reproducibility[tiab] OR psychometrics[MeSH] OR psychometr*[tiab] OR reliability[tiab] OR validity[tiab] OR consistency[tiab] OR cronbach*[tiab] OR agreement[tiab] OR precision[tiab] OR test-retest [tiab] OR (test[tiab] AND retest[tiab]) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatability[tiab] OR replicability[tiab] OR concordance[tiab] OR intraclass[tiab] OR discriminative[tiab] OR discriminant[tiab] OR variability[tiab] OR sensitivity[tiab] OR responsiveness [tiab] OR interpretability[tiab] OR "ceiling effect" [tiab] OR "floor effect" [tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning" [tiab] OR DIF[tiab] OR "computer adaptive testing" [tiab] OR "cross-cultural equivalence" [tiab])

Embase

ENG[LA] and

(scale[tiab] or scales[tiab] or instrument[tiab] or instruments[tiab] or questionnaire[tiab] or questionnaires[tiab] or form[tiab] or forms[tiab] or score[tiab] or scores[tiab] or measurement[tiab] or measurements[tiab]) AND

(foot[MeSH] or foot[Title] or feet[Title]or ankle[Title] or ankles[Title] or toe[Title] or toes[Title] or heel[Title] or heels[Title]) AND

(reproducibility[tiab] OR psychometrics[MeSH] OR psychometr*[tiab] OR reliability[tiab] OR validity[tiab] OR consistency[tiab] OR cronbach*[tiab] OR agreement[tiab] OR precision[tiab] OR test-retest [tiab] OR (test[tiab] AND retest[tiab]) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatability[tiab] OR replicability[tiab] OR concordance[tiab] OR intraclass[tiab] OR discriminative[tiab] OR discriminant[tiab] OR variability[tiab] OR sensitivity[tiab] OR responsiveness [tiab] OR interpretability[tiab] OR "ceiling effect" [tiab] OR "floor effect" [tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning" [tiab] OR DIF[tiab] OR "computer adaptive testing" [tiab] OR "cross-cultural equivalence" [tiab])

Appendix B

See Table 5.

Appendix C: Detailed description of psychometric properties for each included instrument

Manchester-oxford foot questionnaire (MOXFQ)

Ten studies evaluated the psychometric properties of the MOXFQ [28–37]. Due to poor quality, no evidence of content validity or criterion validity was found.

Internal consistency: Moderate positive evidence was found in one study with good quality, for which the Cronbach's α coefficients for the unidimensional walking/

Table 5 Detailed COSMIN scores and ratings

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
No.1	The manchester-oxford foot questionnaire (MOXFQ)																				
	Mousavian, Ebrahimzadeh, Birjandinejad, Omidi-Kashani, Kachooei [28]	1	?	1	+							1	?								
	Garcés, Winson, Goldhahn, Castro, Swords, Grujic, Rammelt, Sands [29]	1	?	3	+	3	+					4	+					3	+		
	Dawson, Boller, Doll, Lavis, Sharp, Cooke, Jenkinson [30]					1	?													1	
	Morley, Jenkinson, Doll, Lavis, Sharp, Cooke, Dawson [31]			1	+							1	?	1	?						
	Dawson, Boller, Doll, Lavis, Sharp, Cooke, Jenkinson [32]																	1	?		
	Maher, Kilmartin [33]																	1	?	1	
	Dawson, Boller, Doll, Lavis, Sharp, Cooke, Jenkinson [34]	1	?	1	+							2	+	2	+						
	Dawson, Coffey, Doll, Lavis, Cooke, Herron, Jenkinson [35]	3	+							3	+	2	+	2	+			2	+		
	Marinozzi, Martinelli, Panasci, Cancilleri, Franceschetti, Vincenzi, Di Martino, Denaro [36]	1	?	1	+							4	−								
	Dawson, Doll, Coffey, Jenkinson [37]																	1	?	1	
No.2	The Manchester Foot Pain and Disability Index (MFPDI)																				

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	Van Der Zwaard, Terwee, Roddy, Terluin, Van Der Horst, Elders [38]	2	–	2	–	2	?					2	+					2	–		
	Menz, Auhl, Ristevski, Frescos, Munteanu [39]																	1	?		
	Gijon-Nogueron, Ndosi, Luque-Suarez, Alcacer-Pitarch, Munuera, Garrow, Redmond [40]																				
	Pedersen, Danneskiold-Samsoe, Garrow, Waehrens, Bliddal, Christensen, Bartels [41]			1	?							1	?	1	?						
	Roddy, Muller, Thomas [42]	2	+	1	–					2	?					1	?				
	Kaoulla, Frescos, Menz [43]	2	–							2	+	1	?								
	Menz, Tiedemann, Kwan, Plumb, Lord [44]	2	–							2	+	1	?								
	Garrow, Papageorgiou, Silman, Thomas, Jayson, Macfarlane [45]	1	?							2	+	1	?			1	?				
No.3	The Foot Function Index (FFI)																				
	Goldstein, Schemitsch, Bhandari, Mathew, Petrisor [61]											1	?								
	Pinsker, Inrig, Daniels, Warmington, Beaton [46]	1	?	1	+	1	?					2	+							1	
	Budiman-Mak, Conrad, Roach [47]	1	–	1	-					2	?	1	?	1	?			1	?		
	Kuyvenhoven, Gorter, Zuithoff, Zuithoff, Budiman-Mak, Conrad, Post [48]	2	–	1	+					2	+							1	?		

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	Agel, Beskin, Brage, Guyton, Kadel, Saltzman, Sands, Sangeorzan, SooHoo, Stroud, Thordarson [49]			1	?																
	SooHoo, Samimi, Vyas, Botzler, Botzler [50]											1	?								
	SooHoo, Vyas, Samimi [51]																	1	?		
	Naal, Impellizzeri, Huber, Rippstein [52]	1	?	1	+	1	?					1	?	1	?						
	Wu, Liang, Hou [53]	1	?	1	+							1	?								
	Pourtier-Piotte, Pereira, Soubrier, Thomas, Gerbaud, Coudeyre [54]	1	−									1	?					1	?		
	Madeley, Wing, Topliss, Penner, Glazebrook, Younger [55]																	1	?		
	Paez-Moguer, Budiman-Mak, Cuesta-Vargas [56]	1	-									1	?								
	Martinelli, Scotto, Sartorelli, Bonifacini, Bianchi, Malerba [57]	1	?	1	+	1	?					2	+	2	−			1	?		
	Jorgensen, Andreasen, Rathleff [58]	1	?	1	+																
	Venditto, Tognolo, Rizzo, Iannuccelli, Di Sante, Trevisan, Maggiolini, Santilli, Ioppolo [59]			1	+							1	?					1	?		
	Saag, Saltzman, Brown, Budiman-Mak [60]			1	1																
No.4	The foot and ankle ability measure (FAAM)																				
	Goldstein, Schemitsch, Bhandari, Mathew, Petrisor [61]											1	?								
	Martin, Irrgang, Burdett, Conti, Van Swearingen [62]	4	−	3	+	3	+			4	+	1	+	1	+			1	?	1	

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	Carcia, Martin, Drouin [63]											1	?			1	?				
	Martin, Hutt, Wukich [64]											1	+	1	+	1	?				
	Mazaheri, Salavati, Negahban, Sohani, Taghizadeh, Feizi, Karimi, Parnianpour [65]	1	?	1	+	1	?					3	−	3	+	1	?				
	Kivlan, Martin, Wukich [66]																	1	+	1	
	Arunakul, Arunakul, Suesiritumrong, Angthong, Chernchujit [67]	1	?	1	+							1	?								
	Weel, Zwiers, Azim, Sierevelt, Haverkamp, van Dijk, Kerkhoffs [68]	1	?	1	+	1	+					2	−	2	+						
	Uematsu, Suzuki, Sasaki, Nagano, Shinozuka, Sunagawa, Fukubayashi [69]	1	?	1	?	1	?					1	?	1	?						
No.5	The foot and ankle outcome score (FAOS)																				
	Roos, Brandsson, Karlsson [70]	2	−	1	+			2	?	3	+	2	+	1	?						
	Karatepe, Gunaydin, Kaya, Karlibas, Ozbek [71]	1	?	1	+							1	?								
	Negahban, Mazaheri, Salavati, Sohani, Askari, Fanian, Parnianpour [72]	1	?	1	+	1	?					2	−								
	Golightly, DeVellis, Roos, Lohmander, Hannan, Nelson, Jordan [73]	4	−	1	+					4	?	1	?								
	van den Akker-Scheek, Seldentuis, Reininga, Stevens [74]	1	?									1	?					1	?		

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	Lee, Chung, Kwon, Sung, Lee, Won, Lee, Lee, Park [75]	2	–	1	+					2	+	1	?								
	Mani, Brown, Nair, Chen, Do, Lyman, Deland, Ellis [76]			2	+			1	?			2	+	2	+			1	?		
	van Bergen, Sierevelt, Hoogervorst, Waizy, van Dijk, Becher [77]	1	?	1	+	1	?					2	+								
	Golightly, Devellis, Nelson, Hannan, Lohmander, Renner, Jordan [78]	4	–	1	–							1	?								
	Angthong [79]	1	?	2	–							1	?								
	Sierevelt, Beimers, van Bergen, Haverkamp, Terwee, Kerkhoffs [80]	1	+	1	+	1	?			1	?	3	+	3	+						
	Sierevelt, van Eekeren, Haverkamp, Reilingh, Terwee, Kerkhoffs [81]																	1	?	1	
	Mani, Do, Vulcano, Hogan, Lyman, Deland, Ellis [82]	1	?									2	–	2	+			1	?		
	Chen, Lyman, Do, Karlsson, Adam, Young, Deland, Ellis [83]	1	?	1	+							2	+	2	+			1	?		
No.6	The foot health status questionnaire (FHSQ)																				
	Menz, Auhl, Ristevski, Frescos, Munteanu [39]																	1	?		
	Cuesta-Vargas, Bennett, Jimenez-Cebrian, Labajos-Manzanares [84]	1	?	1	+					2	+	1	?								
	Bennett, Patterson, Wearing, Baglioni [85]	1	?	1	+					2	+	1	?	1	?						
No.7	The Cumberland Ankle Instability Tool (CAIT)																				
	Hiller, Refshauge, Bundy, Herbert, Kilbreath [86]	3	–	1	+							1	?								

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	De Noronha, Refshauge, Kilbreath, Figueiredo [87]	1	?	1	+													1	?		
	Cruz-Diaz, Hita-Contreras, Lomas-Vega, Osuna-Perez, Martinez-Amat [88]	1	?	1	+					2	+	1	?					1	?		
	Rodriguez-Fernandez, Rebollo-Roldan, Jimenez-Rejano, Gueita-Rodriguez [89]	1	?	1	+													1	?		
	Ko, Rosen, Brown [90]	1	?	1	+					4	+	1	?								
No.8	The Short Musculoskeletal Function Assessment Questionnaire-Dysfunction Index (SMFAQ-DI)																				
	Goldstein, Schemitsch, Bhandari, Mathew, Petrisor [61]											1	?								
	Pinsker, Inrig, Daniels, Warmington, Beaton [46]	1	?	1	+	1	?					1	?								
	Wang, He, Lei, Lin, Li, Wang, Zhai, Xu, Zhang, Lin [91]	4	–							4	+	1	?			1	?				
No.9	The Short Musculoskeletal Function Assessment Questionnaire-Bother Index (SMFAQ-BI)																				
	Goldstein, Schemitsch, Bhandari, Mathew, Petrisor [61]											1	?								
	Wang, He, Lei, Lin, Li, Wang, Zhai, Xu, Zhang, Lin [91]	4	-							4	+	1	?			1	?				
No.10	The ankle osteoarthritis scale (AOS)																				

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	Pinsker, Inrig, Daniels, Warmington, Beaton [46]	1	?	1	+	1	?					1	?								
	Madeley, Wing, Topliss, Penner, Glazebrook, Younger [55]																	1	?		
	Domsic, Saltzman [92]			1	1							1	?								
	McPhail, Williams, Schuetz, Baxter, Tonks, Haines [93]																				
No.11	The Western Ontario and McMaster Universities Osteoarthritis Index (WOMUOI)																				
	Pinsker, Inrig, Daniels, Warmington, Beaton [46]	1	?	1	+	1	?					1	?								
No.12	The Juvenile Arthritis Foot Disability Index (JAFD)																				
	Andre, Hagelberg, Stenstrom [94]	1	?	1	+			1	+			1	?			1	?				
	Esbjörnsson, Iversen, Broström, Hagelberg, André [95]																	1	+		
No.13	The oxford ankle foot questionnaire for children (OxAFQ)																				
	Morris, Doll, Wainwright, Theologis, Fitzpatrick [96]			1	?							1	?			1	?				
	Morris, Doll, Davies, Wainwright, Theologis, Willett, Fitzpatrick [97]			1	+	1	+											1	?	1	
	Martinkevich, Moller-Madsen, Gottliebsen, Kjeldgaard, Rahbek [98]	1	?	1	−	1	-					3	−								
	Martinelli, Romeo, Bonifacini, Viganò, Bianchi, Malerba [99]	1	?	1	+							3	−	3	+			1	?		

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
No.14	The self-administered foot evaluation questionnaire (SAFE-Q)																				
	Niki, Tatsunami, Haraguchi, Aoki, Okuda, Suda, Takao, Tanaka [100]			2	+							1	?			1	?				
	Yano, Ikari, Ochi, Ishida, Sakuma, Yoshida, Koyama, Koenuma, Momohara [101]											1	?					1	?		
No.15	The self-reported foot and ankle score (SEFAS)																				
	Coster, Karlsson, Nilsson, Carlsson [102]	1	?	1	+	1	?					2	+	2	+			1	?		
	Coster, Rosengren, Bremander, Brudin, Karlsson 103]			1	+	1	+											1	+		
	Coster, Bremander, Rosengren, Magnusson, Carlsson, Karlsson [104]			1	+	1	?	1	?			2	+	2	+			1	?		
No.16	ROwan foot pain assessment questionnaire (ROFPAQ)																				
	Rowan [105]	1	?	1	+			1	+			1	+	1	+						
No.17	Foot impact scale for rheumatoid arthritis (FISRA)																				
	Woodburn, Vliet, van der Leeden, Steultjens [106]	1	−																		
	Woodburn, Turner, Rosenbaum, Balint, Korda, Ormos, Szabo, Vliet Vlieland, van der Leeden, Steultjens [107]	2	−																		
No.18	Diabetic foot ulcer scale (DFUS)																				

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	Abetz, Sutton, Brady, McNulty, Gagnon [108]	1	?	1	?			1	+			1	?			1	?	1	?		
No.19	Diabetic Foot Ulcer Scale-Short Form (DFUS-SF)																				
	Bann, Fehnel, Gagnon [109]	3	+	2	−					3	?	3	+					1	?		
	Hui, Yee-Tak, Yam, Yuk Ip [110]	1	?									3	−								
No.20	Diabetes foot self-care behavior scale (DFSCBS)																				
	Chin, Huang [111]	2	+	2	+			1	?	2	−	1	?	1	?						
No.21	The American orthopaedic foot and ankle society clinical rating scale-ankle/hindfoot (AOFAS-AH)																				
	Goldstein, Schemitsch, Bhandari, Mathew, Petrisor [61]											1	?								
	Pinsker, Inrig, Daniels, Warmington, Beaton [46]	1	?	1	+	1	?					1	?							1	
	SooHoo, Vyas, Samimi [51]																	1	?		
	Pena, Agel, Coetzee [112]																	1	?	1	
	Ibrahim, Beiri, Azzabi, Best, Taylor, Menon [113]			1	?							1	?					1	?		
	Madeley, Wing, Topliss, Penner, Glazebrook, Younger [55]																	1	?		
	Coster, Rosengren, Bremander, Brudin, Karlsson [103]			1	+	1	+											1	+		

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
No.22	The American orthopaedic foot and ankle society clinical rating scale-midfoot (AOFAS-M)																				
	SooHoo, Vyas, Samimi [51]																	1	?		
	Ibrahim, Beiri, Azzabi, Best, Taylor, Menon [113]			1	?							1	?					1	?		
	Coster, Rosengren, Bremander, Brudin, Karlsson [103]			1	+	1	+											1	+		
No.23	The American orthopaedic foot and ankle society clinical rating scale-hallux metatarsophalangeal-interphalangeal Joints (AOFAS-HJ)																				
	Baumhauer, Nawoczenski, DiGiovanni, Wilding [114]	1	?	1	+							1	?								
	SooHoo, Vyas, Samimi [51]																	1	?		
	Ibrahim, Beiri, Azzabi, Best, Taylor, Menon [113]			1	?							1	?					1	?		
	Coster, Rosengren, Bremander, Brudin, Karlsson [103]			1	+	1	+											1	+		
No.24	The American orthopaedic foot and ankle society clinical rating scale-lesser metatarsophalangeal-interphalangeal Joints (AOFAS-LJ)																				
	Baumhauer, Nawoczenski, DiGiovanni, Wilding [114]	1	?	1	+							1	?								

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	SooHoo, Vyas, Samimi [51]																	1	?		
	Ibrahim, Beiri, Azzabi, Best, Taylor, Menon [113]			1	?							1	?					1	?		
	Coster, Rosengren, Bremander, Brudin, Karlsson [103]			1	+	1	+											1	+		
No.25	The Foot and Ankle Disability Index (FADI)																				
	Hale, Hertel [115]																	1	?		
No.26	The Foot and Ankle Disability Index Sport (FADI Sport)																				
	Hale, Hertel [115]			1	+													1	?		
No.27	The American orthopaedic foot and ankle society diabetic foot questionnaire (AOFAS-DFQ)																				
	Dhawan, Spratt, Pinzur, Baumhauer, Rudicel, Saltzman [117]	1	?	1	−	1	?														
No.28	The American academy of orthopaedic surgeons foot and ankle questionnaire (AAOS-FAQ)																				
	Goldstein, Schemitsch, Bhandari, Mathew, Petrisor [61]											1	?								
	Kim, Kim, Seo, Lee [118]			1	+							1	?					1	?		
	Boszczyk, Blonski, Pomianowski [119]	1	?	1	+							1	?								
No.29	The sports athlete foot and ankle score (SAFAS)																				
	Morssinkhof, Wang, James, van der Heide, Winson [120]	1	?					1	?			2	+	2	−						

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
No.30	The hand–foot syndrome specific quality of life questionnaire (HFS-14)		+/−																		
	Sibaud, Dalenc, Chevreau, Roché, Delord, Mourey, Lacaze, Rahhali, Taieb [121]	1	?									1	?								
	Mikoshiba, Yamamoto-Mitani, Sato, Asaoka, Ohki, Ohata, Miyashita [122]	4	−	1	+							1	?	1	?	1	?				
No.31	The neuropathy and foot ulcer specific quality of life instrument (NeuroQol)																				
	Vileikyte, Peyrot, Bundy, Rubin, Leventhal, Mora, Shaw, Baker, Boulton [123]	2	−							2	?	2	−								
	Xavier, Foss, Marques Junior, dos Santos, Onofre, Pace [124]	1	?									1	?			1	?				
No.32	The Japanese society of surgery of the foot standard rating system (JSSF standard rating system)																				
	Niki, Aoki, Inokuchi, Ozeki, Kinoshita, Kura, Tanaka, Noguchi, Nomura, Hatori, Tatsunami [116]			1	?																
No.33	The visual analogue scale-foot and ankle (VAS-FA)																				
	Nair, Shamsuddin, John, Hamalainen, Kurien [125]											1	?								

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	Angthong, Chernchujit, Suntharapa, Harnroongroj [126]	1	?	1	+							1	?								
No.34	The cardiff wound impact schedule (CWIS)																				
	Jaksa, Mahoney [127]			1	?							1	?								
No.35	The musculoskeletal functional assessment (MFA)																				
	Pena, Agel, Coetzee [112]																	1	?		
No.36	The questionnaire for usability evaluation of orthopaedic shoes pre-test (QUEOS Pre-test)																				
	Jannink, de Vries, Stewart, Groothoff, Lankhorst [128]	1	?	1	−																
No.37	The questionnaire for usability evaluation of orthopaedic shoes post-test (QUEOS Post-test)																				
	Jannink, de Vries, Stewart, Groothoff, Lankhorst [128]	1	?	1	−																
No.38	The oswestry disability questionnaire (ODQ)																				
	Ferrari [129]																	1	?		
No.39	American college of foot and ankle surgeons scoring scales (ACFAS scoring scale)																				
	Cook, Cook, Rosenblum, Landsman, Roukis, Roukis [130]			1	?			1	?									1	?		
No.40	The Olerud-Molander Ankle Score (OMAS)																				
	Nilsson, Eneroth, Ekdahl [131]	1	?	1	+	1	?					1	?								

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
No.41	The patient reported outcome measurement information system physical function computerized adaptive tests (PROMIS PF CAT)																				
	Hung, Baumhauer, Latt, Saltzman, SooHoo, Hunt [132]	4	+			1	?					1	?			1	+				
No.42	The hand–foot and mucositis symptom and impact questionnaire (HAMSIQ)																				
	Lai, Beaumont, Diaz, Khan, Cella [133]	1	?									2	−			1	+	1	?		
No.43	The hand-foot skin reaction and quality of life questionnaire (HF-Qol)																				
	Anderson, Keating, Doll, Camacho [134]	1	?	1	+					2	?	1	−			1	+	1	?	1	
No.44	The combined foot care confidence scale/foot-care behavior instrument (FCCS-FCB)																				
	Garcia-Inzunza, Valles-Medina, Munoz, Delgadillo-Ramos, Compean-Ortiz [135]	1	?							2	−										
No.45	The salford rheumatoid arthritis foot evaluation part A (SAFE-Part A)																				
	Walmsley, Ravey, Graham, Teh, Williams [136]			1	+							1	?								
No.46	The diabetic foot self-care questionnaire of the university of Malaga (DFSQ-UMA)																				

Table 5 (continued)

		Internal consistency		Reliability		Measurement error		Content validity		Structural validity		Construct validity						Responsiveness		Interpretability	
		Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Rating	Score	Score	Score	Rating	Score	Rating	Score	Rating
	Navarro-Flores, Morales-Asencio, Cervera-Marin, Labajos-Manzanares, Gijon-Nogueron [137]	1	?	1	+					2	+	1	?								
No.47	The questionnaire for diabetes-related foot disease (Q-DFD)																				
	Castillo-Tandazo, Flores-Fortty, Feraud, Tettamanti [138]			1	+							1	?								
No.48	The outcome instrument for the Foot and ankle version 2 (OIFA-2)																				
	Niki, Tatsunami, Haraguchi, Aoki, Okuda, Suda, Takao, Tanaka [139]	2	−							2	+	1	?								
No.49	The chronic ankle instability scale (CAIS)																				
	Eechaute, Vaes, Duquet [140]	1	?	1	+	1	+	1	?			1	+							1	
No.50	The telephone questionnaire																				
	Wulterkens, Aurégan, Letellier, Mebtouche, Levante, Cottin, Bégué [141]	1	?	1	+							1	?								

standing, pain and social interaction domains were 0.92, 0.86, and 0.73, respectively [35].

Reliability: Moderate positive evidence was found for the test-retest reliability in one study with good quality; the intraclass correlation coefficient (ICC) values for the pain, walking/standing and social interaction domains were 0.95, 0.97, and 0.96, respectively [29].

Measurement error: Moderate positive evidence was found in one study that was of good quality [29]. The standard error of the mean (SEM) values were 9.70, 7.49, and 13.97 for the pain, walking/standing, and social interaction domains, respectively, which were smaller than the MDC values of 22.64, 17.47, and 32.59, respectively.

Structural validity: Moderate positive evidence was found in one study with good quality; the three-factor structure explained 65.8% of the total variance [35].

Hypothesis validity: For construct validity, strong but contradictory evidence was found in two studies with excellent quality. In one study using the American Orthopaedic Foot and Ankle Society (AOFAS)—Hindfoot Scale and SF-36 as comparators, only one hypothesis was rejected among 11 in total [29], while in the other study with SF-36 as the comparator, 2 of 5 hypotheses were rejected [36]. Two other studies supported positive construct validity with fair evidence [34, 35]. For discriminant validity, there was moderate positive evidence in two studies with fair quality, in which 4 and 6 hypotheses were all confirmed with the Medical Outcomes Survey (MOS) Short Form—36 (SF-36) as the comparator [34, 35]. There was no evidence for discriminative validity.

Responsiveness: Moderate positive evidence was found in one study with good quality showing the change of the MOXFQ scores could reflect the direction of the transition item [29] and in another study with fair quality showing that the hypothesis related to effect sizes was confirmed [35].

Interpretability: In one study anchor-based MDC values were 13 points for each of the MOXFQ walking/standing, pain, and social interaction domains, while distribution-based MDC values were 11, 12, and 16 score points, respectively [30]. In two other studies, the MIC values were found to be 16 for walking, 12 for pain, and 24 for social interaction [33, 37]. No ceiling or floor effects were observed [28, 29].

Manchester Foot Pain and Disability Index (MFPDI)

Eight studies evaluated the psychometric properties of the MFPDI [38–45]. Due to poor quality, there was no evidence on measurement error, criterion validity or content validity.

Internal consistency: Moderate negative evidence was found in three studies in which unidimensionality was rejected [38, 43, 44], while one study supported the unidimensionality of all the domains with Cronbach's $\alpha > 0.70$ [42].

Reliability: Limited negative evidence was found in one study with fair quality; none of the ICCs exceeded 0.7 for all the domains [38].

Structural validity: Moderate positive evidence was found in three studies with fair quality, with 60.8, 62, and 57% of the total variance explained by a four-factor structure [43–45].

Hypothesis validity: For convergent validity, limited positive evidence was found in one study in which 7 hypotheses were all confirmed using the MOS SF-12 as the comparator [38]. There was no evidence for discriminant validity or discriminant validity.

Responsiveness: Limited negative evidence was found in one study [38]. Only one of 7 hypotheses was confirmed with five comparators.

Interpretability: No ceiling or floor effect was observed [38].

Foot Function Index (FFI)

Sixteen studies evaluated the psychometric properties of the FFI [46–61]. Due to poor quality, no evidence was found on reliability, measurement error, content validity, criterion validity, or responsiveness.

Internal consistency: Limited negative evidence was found in one study with fair quality showing the dimensionality was denied [48].

Structural validity: Limited positive evidence was found in one study with fair quality showing that a two-factor model could explain 65% of the total variance [48].

Hypothesis validity: For convergent validity, moderate positive evidence was found in one study with fair quality showing that high correlation was found between similar domains between the FFI and SF-36 [46] and another study with fair quality in which 6 hypotheses were all confirmed [56]. For discriminant validity, limited negative evidence was found in one study with fair quality in which 3 of 10 hypotheses were rejected [56]. There was no evidence regarding the discriminative validity.

Interpretability: The MDCs of the pain, disability, and activity limitation were 28.36, 19.60, and 21.72, respectively [46]. Five studies explored the floor and ceiling effect [46, 53, 56, 58, 60]. The domain of the activity limitation was found to have floor effect in two studies [46, 53] and ceiling effect in one study [46].

Foot and ankle ability measure (FAAM)

Nine studies evaluated the psychometric properties of the FAAM [61–69]. Due to poor quality, no evidence on

content validity, criterion validity, or responsiveness was found.

Internal consistency: Strong negative evidence was found in one study with excellent quality showing the unidimensionality was rejected in a group of patients [62].

Reliability: Moderate positive evidence was found in one study with good quality. The ICCs for the activities of daily living (ADL) and sports domains were 0.89 and 0.87, respectively [62].

Measurement error: Moderate positive evidence was found in one study with good quality showing the SEMs for the ADL and sports domains were 5.7 and 12.3, respectively [62].

Structural validity: Strong positive evidence was found in one study with excellent quality showing that more than 80% of total variance could be explained by either two- or three-factor model [62].

Hypothesis validity: For convergent validity, moderate negative evidence was found in one study with good quality showing 2 of 6 hypotheses were rejected using SF-36 as the comparator [65] and in another study with fair quality showing that 5 of 10 hypotheses were rejected using the Foot and Ankle Outcome Score (FAOS), Numeric Rating Scale (NRS) for pain and SF-36 as comparators [68]. For discriminant validity, one study with good quality had moderate positive evidence, showing that 4 hypotheses were all confirmed [65] and another study with fair quality showed that 6 hypotheses were all confirmed [68]. There was no evidence regarding the discriminative validity.

Interpretability: The MDCs were 8 and 9 for the ADL and sports domains, respectively [62]. No floor or ceiling effect was found [65, 68].

Foot and ankle outcome score (FAOS)

Fourteen studies evaluated the psychometric properties of the FAOS [70–83]. Due to poor quality, there was no evidence on the measurement error, content validity or responsiveness.

Internal consistency: Strong negative evidence was found in two studies with excellent quality [73, 78] and two studies with fair quality [70, 75] in which unidimensionality was rejected.

Reliability: Limited but conflicting evidence was found in two studies. One study with fair quality provided positive evidence that the ICCs for pain, symptoms, daily activities, sports/recreation, and quality of life were 0.847, 0.787, 0.858, 0.876, and 0.787, respectively [76]. The other study with fair quality provided negative evidence showing that the ICCs for the same domains were only 0.158, 0.428, 0.330, 0.100, and 0.057, respectively [79].

Structural validity: Moderate positive evidence was found in one study with good quality showing the model could explain more than 50% of total variance [70], and in another study with fair quality showing a seven-factor model could explain 69.83% of total variance [75].

Hypothesis validity: For convergent validity, moderate positive evidence was found in 7 studies. One study with good quality provided positive evidence that 35 of 45 hypotheses were confirmed using SF-36, AOFAS, and VAS as the comparators [80]. In four studies with fair quality providing positive evidence, 1 of 1 [70], 15 of 20 [76], 4 of 5 [77], and 16 of 20 [83] hypotheses were confirmed, respectively. On the other hand, two studies provided negative evidence with fair quality, in which only 1 of 10 [71] and 11 of 20 [83] hypotheses were confirmed, respectively. For discriminant validity, positive evidence was found in one study with good quality showing that 20 hypotheses were all confirmed using SF-36, AOFAS, and VAS as the comparators [80]. Meanwhile, positive evidence was also found in three studies with fair quality in which 15 of 20 [76], 15 of 20 [82], and 16 of 20 [83] hypotheses were confirmed. There was no evidence regarding the discriminative validity.

Interpretability: The MICs for the pain, symptoms, daily activities, sports/recreation, and quality of life domains were 7.1/12.5, 15.3/12.5, 17.6/13.9, 22.5/32.5, and 21.9/21.8, respectively, using different anchors in one study [81]. The sports/recreation domain was found to have some ceiling and floor effects in four studies [70, 79, 82, 83].

Foot health status questionnaire (FHSQ)

Three studies evaluated the psychometric properties of the FHSQ [39, 84, 85]. Due to poor quality, no evidence was found on internal consistency, reliability, measurement error, content validity, hypothesis validity, responsiveness, or interpretability. Two studies with fair quality provided positive evidence regarding structural validity, in which 86% [84] and 84% [85] of total variance was explained by a four-factor model.

Cumberland ankle instability tool (CAIT)

The psychometric properties of the CAIT were evaluated by five studies [86–90]. Due to poor quality, no evidence on reliability, measurement error, content validity, hypothesis validity, or responsiveness was found.

Internal consistency: Moderate negative evidence was found in one study with good quality showing unidimensionality was rejected by the Rasch model [86].

Structural validity: In one study with fair quality, 66.4% of total variance was explained by a three-factor model [88]. In another study, 74.4% of total variance could be explained by a two-factor model [90].

Interpretability: No ceiling or floor effects were observed [87, 89]. There was no evidence on the MIC or MDC.

Short Musculoskeletal Function Assessment Questionnaire—Dysfunction Index (SMFAQ-DI)

Three studies evaluated the psychometric properties of the SMFAQ-DI [46, 61, 91]. Due to poor quality, we did not find any evidence regarding the reliability, measurement error, content validity, structural validity, hypothesis validity, or responsiveness.

Internal consistency: Strong negative evidence was found in one study with excellent quality showing that the unidimensionality was rejected [91].

Interpretability: No floor or ceiling effect was observed [46, 91]. MIC or MDC was not explored.

Short Musculoskeletal Function Assessment Questionnaire—Bother Index (SMFAQ-BI)

There were two studies evaluating the psychometric properties of the SMFAQ [61, 91]. Due to poor quality, we found no evidence regarding reliability, measurement error, content validity, structural validity, hypothesis validity, or responsiveness.

Internal consistency: Strong positive evidence was found in one study with excellent quality showing that Cronbach's α for this unidimensional domain was 0.953 [91].

Interpretability: No floor or ceiling effect was observed [61, 91]. MIC or MDC was not explored.

Ankle osteoarthritis scale (AOS)

Four studies examined the psychometric properties of the AOS [46, 55, 92, 93]. Due to poor quality, there was no evidence on internal consistency, reliability, measurement error, content validity, hypothesis validity or responsiveness. No ceiling or floor effects were observed [46].

Western Ontario and McMaster Universities Osteoarthritis Index (WOMUOI)

One study evaluated the psychometric properties of the WOMUOI [46]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, content validity, hypothesis validity, or responsiveness. Floor effect was observed in the pain domain [46].

Juvenile Arthritis Foot Disability Index (JAFD)

There were two studies evaluating the psychometric properties of the JAFD [94, 95]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, content validity, hypothesis validity or responsiveness. Floor effect was observed in the participation restriction domain [94].

Oxford ankle foot questionnaire for children (OxAFQ)

Four studies evaluated the psychometric properties of the OxAFQ [96–99]. Due to poor quality, there was no evidence on internal consistency, reliability, measurement error or responsiveness.

Children's version

Hypothesis validity: For convergent validity, there was strong negative evidence in two studies with good quality. In one study using the Child Health Questionnaire (CHQ) as the comparator, 3 of 7 hypotheses were rejected [98]. In another study also using CHQ, 5 of 9 hypotheses were rejected [99]. For discriminant validity, moderate positive evidence was found in one study with good quality showing that all 24 hypotheses were confirmed [99]. There was no evidence regarding the discriminative validity.

Interpretability: The MIDs for the physical, school and play, and emotional domains were 8.8, 8.3, and 7.3, respectively [97]. Ceiling effects were found in the school and play and emotional domains [98].

Parents' version

Hypothesis validity: For convergent validity, there was moderate but conflicting evidence found in two studies with good quality. In one study using the CHQ as the comparator, only 1 hypothesis in 7 was rejected [98]. In another study also using CHQ as the comparator, 5 of 9 hypotheses were rejected [99]. For discriminant validity, moderate positive evidence was found in one study with good quality showing that all 27 hypotheses were confirmed [99]. There was no evidence regarding the discriminative validity.

Interpretability: The MIDs for the physical, school and play, and emotional domains were 19, 9.9, and 9.7, respectively [97]. Ceiling effects were found in the school and play and emotional domains [98].

Self-administered foot evaluation questionnaire (SAFE-Q)

Two studies evaluated the psychometric properties of the SAFE-Q [100, 101]. Due to poor quality, we did not find

evidence on measurement error, content validity, structural validity, hypothesis validity, responsiveness, or interpretability. Limited positive evidence on reliability was found in one study with fair quality showing that the ICCs for the pain and pain-related, physical functioning and daily living, social functioning, shoe-related, and general health and well-being domains were 0.78, 0.83, 0.72, 0.81, and 0.82, respectively [100].

Self-reported foot and ankle score (SEFAS)

Psychometric properties of the SEFAS were evaluated by three studies [102–104]. Due to poor quality, we did not find evidence on reliability, measurement error, content validity, structural validity, and responsiveness.

Hypothesis validity: For convergent validity, moderate positive evidence was found in two studies with fair quality using SF-36 and FAOS as the comparators in which 1 of 5 and 0 of 5 hypotheses, respectively, were rejected [102, 104]. For discriminant validity, moderate positive evidence was found in two studies with fair quality that both confirmed all 3 hypotheses [102, 104].

Interpretability: Ceiling or floor effects were not found [102, 104].

Rowan foot pain assessment questionnaire (ROFPAQ)

One study evaluated the psychometric properties of the ROFPAQ [105]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness or interpretability.

Foot impact scale for rheumatoid arthritis (FISRA)

Two studies evaluated the psychometric properties of the FISRA [106, 107]. Due to poor quality, we did not find evidence on reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness or interpretability. Limited negative evidence on the internal consistency was found in one study with fair quality showing that unidimensionality was denied [107].

Diabetic Foot Ulcer Scale (DFUS)

One study evaluated the psychometric properties of the DFUS [108]. Due to poor quality, no evidence was found for internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness, or interpretability.

Diabetic Foot Ulcer Scale—short form (DFUS-SF)

There were two studies evaluating the psychometric properties of the DFUS-SF [109, 110]. Due to poor quality, there was no evidence on measurement error, content validity, structural validity, or responsiveness.

Internal consistency: Moderate positive evidence was found in one study with good quality [109]. The Cronbach's α for the leisure, physical health, dependence/daily life, negative emotions, worried about ulcers/feet and bothered by ulcer care domains were 0.90, 0.86, 0.86, 0.92, 0.83, and 0.74, respectively.

Reliability: Limited negative evidence was found in one study with fair quality showing that only the ICC of the dependence/daily life domain was above 0.7 [109].

Hypothesis validity: For convergent validity, there was moderate but conflicting evidence in both studies, with good quality [109, 110]. In one study, the only explicit hypothesis was confirmed [109]. In another study, 2 of 3 hypotheses were rejected [110]. There was no evidence on the discriminant or discriminative validity.

Interpretability: Three domains—dependence/daily life, negative emotions, and bothered by ulcer care—were found to have ceiling effects [110]. No evidence on the MIC or MDC was found.

Diabetes Foot Self-care Behavior Scale (DFSCBS)

The psychometric properties of the DFSCBS were evaluated by one study [111]. Due to poor quality, no evidence was found on internal consistency, measurement error, content validity, hypothesis validity, responsiveness, or interpretability.

Internal consistency: Limited positive evidence was found in one study with fair quality showing that the Cronbach's α was 0.73 for the unidimensional instrument [111].

Reliability: Limited positive evidence was found in one study with fair quality; the ICC was 0.92 for the instrument [111].

Structural validity: Limited negative evidence was found in one study with fair quality showing that only 39% of total variance could be explained [111].

American Orthopaedic Foot and Ankle Society Clinical Rating Scale—ankle/hindfoot (AOFAS-AH)

Seven studies evaluated the psychometric properties of the AOFAS-AH [46, 51, 55, 61, 103, 112, 113]. Due to poor quality, we found no evidence on internal consistency, reliability, measurement error, content validity, structural

validity, hypothesis validity, and responsiveness. Mild ceiling effect was found [112].

American Orthopaedic Foot and Ankle Society Clinical Rating Scale—midfoot (AOFAS-M)

The psychometric properties of the AOFAS-M were evaluated by three studies [51, 103, 113]. Due to poor quality, evidence was lacking on internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness, and interpretability.

American Orthopaedic Foot and Ankle Society Clinical Rating Scale—hallux metatarsophalangeal–interphalangeal joints (AOFAS-HJ)

Four studies evaluated the psychometric properties of the AOFAS-HJ [51, 103, 113, 114]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness, or interpretability.

American Orthopaedic Foot and Ankle Society Clinical Rating Scale—lesser metatarsophalangeal–interphalangeal joints (AOFAS-LJ)

The psychometric properties of the AOFAS-LJ were evaluated by four studies [51, 103, 113, 114]. Due to poor quality, no evidence was found on internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness, or interpretability.

Foot and Ankle Disability Index (FADI)

One study evaluated the psychometric properties of the FADI [115]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness, or interpretability.

Foot and Ankle Disability Index sport (FADI Sport)

The psychometric properties of the FADI Sport were evaluated by only one study [115]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness, or interpretability.

American orthopaedic foot and ankle society diabetic foot questionnaire (AOFAS-DFQ)

There was one study evaluating the psychometric properties of the AOFAS-DFQ [117]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, or responsiveness. The medical care and self-care domains suffered from low ceiling effects, whereas the role emotional, role emotional (SF-36), ADL, role physical, and role physical (SF-36) domains all demonstrated high floor effects [117].

American academy of orthopaedic surgeons foot and ankle questionnaire (AAOS-FAQ)

Three studies evaluated the AAOS-FAQ [61, 118, 119]. Due to poor quality, no evidence was found on internal consistency, reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness, or interpretability.

Sports athlete foot and ankle score (SAFAS)

One study evaluated the psychometric properties of the SAFAS [120]. We did not find evidence on internal consistency, reliability, measurement error, content validity, structural validity, responsiveness or interpretability. For convergent validity, limited positive evidence and limited negative evidence were found in a study with fair quality [120]. No evidence on the discriminative validity was found.

Hand–foot syndrome specific quality of life questionnaire (HFS-14)

There were two studies evaluating the HFS-14 [121, 122]. Due to poor quality, evidence was lacking on reliability, measurement error, content validity, structural validity, hypothesis validity, responsiveness, and interpretability. Strong negative evidence was found in one study with excellent quality showing that unidimensionality was rejected [122].

Neuropathy and foot ulcer specific quality of life instrument (NeuroQol)

The NeuroQol was evaluated by two studies [123, 124]. Owing to poor quality, no evidence was found on reliability, measurement error, structural validity, or responsiveness.

Internal consistency: Limited negative evidence was found in one study in which unidimensionality was rejected [123].

Hypothesis validity: For convergent validity, limited negative evidence was found in one study showing the hypothesis was rejected [123]. There was no evidence regarding the discriminant or discriminative validity.

Interpretability: One study showed modest floor effect [123]; in the other study, the domains of loss/reduction of

sensitivity and limitations in daily activities suffered from floor and ceiling effects, respectively [124].

Japanese society of surgery of the foot standard rating system (JSSF standard rating system)

One study evaluated the psychometric properties of the JSSF standard rating system [116]. No evidence was found, due to poor quality, on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, or interpretability for any of the five scales.

Visual Analogue Scale—foot and ankle (VAS-FA)

There were two studies evaluating the psychometric properties of the VAS-FA [125, 126]. Due to poor quality, we did not find evidence on the internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, or interpretability.

Cardiff wound impact schedule (CWIS)

One study evaluated the CWIS [127]. Due to poor quality, evidence on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, and interpretability was not found.

Musculoskeletal functional assessment (MFA)

One study evaluated the psychometric properties of the MFA [112]. Due to poor quality, evidence was not found on internal consistency, reliability, measurement error, structural validity, hypothesis validity, or responsiveness. Mild ceiling effects were found on each scale [112].

Questionnaire for usability evaluation of orthopaedic shoes pre-test (QUEOS Pre-test)

One study evaluated the psychometric properties of the QUEOS Pre-test [128]. We did not find evidence on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, or interpretability.

Questionnaire for usability evaluation of orthopaedic shoes post-test (QUEOS Post-test)

One study evaluated the psychometric properties of the QUEOS Post-test [128]. We did not find evidence on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, or interpretability.

Oswestry disability questionnaire (ODQ)

In the one study [129], no evidence was found on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, or interpretability, owing to poor quality.

American College of Foot and Ankle Surgeons Scoring Scales (ACFAS scoring scale)

One study evaluated the psychometric properties of the ACFAS scoring scale [130]. Evidence on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, and interpretability was lacking.

Olerud-molander ankle score (OMAS)

One study evaluated the psychometric properties of the OMAS [131]. Due to poor quality, no evidence was found on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, or interpretability.

Patient-reported outcome measurement information system physical function computerized adaptive tests (PROMIS PF CAT)

One study evaluated the PROMIS PF CAT [132], but no evidence was found on reliability, measurement error, structural validity, hypothesis validity, or responsiveness.

Internal consistency: Strong positive evidence was found that Cronbach's α values were >0.9 for all the domains [132].

Interpretability: No ceiling or floor effects were observed [132]. No evidence on the MIC/MDC was found.

Hand–foot and mucositis symptom and impact questionnaire (HAMSIQ)

The psychometric properties of the OMAS were evaluated by one study [133]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, structural validity, responsiveness, or interpretability. Limited negative evidence was found on convergent validity; only one of ten hypotheses was confirmed [133].

Hand–foot skin reaction and quality of life questionnaire (HF-QoL)

HF-QoL was evaluated by one study [134]. Due to poor quality, no evidence was found on internal consistency, reliability, measurement error, structural validity, hypothesis

validity, responsiveness, or interpretability for either symptom scale or daily activities scale.

Combined foot care confidence scale/foot-care behavior instrument (FCCS-FCB)

This instrument's psychometric properties were examined by one study [135]. No evidence was found on internal consistency, reliability, measurement error, hypothesis validity, responsiveness, or interpretability. The two-factor structure could explain only 49.1% of total variance [135].

Salford rheumatoid arthritis foot evaluation part A (SAFE-Part A)

One study evaluated the psychometric properties of SAFE-Part A [136], but no evidence was found on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness or interpretability, owing to poor quality.

Diabetic foot self-care questionnaire of the university of malaga (DFSQ-UMA)

One study evaluated the psychometric properties of the DFSQ-UMA [137]. Due to poor quality, we found no evidence on internal consistency, reliability, measurement error, hypothesis validity, or responsiveness.

Structural validity: Limited positive evidence was found showing that a three-factor structure accounted for 60.88% of total variance [137].

Interpretability: No ceiling or floor effect was observed [137].

Questionnaire for diabetes-related foot disease (Q-DFD)

One study evaluated the psychometric properties of the Q-DFD [138]. No evidence was found on internal consistency, reliability, measurement error, structural validity, hypothesis validity, responsiveness, or interpretability, owing to poor quality.

Outcome instrument for the foot and ankle version 2 (OIFA-2)

One study evaluated the psychometric properties of OIFA-2 [139]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, structural validity, hypothesis validity, or responsiveness.

Internal consistency: Limited negative evidence was found showing that unidimensionality was rejected [139].

Interpretability: No ceiling or floor effect was observed [139].

Chronic Ankle Instability Scale (CAIS)

The psychometric properties of CAIS were evaluated by one study [140]. Due to poor quality, we did not find evidence on internal consistency, reliability, measurement error, structural validity, hypothesis validity, or responsiveness.

Interpretability: No ceiling or floor effect was observed [140].

Telephone questionnaire

The Telephone Questionnaire was evaluated by one study [141]. Due to poor quality, evidence on internal consistency, reliability, measurement error, structural validity, hypothesis validity, and responsiveness was lacking.

References

1. Menz, H. B., & Lord, S. R. (2001). Foot pain impairs balance and functional ability in community-dwelling older people. *Journal of the American Podiatric Medical Association, 91*(5), 222–229.
2. Tinetti, M. E., Speechley, M., & Ginter, S. F. (1988). Risk factors for falls among elderly persons living in the community. *The New England Journal of Medicine, 319*(26), 1701–1707.
3. Benvenuti, F., Ferrucci, L., Guralnik, J. M., Gangemi, S., & Baroni, A. (1995). Foot pain and disability in older persons: an epidemiologic survey. *Journal of the American Geriatrics Society, 43*(5), 479–484.
4. Spahn, G., Schiele, R., Hell, A. K., Klinger, H. M., Jung, R., & Langlotz, A. (2004). The prevalence of pain and deformities in the feet of adolescents. Results of a cross-sectional study. *Z Orthop Ihre Grenzgeb, 142*(4), 389–396.
5. van Wyngarden, T. M. (1997). The painful foot, Part I: Common forefoot deformities. *American Family Physician, 55*(5), 1866–1876.
6. Lardenoye, S., Theunissen, E., Cleffken, B., Brink, P. R., de Bie, R. A., & Poeze, M. (2012). The effect of taping versus semi-rigid bracing on patient outcome and satisfaction in ankle sprains: a prospective, randomized controlled trial. *BMC Musculoskeletal Disorders, 13*, 81.
7. Frey, C. (2000). Foot health and shoewear for women. *Clinical Orthopaedics and Related Research, 372*, 32–44.
8. Garrow, A. P., Silman, A. J., & Macfarlane, G. J. (2004). The Cheshire Foot Pain and Disability Survey: a population survey assessing prevalence and associations. *Pain, 110*(1–2), 378–384.
9. Garratt, A., Schmidt, L., Mackintosh, A., & Fitzpatrick, R. (2002). Quality of life measurement: bibliographic study of patient assessed health outcome measures. *Bmj, 324*(7351), 1417.
10. Eechaute, C., Vaes P. F., Van Aerschot, L., Asman, S., & Duquet, W. (2007). The clinimetric qualities of patient-assessed instruments for measuring chronic ankle instability: a systematic review. *BMC Musculoskelet Disorders, 8*, 6.

11. van der Leeden, M., Terwee, C. B., Rosenbaum, D., Turner, D., Woodburn, J., Dekker, J., et al. (2008). A systematic review of instruments measuring foot function, foot pain, and foot-related disability in patients with rheumatoid arthritis. *Arthritis and Rheumatism, 59*(9), 1257–1269.
12. Hegedus, E. J., McDonough, S. M., Bleakley, C., Baxter, D., & Cook, C. E. (2015). Clinician-friendly lower extremity physical performance tests in athletes: a systematic review of measurement properties and correlation with injury. Part 2–the tests for the hip, thigh, foot and ankle including the star excursion balance test. *British Journal of Sports Medicine, 49*(10), 649–656.
13. Pike, S., Lannin, N. A., Cusick, A., Wales, K., Turner-Stokes, L., & Ashford, S. (2015). A systematic review protocol to evaluate the psychometric properties of measures of function within adult neuro-rehabilitation. *Systematic Reviews, 4*, 86.
14. Mokkink, L. B., Terwee, C. B., Stratford, P. W., Alonso, J., Patrick, D. L., Riphagen, I., et al. (2009). Evaluation of the methodological quality of systematic reviews of health status measurement instruments. *Quality of Life Research, 18*(3), 313–333.
15. Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42.
16. Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research, 21*(4), 651–657.
17. Conijn, A. P., Jens, S., Terwee, C. B., Breek, J. C., & Koelemay, M. J. (2015). Assessing the quality of available patient reported outcome measures for intermittent claudication: a systematic review using the COSMIN checklist. *European Journal of Vascular and Endovascular Surgery, 49*(3), 316–334.
18. Evans, K., Spiby, H., & Morrell, C. J. (2015). A psychometric systematic review of self-report instruments to identify anxiety in pregnancy. *Journal of Advanced Nursing, 71*(9), 1986–2001.
19. Wong, C. K., Lang, B. H., & Lam, C. L. (2016). A systematic review of quality of thyroid-specific health-related quality of life instruments recommends ThyPRO for patients with benign thyroid diseases. *J Clin Epidemiol, 78*, 63–72
20. Wong, C. K., Chen, J., Yu, C. L., Sham, M., & Lam, C. L. (2015). Systematic review recommends the European Organization for Research and Treatment of Cancer colorectal cancer-specific module for measuring quality of life in colorectal cancer patients. *Journal of Clinical Epidemiology, 68*(3), 266–278.
21. Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research, 18*(8), 1115–1123.
22. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2009). *The COSMIN checklist manual*. Amsterdam: VU University Medical Centre.
23. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research, 19*(4), 539–549.
24. Schellingerhout, J. M., Verhagen, A. P., Heymans, M. W., Koes, B. W., de Vet, H. C., & Terwee, C. B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: A systematic review. *Quality of Life Research, 21*(4), 659–670.
25. Huang, H., Grant, J. A., Miller, B. S., Mirza, F. M., & Gagnier, J. J. (2015). A systematic review of the psychometric properties of patient-reported outcome instruments for use in patients with rotator cuff disease. *The American Journal of Sports Medicine, 43*(10), 2572–2582.
26. van Tulder, M., Furlan, A., Bombardier, C., & Bouter, L. (2003). Updated method guidelines for systematic reviews in the cochrane collaboration back review group. *Spine (Phila Pa 1976), 28*(12), 1290–1299.
27. Schellingerhout, J. M., Verhagen, A. P., Heymans, M. W., Koes, B. W., Henrica, C., & Terwee, C. B. (2012). Measurement properties of disease-specific questionnaires in patients with neck pain: A systematic review. *Quality of Life Research, 21*(4), 659–670.
28. Mousavian, A., Ebrahimzadeh, M. H., Birjandinejad, A., Omidi-Kashani, F., & Kachooei, A. R. (2015). Translation and cultural adaptation of the Manchester-Oxford Foot Questionnaire (MOXFQ) into Persian language. *Foot, 25*(4), 224–227.
29. Garcés, J. B. G., Winson, I., Goldhahn, S., Castro, M. D., Swords, M. P., Grujic, L., et al. (2015). Reliability, validity and responsiveness of the Spanish Manchester-Oxford Foot Questionnaire (MOXFQ) in patients with foot or ankle surgery. *Foot and Ankle Surgery*.
30. Dawson, J., Boller, I., Doll, H., Lavis, G., Sharp, R., Cooke, P., et al. (2014). Minimally important change was estimated for the Manchester-Oxford Foot Questionnaire after foot/ankle surgery. *Journal of Clinical Epidemiology, 67*(6), 697–705.
31. Morley, D., Jenkinson, C., Doll, H., Lavis, G., Sharp, R., Cooke, P., et al. (2013). The Manchester-Oxford Foot Questionnaire (MOXFQ): Development and validation of a summary index score. *Bone Joint Research, 2*(4), 66–69.
32. Dawson, J., Boller, I., Doll, H., Lavis, G., Sharp, R., Cooke, P., et al. (2012). Responsiveness of the Manchester-Oxford Foot Questionnaire (MOXFQ) compared with AOFAS, SF-36 and EQ-5D assessments following foot or ankle surgery. *The Journal of Bone and Joint Surgery. British Volume, 94*(2), 215–221.
33. Maher, A. J., & Kilmartin, T. E. (2012). An analysis of Euroqol EQ-5D and Manchester Oxford Foot Questionnaire scores six months following podiatric surgery. *Journal of Foot and Ankle Research, 5*(1), 17.
34. Dawson, J., Boller, I., Doll, H., Lavis, G., Sharp, R., Cooke, P., et al. (2011). The MOXFQ patient-reported questionnaire: assessment of data quality, reliability and validity in relation to foot and ankle surgery. *Foot, 21*(2), 92–102.
35. Dawson, J., Coffey, J., Doll, H., Lavis, G., Cooke, P., Herron, M., et al. (2006). A patient-based questionnaire to assess outcomes of foot surgery: Validation in the context of surgery for hallux valgus. *Quality of Life Research, 15*(7), 1211–1222.
36. Marinozzi, A., Martinelli, N., Panasci, M., Cancilleri, F., Franceschetti, E., Vincenzi, B., et al. (2009). Italian translation of the Manchester-Oxford Foot Questionnaire, with re-assessment of reliability and validity. *Quality of Life Research, 18*(7), 923–927.
37. Dawson, J., Doll, H., Coffey, J., & Jenkinson, C. (2007). Responsiveness and minimally important change for the Manchester-Oxford foot questionnaire (MOXFQ) compared with AOFAS and SF-36 assessments following surgery for hallux valgus. *Osteoarthritis and Cartilage, 15*(8), 918–931.
38. Van Der Zwaard, B. C., Terwee, C. B., Roddy, E., Terluin, B., Van Der Horst, H. E., & Elders, P. J. (2014). Evaluation of the measurement properties of the Manchester foot pain and disability index. *BMC Musculoskelet Disorders, 15*(1).
39. Menz, H. B., Auhl, M., Ristevski, S., Frescos, N., & Munteanu, S. E. (2014). Comparison of the responsiveness of the foot health status questionnaire and the Manchester foot pain and disability index in older people. *Health and Quality of Life Outcomes, 12*(158), 014–0158.
40. Gijon-Nogueron, G., Ndosi, M., Luque-Suarez, A., Alcacer-Pitarch, B., Munuera, P. V., Garrow, A., et al. (2014).

Cross-cultural adaptation and validation of the Manchester Foot Pain and Disability Index into Spanish. *Quality of Life Research, 23*(2), 571–579.
41. Pedersen, C. K., Danneskiold-Samsoe, B., Garrow, A. P., Waehrens, E. E., Bliddal, H., Christensen, R., et al. (2013). Development of a Danish language version of the manchester foot pain and disability index: Reproducibility and construct validity testing. *Pain Research Treatment, 284903*(10), 6.
42. Roddy, E., Muller, S., & Thomas, E. (2009). Defining disabling foot pain in older adults: Further examination of the Manchester Foot Pain and Disability Index. *Rheumatology, 48*(8), 992–996.
43. Kaoulla, P., Frescos, N., & Menz, H. B. (2008). Development and validation of a Greek language version of the Manchester Foot Pain and Disability Index. *Health and Quality of Life Outcomes, 6*(39), 1477–7525.
44. Menz, H. B., Tiedemann, A., Kwan, M. M. S., Plumb, K., & Lord, S. R. (2006). Foot pain in community-dwelling older people: An evaluation of the Manchester Foot Pain and Disability Index. *Rheumatology, 45*(7), 863–867.
45. Garrow, A. P., Papageorgiou, A. C., Silman, A. J., Thomas, E., Jayson, M. I., & Macfarlane, G. J. (2000). Development and validation of a questionnaire to assess disabling foot pain. *Pain, 85*(1–2), 107–113.
46. Pinsker, E., Inrig, T., Daniels, T., Warmington, K., & Beaton, D. E. (2015). Reliability and validity of 6 measures of pain, function, and disability for ankle arthroplasty and arthrodesis. *Foot & Ankle International, 36*(6), 617–625.
47. Budiman-Mak, E., Conrad, K. J., & Roach, K. E. (1991). The Foot Function Index: A measure of foot pain and disability. *Journal of Clinical Epidemiology, 44*(6), 561–570.
48. Kuyvenhoven, M. M., Gorter, K., Zuithoff, P., Zuithoff, P., Budiman-Mak, E., Conrad, K. J., et al. (2002). The foot function index with verbal rating scales (FFI-5pt): A clinimetric evaluation and comparison with the original FFI. *The Journal of Rheumatology, 29*(5), 1023–1028.
49. Agel, J., Beskin, J. L., Brage, M., Guyton, G. P., Kadel, N. J., Saltzman, C. L., et al. (2005). Reliability of the Foot Function Index: A report of the AOFAS Outcomes Committee. *Foot & Ankle International, 26*(11), 962–967.
50. SooHoo, N. F., Samimi, D. B., Vyas, R. M., Botzler, T., & Botzler, T. (2006). Evaluation of the validity of the Foot Function Index in measuring outcomes in patients with foot and ankle disorders. *Foot & Ankle International, 27*(1), 38–42.
51. SooHoo, N. F., Vyas, R., & Samimi, D. (2006). Responsiveness of the foot function index, AOFAS clinical rating systems, and SF-36 after foot and ankle surgery. *Foot & Ankle International, 27*(11), 930–934.
52. Naal, F. D., Impellizzeri, F. M., Huber, M., & Rippstein, P. F. (2008). Cross-cultural adaptation and validation of the Foot Function Index for use in German-speaking patients with foot complaints. *Foot & Ankle International, 29*(12), 1222–1228.
53. Wu, S. H., Liang, H. W., & Hou, W. H. (2008). Reliability and validity of the Taiwan Chinese version of the Foot Function Index. *Journal of the Formosan Medical Association, 107*(2), 111–118.
54. Pourtier-Piotte, C., Pereira, B., Soubrier, M., Thomas, E., Gerbaud, L., & Coudeyre, E. (2015). French validation of the Foot Function Index (FFI). *Annals of Physical Rehabilitation Medicine, 58*(5), 276–282.
55. Madeley, N. J., Wing, K. J., Topliss, C., Penner, M. J., Glazebrook, M. A., & Younger, A. S. (2012). Responsiveness and validity of the SF-36, Ankle Osteoarthritis Scale, AOFAS Ankle Hindfoot Score, and Foot Function Index in end stage ankle arthritis. *Foot & Ankle International, 33*(1), 57–63.
56. Paez-Moguer, J., Budiman-Mak, E., & Cuesta-Vargas, A. I. (2014). Cross-cultural adaptation and validation of the Foot Function Index to Spanish. *Foot and Ankle Surgery, 20*(1), 34–39.
57. Martinelli, N., Scotto, G. M., Sartorelli, E., Bonifacini, C., Bianchi, A., & Malerba, F. (2014). Reliability, validity and responsiveness of the Italian version of the Foot Function Index in patients with foot and ankle diseases. *Quality of Life Research, 23*(1), 277–284.
58. Jorgensen, J. E., Andreasen, J., & Rathleff, M. S. (2015). Translation and validation of the Danish Foot Function Index (FFI-DK). *Scandinavian Journal of Medicine and Science in Sports, 25*(4), 4.
59. Venditto, T., Tognolo, L., Rizzo, R. S., Iannuccelli, C., Di Sante, L., Trevisan, M., et al (2015). 17-Italian Foot Function Index with numerical rating scale: Development, reliability, and validity of a modified version of the original Foot Function Index. *Foot, 25*(1), 12–18.
60. Saag, K. G., Saltzman, C. L., Brown, C. K., & Budiman-Mak, E. (1996). The Foot Function Index for measuring rheumatoid arthritis pain: Evaluating side-to-side reliability. *Foot & Ankle International, 17*(8), 506–510.
61. Goldstein, C. L., Schemitsch, E., Bhandari, M., Mathew, G., & Petrisor, B. A. (2010). Comparison of different outcome instruments following foot and ankle trauma. *Foot & Ankle International, 31*(12), 1075–1080.
62. Martin, R. L., Irrgang, J. J., Burdett, R. G., Conti, S. F., & Van Swearingen, J. M. (2005). Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot & Ankle International, 26*(11), 968–983.
63. Carcia, C. R., Martin, R. l., & Drouin, J. M. (2008). Validity of the Foot and Ankle Ability Measure in athletes with chronic ankle instability. *Journal of Athletic Training, 43*(2), 179–183.
64. Martin, R. L., Hutt, D. M., & Wukich, D. K. (2009). Validity of the Foot and Ankle Ability Measure (FAAM) in Diabetes Mellitus. *Foot & Ankle International, 30*(4), 297–302.
65. Mazaheri, M., Salavati, M., Negahban, H., Sohani, S. M., Taghizadeh, F., Feizi, A., et al. (2010). Reliability and validity of the Persian version of Foot and Ankle Ability Measure (FAAM) to measure functional limitations in patients with foot and ankle disorders. *Osteoarthritis and Cartilage, 18*(6), 755–759.
66. Kivlan, B. R., Martin, R. l., & Wukich, D. K. (2011). Responsiveness of the foot and ankle ability measure (FAAM) in individuals with diabetes. *Foot, 21*(2), 84–87.
67. Arunakul, M., Arunakul, P., Suesiritumrong, C., Angthong, C., & Chernchujit, B. (2015). Validity and reliability of thai version of the foot and ankle ability measure (FAAM) subjective form. *Journal of the Medical Association of Thailand, 98*(6), 561–567.
68. Weel, H., Zwiers, R., Azim, D., Sierevelt, I. N., Haverkamp, D., van Dijk, C. N., et al. (2014). Validity and reliability of a Dutch version of the Foot and Ankle Ability Measure. *Knee Surgery, Sports Traumatology, Arthroscopy, 25*, 25.
69. Uematsu, D., Suzuki, H., Sasaki, S., Nagano, Y., Shinozuka, N., Sunagawa, N., et al. (2015). Evidence of validity for the Japanese version of the foot and ankle ability measure. *Journal of Athletic Training, 50*(1), 65–70.
70. Roos, E. M., Brandsson, S., & Karlsson, J. (2001). Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot & Ankle International, 22*(10), 788–794.
71. Karatepe, A. G., Gunaydin, R., Kaya, T., Karlibas, U., & Ozbek, G. (2009). Validation of the Turkish version of the foot and ankle outcome score. *Rheumatology International, 30*(2), 169–173.
72. Negahban, H., Mazaheri, M., Salavati, M., Sohani, S. M., Askari, M., Fanian, H., et al. (2010). Reliability and validity

of the foot and ankle outcome score: A validation study from Iran. *Clinical Rheumatology, 29*(5), 479–486.
73. Golightly, Y. M., DeVellis, R. F., Roos, E. M., Lohmander, L. S., Hannan, M. T., Nelson, A. E., et al. (2011). Psychometric properties of the foot and ankle outcome score (FAOS) in a community-based osteoarthritis study. *Osteoarthritis and Cartilage, 19*, S153.
74. van den Akker-Scheek, I., Seldentuis, A., Reininga, I. H. F., & Stevens, M. (2013). Reliability and validity of the Dutch version of the Foot and Ankle Outcome Score (FAOS). *BMC Musculoskeletal Disorders, 14*(183), 1471–2474.
75. Lee, K. M., Chung, C. Y., Kwon, S. S., Sung, K. H., Lee, S. Y., Won, S. H., et al. (2013). Transcultural adaptation and testing psychometric properties of the Korean version of the Foot and Ankle Outcome Score (FAOS). *Clinical Rheumatology, 32*(10), 1443–1450.
76. Mani, S. B., Brown, H. C., Nair, P., Chen, L., Do, H. T., Lyman, S., et al. (2013). Validation of the Foot and Ankle Outcome Score in adult acquired flatfoot deformity. *Foot & Ankle International, 34*(8), 1140–1146.
77. van Bergen, C. J., Sierevelt, I. N., Hoogervorst, P., Waizy, H., van Dijk, C. N., & Becher, C. (2014). Translation and validation of the German version of the foot and ankle outcome score. *Archives of Orthopaedic and Trauma Surgery, 134*(7), 897–901.
78. Golightly, Y. M., Devellis, R. F., Nelson, A. E., Hannan, M. T., Lohmander, L. S., Renner, J. B., et al. (2014). Psychometric properties of the foot and ankle outcome score in a community-based study of adults with and without osteoarthritis. *Arthritis Care and Research, 66*(3), 395–403.
79. Angthong, C. (2015). Validity and reliability of Thai version of the Foot and Ankle Outcome Score in patients with arthritis of the foot and ankle. *Foot and Ankle Surgery*.
80. Sierevelt, I. N., Beimers, L., van Bergen, C. J. A., Haverkamp, D., Terwee, C. B., & Kerkhoffs, G. M. M. J. (2015). Validation of the Dutch language version of the Foot and Ankle Outcome Score. *Knee Surgery, Sports Traumatology, Arthroscopy, 23*(8), 2413–2419.
81. Sierevelt, I. N., van Eekeren, I. C., Haverkamp, D., Reilingh, M. L., Terwee, C. B., & Kerkhoffs, G. M. (2015). Evaluation of the Dutch version of the Foot and Ankle Outcome Score (FAOS): Responsiveness and Minimally Important Change. *Knee Surgery, Sports Traumatology, Arthroscopy, 24*, 24.
82. Mani, S. B., Do, H., Vulcano, E., Hogan, M. V., Lyman, S., Deland, J. T., et al. (2015). Evaluation of the foot and ankle outcome score in patients with osteoarthritis of the ankle. *Bone Joint J, 97*, 662–667.
83. Chen, L., Lyman, S., Do, H., Karlsson, J., Adam, S. P., Young, E., et al. (2012). Validation of foot and ankle outcome score for hallux valgus. *Foot & Ankle International, 33*(12), 1145–1155.
84. Cuesta-Vargas, A., Bennett, P., Jimenez-Cebrian, A. M., & Labajos-Manzanares, M. T. (2013). The psychometric properties of the Spanish version of the Foot Health Status Questionnaire. *Quality of Life Research, 22*(7), 1739–1743.
85. Bennett, P. J., Patterson, C., Wearing, S., & Baglioni, T. (1998). Development and validation of a questionnaire designed to measure foot-health status. *Journal of the American Podiatric Medical Association, 88*(9), 419–428.
86. Hiller, C. E., Refshauge, K. M., Bundy, A. C., Herbert, R. D., & Kilbreath, S. L. (2006). The Cumberland ankle instability tool: A report of validity and reliability testing. *Archives of Physical Medicine and Rehabilitation, 87*(9), 1235–1241.
87. De Noronha, M., Refshauge, K. M., Kilbreath, S. L., & Figueiredo, V. G. (2008). Cross-cultural adaptation of the Brazilian-Portuguese version of the Cumberland Ankle Instability Tool (CAIT). *Disability and Rehabilitation, 30*(26), 1959–1965.
88. Cruz-Diaz, D., Hita-Contreras, F., Lomas-Vega, R., Osuna-Perez, M. C., & Martinez-Amat, A. (2013). Cross-cultural adaptation and validation of the Spanish version of the Cumberland Ankle Instability Tool (CAIT): An instrument to assess unilateral chronic ankle instability. *Clinical Rheumatology, 32*(1), 91–98.
89. Rodriguez-Fernandez, A. L., Rebollo-Roldan, J., Jimenez-Rejano, J. J., & Gueita-Rodriguez, J. (2015). Psychometric properties of the Spanish version of the Cumberland Ankle Instability Tool. *Disability and Rehabilitation, 37*(20), 1888–1894.
90. Ko, J., Rosen, A. B., & Brown, C. N. (2015). Cross-Cultural Adaptation and Validation of the Korean Version of the Cumberland Ankle Instability Tool. *International Journal Sports Physical Therapy, 10*(7), 1007–1014.
91. Wang, Y., He, Z., Lei, L., Lin, D., Li, Y., Wang, G., et al. (2015). Reliability and validity of the Chinese version of the Short Musculoskeletal Function Assessment questionnaire in patients with skeletal muscle injury of the upper or lower extremities. *BMC Musculoskelet Disorders, 16*, 161
92. Domsic, R. T., & Saltzman, C. L. (1998). Ankle osteoarthritis scale. *Foot & Ankle International, 19*(7), 466–471.
93. McPhail, S. M., Williams, C. M., Schuetz, M., Baxter, B., Tonks, P., & Haines, T. P. (2014). Development and validation of the ankle fracture outcome of rehabilitation measure (A-FORM). *The Journal of Orthopaedic and Sports Physical Therapy, 44*(7), 488–499.
94. Andre, M., Hagelberg, S., & Stenstrom, C. H. (2004). The juvenile arthritis foot disability index: Development and evaluation of measurement properties. *The Journal of Rheumatology, 31*(12), 2488–2493.
95. Esbjörnsson, A. C., Iversen, M. D., Broström, E. W., Hagelberg, S., & André, M. (2013). PReS-FINAL-2024: Responsiveness of the juvenile arthritis foot disability index. *Pediatric Rheumatology, 11*, P37
96. Morris, C., Doll, H. A., Wainwright, A., Theologis, T., & Fitzpatrick, R. (2008). The Oxford ankle foot questionnaire for children: Scaling, reliability and validity. *The Journal of Bone and Joint Surgery. British volume, 90*(11), 1451–1456.
97. Morris, C., Doll, H., Davies, N., Wainwright, A., Theologis, T., Willett, K., et al. (2009). The Oxford Ankle Foot Questionnaire for children: Responsiveness and longitudinal validity. *Quality of Life Research, 18*(10), 1367–1376.
98. Martinkevich, P., Moller-Madsen, B., Gottliebsen, M., Kjeldgaard, P. L., & Rahbek, O. (2015). Validation of the translated Oxford ankle foot questionnaire in 82 Danish children aged between five and 16 years. *Bone Joint J, 97*, 420–426.
99. Martinelli, N., Romeo, G., Bonifacini, C., Viganò, M., Bianchi, A., & Malerba, F. (2016). Validation of the Italian version of the Oxford Ankle Foot Questionnaire for children. *Quality of Life Research, 25*(1), 117–123.
100. Niki, H., Tatsunami, S., Haraguchi, N., Aoki, T., Okuda, R., Suda, Y., et al. (2013). Validity and reliability of a self-administered foot evaluation questionnaire (SAFE-Q). *Journal of Orthopaedic Science, 18*(2), 298–320.
101. Yano, K., Ikari, K., Ochi, K., Ishida, O., Sakuma, Y., Yoshida, S., et al. (2015). Validity and responsiveness of a self-administered foot evaluation questionnaire in rheumatoid arthritis. *Modern Rheumatology, 25*(3), 358–361.
102. Coster, M. C., Karlsson, M. K., Nilsson, J.-A., & Carlsson, A. K. E. (2012). Validity, reliability, and responsiveness of a self-reported foot and ankle score (SEFAS). *Acta Orthopaedica, 83*(2), 197–203.
103. Coster, M. C., Rosengren, B. E., Bremander, A., Brudin, L., & Karlsson, M. K. (2014). Comparison of the Self-reported Foot and Ankle Score (SEFAS) and the American Orthopedic

Foot and Ankle Society Score (AOFAS). *Foot & Ankle International, 35*(10), 1031–1036.
104. Coster, M. C., Bremander, A., Rosengren, B. E., Magnusson, H., Carlsson, A., & Karlsson, M. K. (2014). Validity, reliability, and responsiveness of the Self-reported Foot and Ankle Score (SEFAS) in forefoot, hindfoot, and ankle disorders. *Acta Orthopaedica, 85*(2), 187–194.
105. Rowan, K. (2001). The development and validation of a multi-dimensional measure of chronic foot pain: The ROwan Foot Pain Assessment Questionnaire (ROFPAQ). *Foot & Ankle International, 22*(10), 795–809.
106. Woodburn, J., Vliet, V. T. P., van der Leeden, M., & Steultjens, M. P. M. (2011). Rasch analysis of Dutch-translated version of the Foot Impact Scale for rheumatoid arthritis. *Rheumatology, 50*(7), 1315–1319.
107. Woodburn, J., Turner, D. E., Rosenbaum, D., Balint, G., Korda, J., Ormos, G., et al. (2012). Adaptation and cross-cultural validation of the Foot Impact Scale for Rheumatoid Arthritis using Rasch analysis. *Arthritis Care and Research, 64*(7), 986–992.
108. Abetz, L., Sutton, M., Brady, L., McNulty, P., & Gagnon, D. D. (2002). The Diabetic Foot Ulcer Scale (DFS): A quality of life instrument for use in clinical trials. *Practical Diabetes International, 19*(6), 167–175.
109. Bann, C. M., Fehnel, S. E., & Gagnon, D. D. (2003). Development and validation of the Diabetic Foot Ulcer Scale-short form (DFS-SF). *PharmacoEconomics, 21*(17), 1277–1290.
110. Hui, L. F., Yee-Tak, F. D., Yam, M., & Yuk Ip, W. (2008). Translation and validation of the Chinese diabetic foot ulcer scale—short form. *Patient, 1*(2), 137–145.
111. Chin, Y. F., & Huang, T. T. (2013). Development and validation of a diabetes foot self-care behavior scale. *The Journal of Nursing Research: JNR, 21*(1), 19–25.
112. Pena, F., Agel, J., & Coetzee, J. C. (2007). Comparison of the MFA to the AOFAS outcome tool in a population undergoing total ankle replacement. *Foot & Ankle International, 28*(7), 788–793.
113. Ibrahim, T., Beiri, A., Azzabi, M., Best, A. J., Taylor, G. J., & Menon, D. K. (2007). Reliability and validity of the subjective component of the American Orthopaedic Foot and Ankle Society clinical rating scales. *The Journal of Foot and Ankle Surgery, 46*(2), 65–74.
114. Baumhauer, J. F., Nawoczenski, D. A., DiGiovanni, B. F., & Wilding, G. E. (2006). Reliability and validity of the American Orthopaedic Foot and Ankle Society Clinical Rating Scale: A pilot study for the hallux and lesser toes. *Foot & Ankle International, 27*(12), 1014–1019.
115. Hale, S. A., & Hertel, J. (2005). Reliability and sensitivity of the foot and ankle disability index in subjects with chronic ankle instability. *Journal of Athletic Training, 40*(1), 35–40.
116. Niki, H., Aoki, H., Inokuchi, S., Ozeki, S., Kinoshita, M., Kura, H., et al. (2005). Development and reliability of a standard rating system for outcome measurement of foot and ankle disorders II: interclinician and intraclinician reliability and validity of the newly established standard rating scales and Japanese Orthopaedic Association rating scale. *Journal of Orthopaedic Science, 10*(5), 466–474.
117. Dhawan, V., Spratt, K. F., Pinzur, M. S., Baumhauer, J., Rudicel, S., & Saltzman, C. L. (2005). Reliability of AOFAS diabetic foot questionnaire in Charcot arthropathy: stability, internal consistency, and measurable difference. *Foot & Ankle International, 26*(9), 717–731.
118. Kim, J. B., Kim, J. K., Seo, S. G., & Lee, D. Y. (2015). Validity, reliability, and responsiveness of the Korean version of American Academy of Orthopedic Surgeons Foot and Ankle questionnaire. *The Journal of Foot and Ankle Surgery, 54*(1), 46–50.
119. Boszczyk, A., Blonski, M., & Pomianowski, S. (2015). Translation, cultural adaptation and validation of Polish version of foot and ankle outcomes questionnaire. *Ortopedia, Traumatologia, Rehabilitacja, 17*(2), 175–187.
120. Morssinkhof, M. L., Wang, O., James, L., van der Heide, H. J. L., & Winson, I. G. (2013). Development and validation of the Sports Athlete Foot and Ankle Score: An instrument for sports-related ankle injuries. *Foot and Ankle Surgery, 19*(3), 162–167.
121. Sibaud, V., Dalenc, F., Chevreau, C., Roché, H., Delord, J. P., Mourey, L., et al. (2011). HFS-14, a specific quality of life scale developed for patients suffering from hand-foot syndrome. *Melanoma Research, 21*, e12–e13.
122. Mikoshiba, N., Yamamoto-Mitani, N., Sato, K., Asaoka, Y., Ohki, T., Ohata, M., et al. (2015). Validation of the Japanese version of HFS-14, a disease-specific quality of life scale for patients suffering from hand-foot syndrome. *Supportive Care in Cancer, 23*(9), 2739–2745.
123. Vileikyte, L., Peyrot, M., Bundy, C., Rubin, R. R., Leventhal, H., Mora, P., et al. (2003). The development and validation of a neuropathy- and foot ulcer-specific quality of life instrument. *Diabetes Care, 26*(9), 2549–2555.
124. Xavier, A. T., Foss, M. C., Marques Junior, W., dos Santos, C. B., Onofre, P. T., & Pace, A. E. (2011). Cultural adaptation and validation of the Neuropathy - and Foot Ulcer - Specific Quality of Life instrument (NeuroQol) for Brazilian Portuguese - Phase 1. *Revista latino-americana de enfermagem, 19*(6), 1352–1361.
125. Nair, A. V., Shamsuddin, K., John, P. S., Hamalainen, J. A., & Kurien, M. A. (2015). Correlation of visual analogue scale foot and ankle (VAS-FA) to AOFAS score in malleolar fractures using Indian language questionnare. *Foot and Ankle Surgery, 21*(2), 125–131.
126. Angthong, C., Chernchujit, B., Suntharapa, T., & Harnroongroj, T. (2011). Visual analogue scale foot and ankle: validity and reliability of Thai version of the new outcome score in subjective form. *Journal of the Medical Association of Thailand, 94*(8), 952–957.
127. Jaksa, P. J., & Mahoney, J. L. (2010). Quality of life in patients with diabetic foot ulcers: Validation of the Cardiff Wound Impact Schedule in a Canadian population. *International Wound Journal, 7*(6), 502–507.
128. Jannink, M. J., de Vries, J., Stewart, R. E., Groothoff, J. W., & Lankhorst, G. J. (2004). Questionnaire for usability evaluation of orthopaedic shoes: construction and reliability in patients with degenerative disorders of the foot. *Journal of Rehabilitation Medicine, 36*(6), 242–248.
129. Ferrari, R. (2007). Responsiveness of the short-form 36 and oswestry disability questionnaire in chronic nonspecific low back and lower limb pain treated with customized foot orthotics. *Journal of Manipulative and Physiological Therapeutics, 30*(6), 456–458.
130. Cook, J. J., Cook, E. A., Rosenblum, B. I., Landsman, A. S., Roukis, T. S., & Roukis, T. S. (2011). Validation of the American College of Foot and Ankle Surgeons Scoring Scales. *The Journal of Foot and Ankle Surgery, 50*(4), 420–429.
131. Nilsson, G. M., Eneroth, M., & Ekdahl, C. S. (2013). The Swedish version of OMAS is a reliable and valid outcome measure for patients with ankle fractures. *BMC Musculoskeletal Disorders, 14*(109), 1471–2474.
132. Hung, M., Baumhauer, J. F., Latt, L. D., Saltzman, C. L., SooHoo, N. F., & Hunt, K. J. (2013). Validation of PROMIS (R) Physical Function computerized adaptive tests for orthopaedic foot and ankle outcome research. *Clinical Orthopaedics and Related Research, 471*(11), 3466–3474.
133. Lai, J. S., Beaumont, J. L., Diaz, J., Khan, S., & Cella, D. (2016). Validation of a short questionnaire to measure symptoms and functional limitations associated with hand-foot

syndrome and mucositis in patients with metastatic renal cell carcinoma. *Cancer, 122*(2), 287–295.
134. Anderson, R. T., Keating, K. N., Doll, H. A., & Camacho, F. (2015). The Hand-Foot Skin Reaction and Quality of Life Questionnaire: An Assessment Tool for Oncology. *The oncologist, 20*(7), 831–838.
135. Garcia-Inzunza, J. A., Valles-Medina, A. M., Munoz, F. A., Delgadillo-Ramos, G., & Compean-Ortiz, L. G. (2015). Validity of the Mexican version of the combined Foot Care Confidence / Foot-Care Behavior scale for diabetes. *Revista Panamericana de Salud Publica, 38*(1), 35–41.
136. Walmsley, S., Ravey, M., Graham, A., Teh, L. S., & Williams, A. E. (2012). Development of a patient-reported outcome measure for the foot affected by rheumatoid arthritis. *Journal of Clinical Epidemiology, 65*(4), 413–422.
137. Navarro-Flores, E., Morales-Asencio, J. M., Cervera-Marin, J. A., Labajos-Manzanares, M. T., & Gijon-Nogueron, G. (2015). Development, validation and psychometric analysis of the diabetic foot self-care questionnaire of the University of Malaga, Spain (DFSQ-UMA). *Journal of Tissue Viability, 24*(1), 24–34.
138. Castillo-Tandazo, W., Flores-Fortty, A., Feraud, L., & Tettamanti, D. (2013). Spanish translation, cross-cultural adaptation, and validation of the Questionnaire for Diabetes-Related Foot Disease (Q-DFD). *Vascular Health and Risk Management, 9*, 501–508.
139. Niki, H., Tatsunami, S., Haraguchi, N., Aoki, T., Okuda, R., Suda, Y., et al. (2011). Development of the patient-based outcome instrument for foot and ankle: Part 2: Results from the second field survey: validity of the Outcome Instrument for the foot and ankle version 2. *Journal of Orthopaedic Science, 16*(5), 556–564.
140. Eechaute, C., Vaes, P., & Duquet, W. (2008). The chronic ankle instability scale: clinimetric properties of a multidimensional, patient-assessed instrument. *Physical Therapy in Sport, 9*(2), 57–66.
141. Wulterkens, L., Aurégan, J. C., Letellier, T., Mebtouche, N., Levante, S., Cottin, P., et al. (2015). A telephone questionnaire in order to assess functional outcome after post-traumatic limb salvage surgery: Development and preliminary validation. *Injury, 46*(12), 2452–2456.
142. Roos, E. M., Roos, H. P., Lohmander, L. S., Ekdahl, C., & Beynnon, B. D. (1998). Knee Injury and Osteoarthritis Outcome Score (KOOS)–development of a self-administered outcome measure. *The Journal of Orthopaedic and Sports Physical Therapy, 28*(2), 88–96.
143. Swiontkowski, M. F., Engelberg, R., Martin, D. P., & Agel, J. (1999). Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. *The Journal of Bone and Joint Surgery. American Volume, 81*(9), 1245–1260.
144. Bellamy, N., Buchanan, W. W., Goldsmith, C. H., Campbell, J., & Stitt, L. W. (1988). Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *The Journal of Rheumatology, 15*(12), 1833–1840.
145. Niki, H., Tatsunami, S., Haraguchi, N., Aoki, T., Okuda, R., Suda, Y., et al. (2011). Development of the patient-based outcome instrument for the foot and ankle. Part 1: project description and evaluation of the Outcome Instrument version 1. *Journal of Orthopaedic Science, 16*(5), 536–555.
146. Hosman, A. H., Mason, R. B., Hobbs, T., & Rothwell, A. G. (2007). A New Zealand national joint registry review of 202 total ankle replacements followed for up to 6 years. *Acta Orthopaedica, 78*(5), 584–591.
147. Helliwell, P., Reay, N., Gilworth, G., Redmond, A., Slade, A., Tennant, A., et al. (2005). Development of a foot impact scale for rheumatoid arthritis. *Arthritis and Rheumatism, 53*(3), 418–422.
148. Kitaoka, H. B., Alexander, I. J., Adelaar, R. S., Nunley, J. A., Myerson, M. S., & Sanders, M. (1994). Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. *Foot & Ankle International, 15*(7), 349–353.
149. Martin, R. L., Burdett, R. G., & Irrgang, J. J. (1999). Development of the Foot and Ankle Disability Index (FADI). *The Journal of Orthopaedic and Sports Physical Therapy, 29*, A32–A33.
150. Greenfield, S., Kaplan, S. H., Silliman, R. A., Sullivan, L., Manning, W., D'Agostino, R., et al. (1994). The uses of outcomes research for medical effectiveness, quality of care, and reimbursement in type II diabetes. *Diabetes Care, 17 (Suppl 1)*, 32–39.
151. Sibaud, V., Dalenc, F., Chevreau, C., Roche, H., Delord, J. P., Mourey, L., et al. (2011). HFS-14, a specific quality of life scale developed for patients suffering from hand-foot syndrome. *The Oncologist, 16*(10), 1469–1478.
152. Richter, M., Zech, S., Geerling, J., Frink, M., Knobloch, K., & Krettek, C. (2006). A new foot and ankle outcome score: Questionnaire based, subjective, Visual-Analogue-Scale, validated and computerized. *Foot and Ankle Surgery, 12*(4), 191–199.
153. Price, P., & Harding, K. (2004). 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. *International Wound Journal, 1*(1), 10–17.
154. Martin, D. P., Engelberg, R., Agel, J., Snapp, D., & Swiontkowski, M. F. (1996). Development of a musculoskeletal extremity health status instrument: the Musculoskeletal Function Assessment instrument. *Journal of Orthopaedic Research, 14*(2), 173–181.
155. Fairbank, J. C., Couper, J., Davies, J. B., & O'Brien, J. P. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy, 66*(8), 271–273.
156. Zlotoff, H. J., Christensen, J. C., Mendicino, R. W., Schuberth, J. M., Schwartz, N. H., Thomas, J. L., et al. (2002). ACFAS Universal Foot and Ankle Scoring System: First Metatarsophalangeal Joint and First Ray (module 1). *The Journal of Foot and Ankle Surgery, 41*(1), 2–5.
157. Olerud, C., & Molander, H. (1984). A scoring scale for symptom evaluation after ankle fracture. *Archives of Orthopaedic and Trauma Surgery, 103*(3), 190–194.
158. Cella, D., Yount, S., Rothrock, N., Gershon, R., Cook, K., Reeve, B., et al. (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care, 45*(5 Suppl 1), S3–S11.
159. Wong, M. K., Mohamed, A. F., Hauber, A. B., Yang, J. C., Liu, Z., Rogerio, J., et al. (2012). Patients rank toxicity against progression free survival in second-line treatment of advanced renal cell carcinoma. *Journal of Medical Economics, 15*(6), 1139–1148.
160. Perrin, B. M., Swerissen, H., & Payne, C. (2009). The association between foot-care self efficacy beliefs and actual foot-care behaviour in people with peripheral neuropathy: A cross-sectional study. *Journal of Foot and Ankle Research, 2*, 3.
161. Bergin, S. M., Brand, C. A., Colman, P. G., & Campbell, D. A. (2009). A questionnaire for determining prevalence of diabetes related foot disease (Q-DFD): Construction and validation. *Journal of Foot and Ankle Research, 2*, 34.
162. de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). Measurement in Medicine: A Practical Guide (Practical Guides to Biostatistics and Epidemiology). Cambridge: Cambridge University Press, pp, 347.

163. Elbers, R. G., Rietberg, M. B., van Wegen, E. E., Verhoef, J., Kramer, S. F., Terwee, C. B., et al. (2012). Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and stroke: a systematic review of measurement properties. *Quality of Life Research, 21*(6), 925–944.
164. Uijen, A. A., Heinst, C. W., Schellevis, F. G., van den Bosch, W. J., van de Laar, F. A., Terwee, C. B., et al. (2012). Measurement properties of questionnaires measuring continuity of care: A systematic review. *PLoS ONE, 7*(7), e42256.
165. Moher, D., Pham, B., Lawson, M. L., & Klassen, T. P. (2003). The inclusion of reports of randomised trials published in languages other than English in systematic reviews. *Health Technology Assessment, 7*(41), 1–90.
166. Morrison, A., Polisena, J., Husereau, D., Moulton, K., Clark, M., Fiander, M., et al. (2012). The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. *International Journal of Technology Assessment in Health Care, 28*(2), 138–144.
167. Williamson, P., & Clarke, M. (2012). The COMET (Core Outcome Measures in Effectiveness Trials) initiative: Its Role in improving cochrane reviews. *Cochrane Database of Systematic Reviews (Online), 5*, ED000041.