

The WOUND-Q Function and Symptoms Scales for Chronic Lower Extremity Wounds: A Validation Study

Nina Vestergaard Simonsen,^{1,*} Anne F. Klassen,² Charlene Rae,² Lily R. Mundy,³ Lotte Poulsen,⁴ Andrea L. Pusic,⁵ Kenneth L. Fan,⁶ and Jens Ahm Sørensen¹

¹Research Unit for Plastic Surgery, Department of Plastic Surgery, Odense University Hospital, University of Southern Denmark, Odense, Denmark.

²Department of Pediatrics, McMaster University, Hamilton, Canada.

³Department of Plastic and Reconstructive Surgery, Johns Hopkins School of Medicine, Baltimore, Maryland, USA.

⁴Research Unit for Plastic Surgery, Odense University Hospital, University of Southern Denmark, Denmark.

⁵Department of Surgery and Patient-Reported Outcomes, Value & Experience (PROVE) Center, Brigham and Women's Hospital, Harvard Medical School, Boston, Maryland, USA.

⁶Department of Plastic and Reconstructive Surgery, MedStar, MedStar Plastic and Reconstructive Surgery, Washington, USA.



Open camera or QR reader and scan code to access this article and other resources online.

Objective: Determine the validity and reliability of the LIMB-Q scales, Function, and Symptoms in patients with chronic lower extremity wounds.

Approach: Cognitive debriefing interviews with people with current or previous wounds were conducted to examine content validity. Scales were field-tested in an international sample of people with chronic lower extremity wounds sourced from an online platform (*i.e.*, Prolific). Psychometric properties were examined using the Rasch Measurement Theory analysis. A test-retest reproducibility study was performed, and construct validity was examined.

Results: Content validity was established after 10 cognitive interviews. A total of 233 people with lower extremity wounds (age 19–80 years, mean 39.3) participated in the field test. All 25 items tested demonstrated good fit to the Rasch model with ordered thresholds. One item had a fit residual outside ± 2.5 , but no items had significant χ^2 values after Bonferroni adjustment. Reliability was high with the person separation index, Cronbach alpha, and intraclass correlation coefficient values >0.8 . Strong correlations were found between the Function and Symptoms scales and EQ-5D dimensions measuring similar constructs as well as the EQ-5D global score. All hypotheses for construct validity were confirmed.

Innovation: Patient-reported outcome measures are an important component of patient-centered care, as they capture the patient's perspective in a rigorous and reproducible way. Adding these two scales to the WOUND-Q provides a means to measure function and symptoms associated with lower extremity wounds.

Conclusion: These new WOUND-Q scales can be used to measure outcomes important to patients with lower extremity wounds in clinical settings and research studies.

Keywords: leg wound, psychometrics, patient-reported outcome, health-related quality of life, reliability, amputation



Nina Vestergaard Simonsen, MD, PhD student

Submitted for publication February 09, 2024.
Accepted in revised form May 12, 2024.

*Correspondence: MD, PhD student, J.B. Winsløvs Vej 4, Odense C, 5000, Denmark
(e-mails: Nina.vestergaard.simonsen2@rsyd.dk; Nina6533@hotmail.com)

INTRODUCTION

Chronic wounds are a growing public health challenge.^{1,2} With a prevalence of 2.21 per 1,000 people,³ chronic wounds impact millions of people worldwide. The most common type of chronic wounds are lower extremity wounds,³ which tend to be caused by arterial insufficiency, venous insufficiency, mixed arterial and venous insufficiency, pressure injuries, and/or diabetes.⁴ The risk of developing a lower extremity wound increases in older people, those with obesity, and those with diabetes.^{1,5–7} The increasing incidence of chronic lower extremity wounds, linked to an aging population and rising obesity rates, places a considerable burden on the health care systems.^{1,2,8,9}

In addition to the high burden that lower extremity wounds have on health systems, they also have a multifaceted and often negative impact on patients.^{8,10–14} Wound-related concerns such as pain,^{8,10,15,16} foul odor,^{10,14} impaired mobility, sleep disturbances, social isolation, and financial and psychological distress^{8,10,11,14–17} contribute to worse health-related quality of life (HRQL). Given the significant impact lower extremity wounds can have on HRQL, it is imperative to understand how HRQL is affected to improve it.¹⁸ Patient-reported outcome measures (PROMs) are designed to measure patients' self-reported experiences, symptoms, functioning, and quality of life in a quantitative manner. These outcomes, also known as patient-reported outcomes (PRO), are defined by the Food and Drug Administration as "any report coming directly from patients about health condition and its treatment."¹⁹ The implementation of PROMs in clinical practice has been found to enhance the quality of health care through improvements in diagnosis, patients' perceptions of communication, disease control, and patient's quality of life.²⁰ PROMs can be divided into two main categories, that is, generic and condition-specific. Generic PROMs can be used across a wide range of health conditions; in contrast, condition-specific PROMs are developed to assess outcomes for a specific condition. When PROMs are designed by eliciting concepts important to patients, it ensures that only patient-centered outcomes relevant to patients are measured. A condition-specific PROM further ensures that important and relevant outcomes to the specific patient population are measured. The choice between these two types of PROMs depends on the objective of the assessments.^{19,21}

INNOVATION

Lower extremity wounds are common and associated with a significant symptom burden and

impact on function and mobility. We have added two new scales to the condition-specific PROM WOUND-Q (*i.e.*, Function, Symptoms) for patients with lower extremity wounds (Fig. 1). Adding these scales to the WOUND-Q fills an important gap in the measurement of outcomes for lower extremity wounds. Using the WOUND-Q with these two new scales in research and clinical settings has the potential to improve patient care.

CLINICAL PROBLEM ADDRESSED

The WOUND-Q is a condition-specific PROM developed and validated to measure PROs important to people with chronic wounds. It has recently been identified as the highest quality PROM for chronic wounds.²² The development and validation of the WOUND-Q have been described previously.^{23–27} Given its modular design, WOUND-Q offers the flexibility to expand by adding new scales that enhance its coverage. The original version of WOUND-Q did not include scales for assessing function and symptoms related to lower extremity wounds. These are important concepts for patients receiving wound treatment, which could be integrated to enrich the PROM's comprehensiveness.¹⁶

Our group previously developed and validated the LIMB-Q, a PROM for lower extremity trauma patients.^{28–32} We followed the same international, rigorous PROM development guidelines as we did for the WOUND-Q to develop the LIMB-Q.^{19,23,28,33–36} Although the LIMB-Q was developed for lower extremity trauma, patients with limb-threatening soft tissue wounds, independent of the etiology, often undergo very similar treatment pathways. For this reason, our group hypothesized that the LIMB-Q Function and Symptoms scales would be relevant to patients with lower extremity wounds. Therefore, the primary aim of this study was to evaluate the content validity and psychometric performance of the LIMB-Q Function and Symptoms scales in patients with lower extremity wounds.

MATERIALS AND METHODS

This study was performed in two parts: an initial qualitative phase followed by a quantitative phase. Ethics approval for the qualitative phase was obtained at Georgetown University (IRB STUDY 00001591) and for the quantitative phase, from McMaster University (Hamilton Integrated Research Ethics Board [#14946]). A license agreement was obtained to use the EQ-5D-5L.³⁷

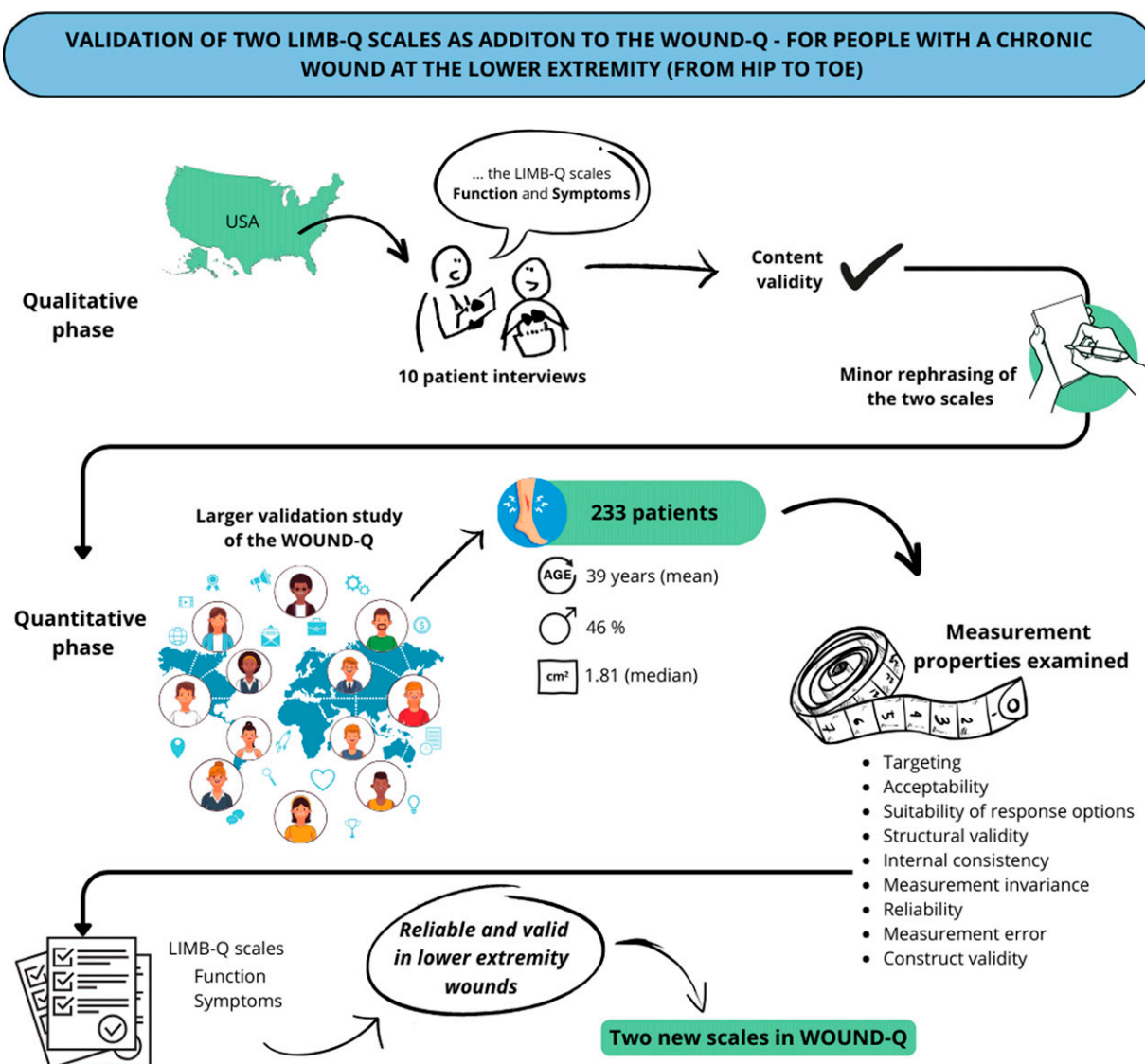


Figure 1. Summary graphic.

The Function and Symptoms scales

The development of the LIMB-Q has been published previously^{28–32} and followed international guidelines for PROM development.^{19,33–36} Briefly, the LIMB-Q contains 16 independently functioning scales. These scales were developed from 33 qualitative patient interviews, further refined through 12 cognitive debriefing interviews and insights from 43 experts and field-tested in an international sample of 713 patients. The field-test versions of the Function and Symptoms scales had 37 and 17 items, respectively, which were reduced to 15 and 10 items following the psychometric analysis using the Rasch Measurement Theory (RMT).³² In the qualitative part of this study, we examined content validity for all 54 items. For the RMT analysis, we performed the psychometric analysis for the final set of 25 items.

Content validity: Qualitative phase. Content validity was established through a series of cognitive debriefing interviews. The sample included patients with chronic lower extremity wounds recruited from a multidisciplinary outpatient wound clinic at Georgetown University. This institution has a unique chronic wound and limb-salvage clinic. It is overseen by plastic surgeons but is run by a multidisciplinary team that includes vascular surgery, wound care, physical medicine and rehabilitation, and prosthetists. This clinic manages the chronic and diabetic wounds for the health system. To maximize variability in clinical and demographic variables, purposeful sampling was used. Recruitment took place between June and September 2020. As only the field-test version was available at the start of the study, this version was included in the qualitative and quantitative

phases of this study. To adapt the LIMB-Q scales for the chronic lower extremity wound population, minor adjustments were made in the instructions, such as replacing “injured lower limb” with “lower limb.” These modifications are shown in Supplementary Appendix A. Interviews were conducted using an interview guide and performed by one interviewer (TS). Participants were shown the two LIMB-Q scales, Function and Symptoms, during their interview. Feedback was solicited on the comprehensibility and relevance of the instructions, items, and response options, and participants were encouraged to suggest any missing content. All interviews were audio-recorded, transcribed, and coded line-by-line. Coding was conducted in Microsoft Word (Version 2307; Redmond, WA) by N.V.S. and transferred to a Microsoft Excel (Version 2307; Redmond, WA) worksheet for analysis using DocTools.

Field testing: Quantitative phase.

Sample and recruitment. The LIMB-Q Function and Symptoms scales were administered to an international sample of patients with chronic lower extremity wounds identified in Prolific Academic (“Prolific” www.prolific.com), an online patient research platform. Data were collected using a REDCap (Research Electronic Data Capture) survey and collected as part of a larger study examining the psychometric performance of 9 WOUND-Q scales across various chronic wound types in an international community-based sample.²⁵ In Prolific, an initial screen was administered in August 2022 to identify people aged 18 years and older who were fluent in English and had one or more chronic wounds located anywhere on their body. A chronic wound was defined as a wound that had lasted at least 3 months. Eligible patients were then invited to complete the survey in September 2022. Participants who reported a wound on their leg, ankle, foot, or toe using branching logic were invited to answer the LIMB-Q Function and Symptoms scale. The survey did not specify the anatomical definition of the leg. Based on common knowledge, the leg should be understood as the region from the hip to the ankle. In addition to completing the scales, participants answered questions about wound characteristics, treatments, and demographic information. They also completed the EQ-5D-5L,³⁷ a generic HRQL PROM that has been validated in venous leg ulcers.^{38–40} The EQ-5D-5L consists of two sections. One part describes health states by five dimensions, rated from level 1 (no problem) to 5 (unable/extreme problems). The second part is the EQ-VAS, where participants rate

their overall health state from 0 (worst) to 100 (best). EQ-5D assessments were used to examine convergent validity. At the end of the survey, participants were invited to participate in a test–retest reproducibility (TRT) study seven days after the initial assessment. We aimed to include about 100 assessments in REDCap for the test–retest survey. Since the number of assessments for the TRT survey was limited, participation was on a first-come-first-serve basis. In the main survey and TRT, participants were able to skip answering any of the scale items. All participants were paid a minimum of 10 GBP per hour.

Scoring of PROMs

Scores for the Function and Symptoms scales were computed by adding the raw scores of the scales to get a total raw score. For cases with missing data with less than 50% of the scale’s items, the within-person mean for completed items was imputed before computing the total raw score. The total raw scores were converted to the continuous score (0–100) using the LIMB-Q scoring key.³² EQ-5D index scores for each dimension and the overall global score were computed using the UK normative values.⁴¹

STATISTICAL ANALYSIS

Psychometric analyses were performed on the final versions of the two LIMB-Q scales to assess the applicability of these scales across clinical conditions. Psychometric analysis used RMT analysis and followed COSMIN (CONsensus-based Standards for the selection of health Measurement INSTRUMENTs) standards for psychometric performance.⁴² RMT analysis examines how responses to items on a scale align with what is expected by the measurement model.⁴³ In the Rasch model, scales should be unidimensional, and items should map out a clinical hierarchy. When data fit the Rasch model, the scale has interval-level measurement properties, making it suitable for parametric statistics.⁴³ The RMT analysis was conducted with RUMM2030 software (RUMM version 2030, RUMM Laboratory Pty Ltd, Duncraig, Western Australia, Australia, 1998–2023),⁴⁴ using the unrestricted Rasch model for polytomous data.⁴⁵ Descriptive statistics, test–retest analysis, and construct validity analyses were performed using SPSS (Version 28; IBM Corporation, Armonk, NY, USA). The following psychometric tests were carried out:

- (1) Fit statistics: Three-item fit statistics—(1) item characteristic curves (IC), (2) item fit residuals (item–person interaction), and (3)

χ^2 values (item–trait interaction)—were performed to examine if the scale items measure the same underlying construct. IC were visually inspected. Fit residuals are considered ideal within ± 2.5 . Nonsignificant Bonferroni-adjusted χ^2 values indicate that the items fit the Rasch model.

- (2) Category threshold order: Threshold maps were examined to ensure that the response options functioned as intended, for example, lower values on the scale representing lower positions on the continuum.
- (3) Targeting: Person–item threshold distribution (item-map) and the proportion of participants scoring outside the scale were assessed graphically and statistically, to evaluate if the items within each scale effectively measured the construct experienced by the sample. Persons and item locations should be mirrored on the item-map for a scale to be targeted. The proportion of participants with extreme scores (ceiling or floor effects) and those skipping items (missing data) were calculated for each scale. A floor/ceiling effect was considered significant if $\geq 15\%$,⁴⁶ whereas the item-level missing data should be $< 10\%$.⁴⁷
- (4) Stability: Differential item functioning (DIF) was examined to determine if the items were responded to differently by different subgroups in the sample. We examined DIF by age group (< 35 vs. ≥ 35 years), gender (man vs. woman), and sample (current sample vs. LIMB-Q sample). Random samples were selected to ensure a balanced size in each group. We performed DIF for scales with at least 75 participants in each subgroup. The analysis was conducted three times for each scale to determine if any DIF detected was stable. Items with significant Bonferroni-adjusted χ^2 values were split by the relevant patient characteristics. To judge if the items with DIF impacted the scale scoring, we computed Pearson correlations for original and adjusted person locations. DIF by sample was examined to evaluate if the LIMB-Q scoring keys were valid in the study sample.
- (5) Dependency: Correlations between items in each scale were conducted to examine if a person's response to one item influenced the response to another item. Item pairs with correlations > 0.2 were included in

subtest analysis to assess their impact on scale reliability.

- (6) Reliability and reproducibility: Internal consistency was evaluated by Cronbach alpha.⁴⁸ Person Separation Index (PSI) was used to examine the scales' ability to differentiate between people in the sample. TRT was evaluated by one-sample *t*-test to determine if the mean difference in the scale scores between the first (T1) and the second (T2) assessment was different from zero and by intraclass correlation coefficients (ICCs) using the two-way mixed-effect model, based on average measurements as we deal with group statistics, evaluating for absolute agreement between scale scores.^{49,50} The mixed-effect model was chosen, as the time point was fixed. Absolute agreement was chosen, as agreement between scale scores at T1 and T2 and systematic differences in the participant's scores were of interest. Systematic bias is not included when evaluating for consistency. Only participants who indicated that they did not experience an important change in the construct measured by the scale between the T1 and T2 were included in the TRT analysis. Reliability coefficients ≥ 0.70 were considered acceptable.^{42,46}
- (6.1) Measurement error: The standard error of measurement (SEM) was conducted based on participants stating no important change between T1 and T2, using the formula $SEM_{\text{consistency}} = ((T1(SD) + T2(SD))/2) \times \sqrt{(1 - ICC_{\text{agreement}})}$. The SEM was converted into the smallest detectable change (SDC). The SDC was computed on a group level with the formula $SDC_{\text{group}} = (1.96 \times \sqrt{2} \times SEM) / \sqrt{n}$.^{46,51,52} The SDC_{group} indicates the minimum score change needed to ensure it is true and not a product of measurement error between groups. The SDC is based on the SEM, which is the error around each person's scale score. Regardless of the person's location on the scale's continuum (0–100), the SEM is assumed to be constant. However, scores at the floor and ceiling are less precise than scores at the center. Furthermore, the distribution of scale scores (standard deviation), which is dependent on the study population, affects the SEM. As a result of these limitations, the SDC is not recommended for

decision-making at the individual patient level.⁵¹

Construct validity: To examine convergent validity, correlations were computed between the scores of the Function, Symptoms, EQ-5D dimensions, EQ-5D global, and the EQ-VAS. COSMIN criteria state that correlations should be ≥ 0.50 for scales measuring similar constructs, $0.3\text{--}0.5$ for scales with related but dissimilar constructs, and <0.3 for scales with unrelated constructs.⁴² Hypotheses for convergent validity are available in Supplementary Appendix B. Furthermore, we hypothesized that the number of wounds would correlate <0.3 with the LIMB-Q Function and Symptom scores. Lastly, we hypothesized that participants with lower levels of mobility, measured by the EQ-5D Mobility dimension, would have lower scores on the Function scale. The last hypothesis of discriminative ability was examined by one-way analysis of variance (ANOVA) and *post hoc* analysis. Participants were classified into four mobility groups: no problems walking about, slight problems, moderate problems, and severe problems. Homogeneity of variance between the groups was tested by Levene's test. For construct validity analysis, parametric and non-parametric tests were used depending on data normality. Continuous data were inspected for normal distribution by QQ plots and statistically by skewness and kurtosis. Kurtosis and skewness should be within ± 2 for data to be considered normally distributed.⁵³

RESULTS

Qualitative phase

We performed 10 cognitive interviews with 5 men and 5 women. The mean age of the sample was 59.7 years (range 37 to 71 years). The sample included one person who had an active wound, two with a wound that had healed, and seven who had an amputation for a lower extremity wound. Most of the latter group had amputations because of a diabetic foot wound ($n = 6$). Participant demographics and quotes about relevance, comprehensibility, and comprehensiveness are in Supplementary Appendix C. Most participants found the content of the two scales relevant and easy to understand. A participant who did not have symptoms reported that the Symptoms scale was not relevant. The remaining participants found the two scales comprehensive and relevant to their lower extremity wound or subsequent amputation.

Quantitative phase

Study participants. Of the 421 participants recruited for the WOUND-Q study, 233 participants had a chronic lower extremity wound and completed one or both of the two new scales.²⁵ The sample consisted of 108 (46%) men, 122 (52%) women, and 3 (1%) people who identified as another gender. The mean age was 39.3 years (standard deviation [SD] 14.2 years, range 19–80 years). A total of 125 (54%) participants were ≥ 35 years old. Most people identified as Whites ($n = 156$, 67%), followed by Blacks ($n = 49$, 21%). In the sample, 66 (28.3%) participants currently smoked or vaped. A third ($n = 73$, 31%) reported no comorbidities, 86 (37%) reported 1, and 74 (32%) reported >1 . Dependency of a supportive device (e.g., crutches, prosthesis, wheelchair) to move around was reported by 41 (18%) participants. Most participants had one chronic wound (68%). The median self-reported wound size (width \times length) was 1.8 cm^2 . The leading cause for the wounds was trauma (35%), and for most participants ($n = 123$, 53%), the duration of the wound was between 3 and 6 months. Table 1 presents further demographic, clinical, and wound characteristics.

Psychometric findings. Psychometric data derived from the RMT analysis are presented in Table 2, Table 3, and Supplementary Appendix D and are described next.

Fit statistics and category threshold order: Data for the two scales fit the Rasch model. Of the 25 items tested, 24 had fit residuals within ± 2.5 . All items had nonsignificant χ^2 values after Bonferroni adjustment and ordered thresholds.

Targeting: Most of the sample scored within the measurement range of the scales, as illustrated in the person-item threshold maps in Supplementary Appendix E. The maps show the distribution of persons (upper histogram) and item locations (lower histogram) on the latent trait. As the histograms are mirrored, they provide evidence that the scales accurately measure the constructs as experienced by the sample. No ceiling or floor effect occurred, but most participants who scored outside the scale scored at the top of the scale (to the right side of the figure). All scales were acceptable in terms of missing data, which was $<10\%$.

Stability: No DIF was detected for age-group or gender for any item. DIF occurred between this sample and the LIMB-Q sample for four items in the Function scale and four items in the Symptoms scale, as seen in Supplementary Appendix D.

Table 1. Demographic and clinical characteristics of the 233 participants

	n	%
BMI		
(Mean; SD)	28.0	SD 7.9
Underweight <18.5	17	7.3
Normal weight 18.5–24.9	68	29.2
Overweight 25–29.9	70	30
Obese ≥30	75	32.2
Missing	3	1.3
Country		
Canada	7	3
Poland	8	3.4
Portugal	8	3.4
South Africa	44	18.9
The United Kingdom	82	35.2
The United States	58	24.9
Other	25	10.7
Prefer not to answer	1	0.4
Marital status		
Married, living common law	105	45.1
Never married	103	44.2
Divorced, separated	20	8.6
Other	5	2.1
Highest level of education		
Incomplete high school	4	1.7
High school	30	12.9
Incomplete trade school/college/university	48	20.6
Trade school/college/university	95	40.8
Incomplete masters/doctoral degree	19	8.2
Masters/doctoral degree	36	15.5
Prefer not to answer	1	0.4
Work		
Full time (≥37 h)	132	56.7
Part time (<37 h)	30	12.9
Unemployed	24	10.3
Student	7	3
Retired	13	5.6
Sick leave	8	3.4
More than one occupation	12	5.2
Other	6	2.6
Prefer not to answer	1	0.4
Comorbidities*		
Arthritis	29	12.4
Cancer	4	1.7
Chronic obstructive pulmonary disease (COPD)	8	3.4
Diabetes	69	29.6
Heart failure	5	2.1
Hypertension	58	24.9
Ischemic heart disease	9	3.9
Neuropathy	30	12.9
Paralysis	4	1.7
Peripheral artery disease (PAD)	15	6.4
Peripheral venous disease (PVD)	24	10.3
Skin disease (inflammatory and autoimmune)	6	2.6
Stroke	5	2.1
Other	33	14.2
Amputated		
Yes	68	29.2
Reason		
Chronic lower extremity wound**	3	1.3
Chronic lower extremity wound** + another reason	6	2.6

(continued)

Table 1. (Continued)

	n	%
Diabetes	15	6.4
Injury	14	6.0
Infection	8	3.4
Ischemic	4	1.7
>1 reason	9	3.9
Other reason	6	2.6
Not sure	3	1.3
Number of used supportive devices to be able to move around		
0	192	82.4
1	29	12.4
2	10	4.3
3	2	0.9
Type of supportive device used*		
Boot	6	2.6
Crutches	20	8.6
Walker	4	1.7
Leg prosthesis	6	2.6
Wheelchair	12	5.2
Other device	7	3.0
Wound location		
Leg***	96	41.2
Foot	70	30
Toe	31	13.3
Leg*** + foot	7	3
Foot + toe	10	4.3
Missing	19	8.2
Number of chronic wounds		
1	159	68.2
2	48	20.6
3	18	7.7
≥4	8	3.4
Wound type		
Arterial ulcer	8	3.4
Diabetic foot ulcer	37	15.9
Hidradenitis suppurativa	11	4.7
Pressure ulcer	7	3
Surgery	18	7.7
Trauma/injury	82	35.2
Venous ulcer	8	3.4
Multiple	30	12.9
Not sure	16	6.9
Other	16	6.9
Wound age		
3–6 months	123	52.8
7–12 months	31	13.3
1–2 years	36	15.5
3–4 years	16	6.9
5–10 years	22	9.4
11 to >30 years	5	2.1
Wound size (cm ²)		
<1	73	31.3
1–2.4	38	16.3
2.5–4.9	28	12
5–9.9	16	6.9
10–24.9	19	8.2
>24.9	31	13.3
Missing	28	12

*Able to choose multiple.

**From hip to toes.

***From hip to ankle.

SD, standard deviation.

Table 2. Function and Symptoms scales and scores

Scale	No. of items in scale	No. completed scale	Floor ^a (%)	Ceiling ^b (%)	Percent respondent with missing data ^c (%)	Scale scores (0–100)	
						Mean	Standard deviation
Function	15	227	4	11	9.1	56.1	25.2
Symptoms	10	225	0.4	11.6	6	63.3	20.3

^aPercent of respondents with 0 in score.^bPercent of respondents with 100 in score.^cPercent of respondents with at least one missing item (item-level).
No., Number.

The items with DIF did not impact the scale scoring, as all correlations between the original and the adjusted person locations were ≥ 0.998 .

Dependency: One pair of item in the Symptoms scales had residual correlations > 0.2 . In the Function scale, 12 items correlated with other items > 0.2 . The maximum drop in PSI after subtest analysis for correlated items in the two scales was 0.02, indicating that the residual correlations had a minor influence on scale reliability.

Reliability: Scale reliability was high for both scales, with PSI values ≥ 0.80 and Cronbach alpha ≥ 0.9 (Table 3).

Test-retest: For the TRT study, 89 participants with a lower extremity wound completed the Function and Symptoms scales. All participants completed the second assessment seven days after the first assessment. Owing to reported change, two participants were excluded from the analysis for TRT for the Function scale and six from the Symptoms scale. ICC values with and without outliers were ≥ 0.9 for the Function scale and ≥ 0.86 for the Symptoms scale. All ICC confidence intervals were ≥ 0.8 . Outliers were identified on boxplots. There was no significant difference between scale scores at T1 and T2 for the two scales using a one-sample *t*-test. TRT statistics, SEM, and SDC values are shown in Table 4.

Measurement error: Depending on the scale, the SEM values with outliers ranged from 6.8 to 7.2 and without outliers from 4.6 to 4.9. The minimal change that should occur in a score to be beyond measurement error at a group level (SDC_{group}) is listed in Table 4.

Construct validity: All hypotheses were confirmed. For the Function and Symptoms scale, higher scores were associated with higher health states on the EQ-5D. Lower scores on the Function scale were associated with increasing restrictions in walking ability and usual activities on the relevant EQ-5D dimensions. Lower scores on the Symptoms scale were associated with higher levels on the EQ-5D dimension that measures pain or discomfort. Further results of convergent validity are shown in Supplementary Appendix B. A higher number of wounds correlated weakly with worse scores for Function ($r = -0.21$ [$p < 0.002$]) and Symptoms ($r = -0.20$ [$p < 0.001$]) scales. The Function scale could discriminate the four mobility groups as the mean scores differed significantly between the four categories, Welch's ANOVA $F(3, 72.9) = 63.1$, $p < 0.001$. The Function scale scores worsened with higher walking restrictions (Fig. 2). Games–Howell *post hoc* analysis revealed that the mean score significantly differed between all possible group comparisons with *p*-values < 0.001 , except for the subgroups moderate and severe walking problems with a smaller but still significant *p*-value = 0.038.

DISCUSSION

This mixed-methods study examined the psychometric properties of the two LIMB-Q scales, Function and Symptoms, in 233 patients with chronic lower extremity wounds. The 10 cognitive interviews with patients with a history of a chronic wound provided evidence that the scales measured relevant concepts and that nothing

Table 3. RMT scale-level statistics

Scale	No. included in RMT	Scored on scale	χ^2	DF	p-Value	PSI + Extr	PSI – Extr	α + Extr	α – Extr	Residuals $> 0.20^a$
Function	193	85%	33.84	30	0.94	0.94	0.94	0.97	0.95	12 items (8 pairs)
Symptoms	198	88%	19.55	20	0.49	0.86	0.86	0.92	0.90	2 items (1 pair)

^aTotal items and number of item pairs with residuals > 0.2 . α , Cronbach alpha; DF, degrees of freedom; extr, extremes; No., number; PSI, Person Separation Index; RMT, Rasch Measurement Theory.

Table 4. Test-retest statistics and measurement error statistics

Scale	Excluded ^a (n)		Included (n)	T1 Mean (SD)	T2 Mean (SD)	Change Mean (SD)	One-sample t-test ^b	ICC ^c (95% CI)	SEM	SDC _{group}
Function	2	+ extr	87	53.5 (24.4)	54.4 (20.5)	−0.95 (13.8)	$p = 0.52$	0.90 (0.84–0.93)	7.23	2.15
		− extr	82	51.8 (22.2)	54 (20.9)	−2.23 (11)	$p = 0.07$	0.92 (0.87–0.95)	4.60	1.41
Symptoms	6	+ extr	83	60.8 (19.6)	60.5 (17.4)	0.28 (12.9)	$p = 0.85$	0.86 (0.79–0.91)	6.82	2.07
		− extr	77	59.2 (18.4)	60.7 (16.1)	−1.51 (9.5)	$p = 0.17$	0.92 (0.87–0.95)	4.94	1.56

^aOwing to change in scale construct.

^bTesting if mean change between T1 and T2 is different from zero, two-tailed p value.

^cAverage measure as we deal with group statistic.

CI, confidence interval; ICC, intraclass correlation coefficient; SDC, smallest detectable change; SEM, standard error of measurement; T1, first assessment; T2, second assessment, seven days later.

important was missing from them. Our field-test study provides evidence that the Function and Symptoms scales were psychometrically valid and reliable outcome measures in patients with all types of chronic lower extremity wounds.

The importance of PROMs for benchmarking, patient-centered care, research, and outcome monitoring in chronic wounds is well grounded.^{21,54,55} Numerous different condition-specific PROMs have been developed for wound patients, with most focused on specific types of wounds.^{22,24} Systematic reviews have identified a need for well-developed and valid PROMs for patients with diabetic foot or venous leg ulcers.^{56,57} The WOUND-Q is a promising PROM designed for all types of chronic wounds. In a review of 33 wound-specific PROMs, the WOUND-Q was recommended for use in research and clinical

care.²² Currently, this PROM is used in the statewide program “Leading Better Value Care” in New South Wales, Australia.⁵⁸

When incorporating new scales into a PROM, their psychometric performance must be examined in the target population to assess whether valid conclusions can be drawn from these measures. Both scales we tested fulfilled the demands of the Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) criteria on measurement properties.⁴² The RMT analysis revealed that each scale’s items successfully mapped out a clinical hierarchy covering the experiences and abilities of the population (target to sample). The LIMB-Q scoring key was found valid for use in chronic lower extremity wound patients as the items with DIF did not impact the scoring of the two scales. Having the

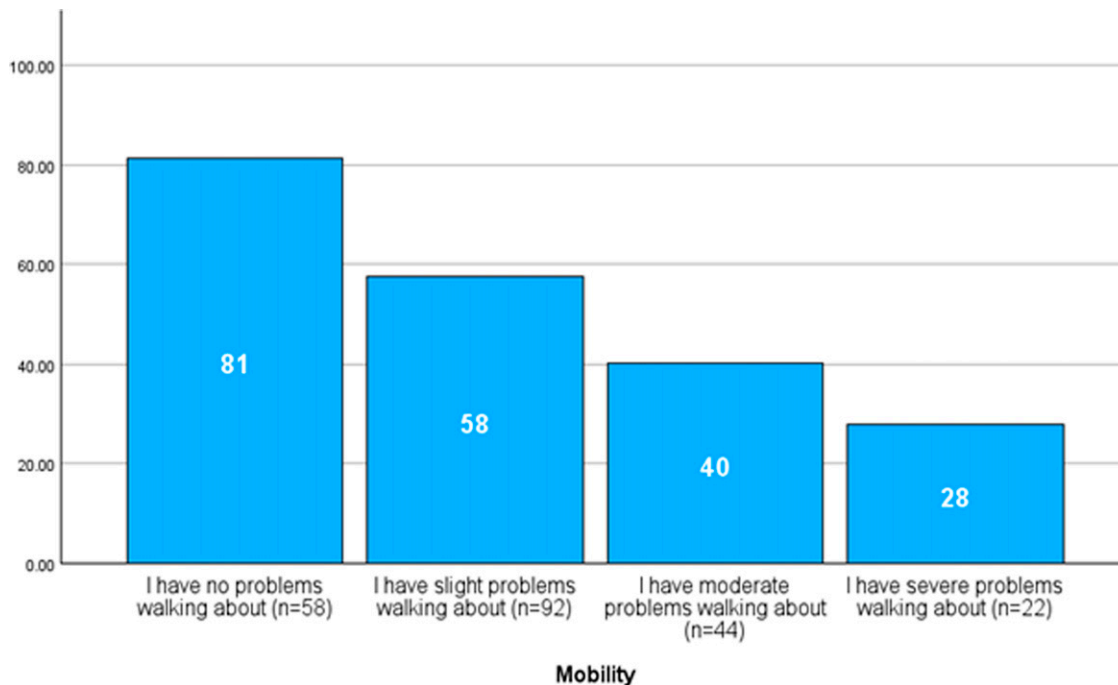


Figure 2. Mean score of the Function scale by four mobility groups. All mean scores differed significantly ($p < 0.05$) between the four groups based on Welch’s ANOVA and Games–Howell *post hoc* analysis. ANOVA, analysis of variance.

same scoring key makes it possible to directly compare patients with lower extremity wounds with patients undergoing reconstruction or amputations due to trauma. Regarding reliability, the Function scale had some items that evidenced local dependency, although this did not have an important impact on scale reliability. Based on ICC estimates, the scales were found applicable for decision-making at group level, as ICC estimates and 95% confidence interval (CI) were >0.7 .^{59,60} The SDC is a valuable tool to help interpret if there is a change in score outside of measurement error, but opposite to the minimal important difference (MID), it does not report if the patients find the change important.^{52,61} However, anchor-based MIDs do not consider measurement error, and to conclude that a change in score is clinically important and not a result of measurement error, the SDC should ideally be smaller than the MID.^{61–63}

This study has some methodological considerations. The cognitive interviews were performed at a single site in the United States, which can limit its generalizability. Also, the participants consisted mainly of people with prior wounds that had either healed or been managed with an amputation. It is possible there could have been recall bias when judging the relevance and comprehensiveness of the scales. We feel this is unlikely as participants had rich experiences with lower extremity wounds, and those with amputations would have gone through a lengthy attempt at limb salvage. Given the sample, our findings suggest that the scales apply to patients before and after amputation. Further research examining the finalized scales' content validity in a sample of patients with lower extremity wounds could confirm these findings. Having scales that are relevant to patients before and after amputation is essential, as amputations among patients with diabetes are preceded by a wound in more than 80% of cases.^{64,65} It has been found that amputations can improve mobility, reduce symptoms in the lower extremities, and enhance HRQL compared with living with an ulcer. However, the evidence is contradictory, and there is a need for more knowledge about the patient's perspective to guide decisions regarding

KEY FINDINGS

- The WOUND-Q is a rigorously developed PROM for all types of chronic wounds. However, it does not consist of scales assessing physical function and leg symptoms.
- This mixed-method study examined the psychometric properties of the scales Function and Symptoms in an international sample of patients with lower extremity wounds.
- The two scales were found relevant and easy to understand by patients with a current or history of a chronic lower extremity wound and in patients with amputation due to wounds.
- RMT analysis provided evidence for the two scales' reliability, validity, and interval-level measurement properties.
- The scores of the WOUND-Q Function and Symptoms scales can be compared with the LIMB-Q Function and Symptoms scores.
- The Function and Symptoms scales are integrated into the WOUND-Q and can be used separately in research, quality improvement initiatives, benchmarking, and patient care.

nonemergency lower extremity amputations.⁶⁶ The Function and Symptoms scales could provide important insights throughout this trajectory. The quantitative part of the study tested the scales in an international online sample. In using the online research platform Prolific, individuals self-select to participate, and all data were patient-reported, which we could not verify. Data from crowdsourcing platforms have been found reliable and valid,^{67–70} but participants were paid for their time, which might have motivated some to participate.^{71,72} Our chronic wound sample was relatively young, resulting in a higher proportion of people with traumatic wounds. However, a 50% increase in the proportion of younger persons <65 years with a wound was found in the United Kingdom between 2012 and 2018, stating that wounds are not only a condition in older people.² Lastly, the Function scale is only applicable to patients who can walk. Owing to the small sample size, DIF could not be examined for subgroups based on the type of assistive device used (e.g., boots, crutches, walker), neuropathy, or by treatment approach (e.g., amputation, reconstruction), factors that can impact balance and walking ability. However, no DIF was found in a sample of patients with lower extremity traumas who were treated with either fracture surgery or reconstruction/amputation, providing evidence that the Function scale works the same across these treatment types.³²

As wound closure can take months to years or may never be achieved, it is critical to provide wound management that can improve outcomes that are important to patients. By validating and adding the Function and Symptoms scales to the WOUND-Q, it is possible to measure how chronic

lower extremity wounds and their treatment affect patients from their perspective. Implementing these scales in clinical care may improve patient–provider communication, shared decision-making, and better treatment outcomes.^{20,21,55} In conclusion, we have demonstrated the psychometric properties of the Function and Symptoms scales in adults with chronic lower extremity wounds. The scales were well-targeted, reliable, and valid, making them suitable for integration into the WOUND-Q. The WOUND-Q now comprises 15 independently functioning scales with these additions. The WOUND-Q can be accessed at qportfolio.org/woundq/

ACKNOWLEDGMENTS AND FUNDING SOURCES

The authors want to acknowledge the patients' interest and participation in this study. Furthermore, they would like to thank Tanvee Singh and Paige Dekker, who recruited and conducted the cognitive interviews. A.F.K. received funds from the Department of Pediatrics, McMaster University, for data collection. Grants were given from Odense University Hospital (grant number A4774) and the Region of Southern Denmark (grant number 21/58368) to N.V.S.

AUTHORS' CONTRIBUTIONS

The authors confirm their contributions to the article as follows: study conception and design: N.V.S. and A.F.K.; data collection: N.V.S., A.F.K., C.R., L.R.M., and K.L.F.; analysis and interpretation of results: N.V.S., A.F.K., and C.R.; draft article preparation: N.V.S., L.R.M., A.F.K., C.R., and L.R.P. All authors reviewed the results and approved the final version of the article.

AUTHOR DISCLOSURE AND GHOSTWRITING

L.R.M. is a codeveloper of the LIMB-Q, and A.F.K. and A.L.P. are codevelopers of the

WOUND-Q and LIMB-Q and would receive a share of any license revenue in for-profit studies. A.F.K. provides research consulting services to pharmaceutical companies through EVENTUM Research. The rest of the authors have nothing to disclose.

ABOUT THE AUTHORS

Nina V. Simonsen, MD, PhD student, has since 2017 been involved in PRO research and connected to the Research Unit for Plastic Surgery, University of Southern Denmark (SDU) at Odense University Hospital. **Anne F. Klassen, DPhil**, professor in the Department of Pediatrics at McMaster University in Canada, Associate Member in the Departments of Surgery and HEI, a Scientist in CanChild, and an Adjunct Professor at the SDU, is involved in the development of the Q-Portfolio series of adult and pediatric PROMs. **Lotte Poulsen, MD, PhD**, associate professor at the Research Unit for Plastic Surgery, focuses on research in PROs and plastic surgery. **Charlene Rae, PhD**, research associate in the Department of Pediatrics at McMaster University, mainly focuses on developing and validating PROMs. **Lily R. Mundy, MD**, assistant professor of Plastic and Reconstructive Surgery at Johns Hopkins University, has a clinical and research focus in lower extremity reconstruction, oncologic breast reconstruction, and PROs. **Jens Ahm Sørensen, MD, PhD**, professor, and head of the Research Unit for Plastic Surgery, SDU, has a particular interest in microsurgery, head and neck, lymphedema, and extremity reconstruction. His research focuses on PROs, adipose-derived stem cells, head and neck, and lymphedema.

SUPPLEMENTARY MATERIAL

Supplementary Appendix A
Supplementary Appendix B
Supplementary Appendix C
Supplementary Appendix D
Supplementary Appendix E

REFERENCES

1. Sen CK, Gordillo GM, Roy S, et al. Human skin wounds: A major and snowballing threat to public health and the economy. *Wound Repair Regen* 2009; 17(6):763–771; doi: 10.1111/j.1524-475X.2009.00543.x
2. Guest JF, Fuller GW, Vowden P. Cohort study evaluating the burden of wounds to the UK's National Health Service in 2017/2018: Update from 2012/2013. *BMJ Open* 2020;10(12):e045253; doi: 10.1136/bmjopen-2020-045253
3. Martinengo L, Olsson M, Bajpai R, et al. Prevalence of chronic wounds in the general population: Systematic review and meta-analysis of observational studies. *Ann Epidemiol* 2019;29: 8–15; doi: 10.1016/j.annepidem.2018.10.005
4. Kirsli I, Elena C, Leanne A, et al. Lower leg ulcer diagnosis and principles of treatment. *Journal of Wound Management* 2023;24(2); doi: 10.35279/jowm2023.24.02.sup01
5. Probst S, Seppänen S, Gerber V, et al. EWMA Document: Home care-wound care: Overview, challenges and perspectives. *J Wound Care* 2014; 23 Suppl 5a:S1–S41; doi: 10.12968/jowc.2014.23.Sup5a.S1
6. Heyer K, Herberger K, Protz K, et al. Epidemiology of chronic wounds in Germany: Analysis of statutory health insurance data. *Wound Repair Regen* 2016;24(2):434–442; doi: 10.1111/wrr.12387

7. Alam W, Hasson J, Reed M. Clinical approach to chronic wound management in older adults. *J Am Geriatr Soc* 2021;69(8):2327–2334; doi: 10.1111/jgs.17177
8. Olsson M, Järbrink K, Divakar U, et al. The humanistic and economic burden of chronic wounds: A systematic review. *Wound Repair Regen* 2019; 27(1):114–125; doi: 10.1111/wrr.12683
9. Guest JF, Ayoub N, McIlwraith T, et al. Health economic burden that wounds impose on the National Health Service in the UK. *BMJ Open* 2015;5(12):e009283; doi: 10.1136/bmjopen-2015-009283
10. Green J, Jester R, McKinley R, et al. The impact of chronic venous leg ulcers: A systematic review. *J Wound Care* 2014;23(12):601–612; doi: 10.12968/jowc.2014.23.12.601
11. Kapp S, Santamaria N. The financial and quality-of-life cost to patients living with a chronic wound in the community. *Int Wound J* 2017;14(6): 1108–1119; doi: 10.1111/iwj.12767
12. Khunkaew S, Fernandez R, Sim J. Health-related quality of life among adults living with diabetic foot ulcers: A meta-analysis. *Qual Life Res* 2019; 28(6):1413–1427; doi: 10.1007/s11136-018-2082-2
13. Gonzalez-Consuegra RV, Verdu J. Quality of life in people with venous leg ulcers: An integrative review. *J Adv Nurs* 2011;67(5):926–944; doi: 10.1111/j.1365-2648.2010.05568.x
14. Herber OR, Schnepf W, Rieger MA. A systematic review on the impact of leg ulceration on patients' quality of life. *Health Qual Life Outcomes* 2007;5:44; doi: 10.1186/1477-7525-5-44
15. Hellström A, Nilsson C, Nilsson A, et al. Leg ulcers in older people: A national study addressing variation in diagnosis, pain and sleep disturbance. *BMC Geriatr* 2016;16:25; doi: 10.1186/s12877-016-0198-1
16. Gould LJ, Liu J, Wan R, et al. Evidence supporting wound care end points relevant to clinical practice and patients' lives. Part 3: The patient survey. *Wound Repair Regen* 2021;29(1):60–69; doi: 10.1111/wrr.12872
17. Crocker RM, Palmer KNB, Marrero DG, et al. Patient perspectives on the physical, psychosocial, and financial impacts of diabetic foot ulceration and amputation. *J Diabetes Complications* 2021;35(8):107960; doi: 10.1016/j.jdiacomp.2021.107960
18. Wukich DK, Raspovic KM. Assessing health-related quality of life in patients with diabetic foot disease: Why is it important and how can we improve? The 2017 Roger E. Pecoraro Award Lecture. *Diabetes Care* 2018;41(3):391–397; doi: 10.2337/dci17-0029
19. Health USDo, Human Services FDACfDE, Research, et al. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. *Health and Quality of Life Outcomes* 2006;4:79; doi: 10.1186/1477-7525-4-79
20. Gibbons C, Porter I, Gonçalves-Bradley DC, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. *Cochrane Database Syst Rev* 2021;10(10):CD011589; doi: 10.1002/14651858.CD011589.pub2
21. Williams K, Sansoni J, Morris D, et al. Patient-reported outcome measures Literature review ACSQHC: Available from: <http://www.safetyandquality.gov.au/>; 2016
22. van Alphen TC, Ter Brugge F, van Haren E, et al. SCI-QOL and WOUND-Q have the best patient-reported outcome measure design: A systematic literature review of PROMs used in chronic wounds. *Plast Reconstr Surg Glob Open* 2023; 11(1):e4723; doi: 10.1097/GOX.0000000000004723
23. Klassen A, van Haren EL, Cross K, et al. International mixed methods study protocol to develop a patient-reported outcome measure for all types of chronic wounds (the WOUND-Q). *BMJ Open* 2020;10(3):e032332; doi: 10.1136/bmjopen-2019-032332
24. Klassen AF, van Haren E, van Alphen TC, et al. International study to develop the WOUND-Q patient-reported outcome measure for all types of chronic wounds. *Int Wound J* 2021;18(4):487–509; doi: 10.1111/iwj.13549
25. Simonsen NV, Klassen AF, Rae C, et al. Further psychometric validation and test-retest reproducibility of the WOUND-Q. *Int Wound J* 2023;21(1): e14354; doi: 10.1111/iwj.14354
26. van Alphen TC, Poulsen L, van Haren ELWG, et al. Danish and Dutch linguistic validation and cultural adaptation of the WOUND-Q, a PROM for chronic wounds. *Eur J Plast Surg* 2019;42(5):495–504; doi: 10.1007/s00238-019-01529-7
27. Tsangaris E, van Haren EL, Poulsen L, et al. Identifying health-related quality of life concepts to inform the development of the WOUND-Q. *J Wound Care* 2024;33(1):28–38; doi: 10.12968/jowc.2024.33.1.28
28. Mundy LR, Klassen A, Grier J, et al. Development of a patient-reported outcome instrument for patients with severe lower extremity trauma (LIMB-Q): Protocol for a multiphase mixed methods study. *JMIR Res Protoc* 2019;8(10):e14397; doi: 10.2196/14397
29. Mundy LR, Klassen A, Grier AJ, et al. Identifying factors most important to lower extremity trauma patients: Key concepts from the development of a patient-reported outcome instrument for lower extremity trauma, The LIMB-Q. *Plast Reconstr Surg* 2020;145(5):1292–1301; doi: 10.1097/PRS.00000000000006760
30. Mundy LR, Grier AJ, Weissler EH, et al. Patient-reported outcome instruments in lower extremity trauma: A systematic review of the literature. *Plast Reconstr Surg Glob Open* 2019;7(5):e2218; doi: 10.1097/GOX.0000000000002218
31. Mundy LR, Klassen A, Sergesketter AR, et al. Content validity of the LIMB-Q: A patient-reported outcome instrument for lower extremity trauma patients. *J Reconstr Microsurg* 2020;36(9):625–633; doi: 10.1055/s-0040-1713669
32. Mundy LR, Klassen AF, Pusic AL, et al. The LIMB-Q: Reliability and validity of a novel patient-reported outcome measure for lower extremity trauma patients. *Plast Reconstr Surg* 2024; doi: 10.1097/PRS.00000000000011293
33. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: Attributes and review criteria. *Qual Life Res* 2002; 11(3):193–205.
34. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1—eliciting concepts for a new PRO instrument. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 2011;14(8):967–977; doi: 10.1016/j.jval.2011.06.014
35. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity—establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO Good Research Practices Task Force report: Part 2—assessing respondent understanding. *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 2011;14(8):978–988; doi: 10.1016/j.jval.2011.06.013
36. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value Health* 2005;8(2): 94–104; doi: 10.1111/j.1524-4733.2005.04054.x
37. EQ-5D. 2023. Available from: <https://euroqol.org/> [Last accessed: March, 21].
38. Cheng Q, Kularatna S, Lee XJ, et al. Comparison of EQ-5D-5L and SPVU-5D for measuring quality of life in patients with venous leg ulcers in an Australian setting. *Qual Life Res* 2019;28(7): 1903–1911; doi: 10.1007/s11136-019-02128-6
39. Yang Y, Brazier J, Longworth L. EQ-5D in skin conditions: An assessment of validity and responsiveness. *Eur J Health Econ* 2015;16(9):927–939; doi: 10.1007/s10198-014-0638-9
40. Iglesias CP, Birks Y, Nelson EA, et al. Quality of life of people with venous leg ulcers: A comparison of the discriminative and responsive characteristics of two generic and a disease specific instruments. *Qual Life Res* 2005;14(7):1705–1718; doi: 10.1007/s11136-005-2751-9
41. Devlin NJ, Shah KK, Feng Y, et al. Valuing health-related quality of life: An EQ-5D-5L value set for England. *Health Econ* 2018;27(1):7–22; doi: 10.1002/hec.3564
42. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018; 27(5):1147–1157; doi: 10.1007/s11136-018-1798-3

43. Pallant JF, Tennant A. An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 2007;46(Pt 1):1–18; doi: 10.1348/014466506x96931
44. Andrich D, Sheridan B. RUMM2030 (software program). Perth, Australia: RUMM Laboratory, 1997–2011.
45. Rasch G. Studies in mathematical psychology: 1. Copenhagen: Danmarks pædagogiske Institut.; 1960.
46. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; 60(1):34–42; doi: 10.1016/j.jclinepi.2006.03.012
47. The World Health Organization Quality of Life Assessment (WHOQOL): Development and general psychometric properties. *Soc Sci Med* 1998;46(12): 1569–1585; doi: 10.1016/s0277-9536(98)00009-4
48. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Med Educ* 2011;2(5):53–55; doi: 10.5116/ijme.4d4b.8d4d
49. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016;15(2): 155–163; doi: 10.1016/j.jcm.2016.02.012
50. Qin S, Nelson L, McLeod L, et al. Assessing test-retest reliability of patient-reported outcome measures using intraclass correlation coefficients: Recommendations for selecting and documenting the analytical formula. *Qual Life Res* 2019;28(4): 1029–1033; doi: 10.1007/s11136-018-2076-0
51. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technol Assess* 2009;13(12):iii, ix–ix-x, 1-177; doi: 10.3310/hta13120
52. de Vet HC, Terwee CB, Ostelo RW, et al. Minimal changes in health status questionnaires: Distinction between minimally detectable change and minimally important change. *Health Qual Life Outcomes* 2006;4:54; doi: 10.1186/1477-7525-4-54
53. Kim HY. Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod* 2013;38(1):52–54; doi: 10.5395/rde.2013.38.1.52
54. Health NMo. Leading better value care standards for wound management. 2019. Available from: https://aci.health.nsw.gov.au/__data/assets/pdf_file/0010/558352/NSW-Health-Wound-Standards-September-2019.PDF [November 8].
55. Quality AHSSa. Evidence on PROMs. 2023. Available from: <https://www.safetyandquality.gov.au/our-work/indicators-measurement-and-reporting/patient-reported-outcomes/evidence-proms> [November 8].
56. Romero-Collado À, Hernández-Martínez-Esparza E, Zabaleta-Del-Olmo E, et al. Patient-reported outcome measures of quality of life in people affected by diabetic foot: A psychometric systematic review. *Value Health* 2022;25(9):1602–1618; doi: 10.1016/j.jval.2022.04.1737
57. Poku E, Aber A, Phillips P, et al. Systematic review assessing the measurement properties of patient-reported outcomes for venous leg ulcers. *BJS Open* 2017;1(5):138–147; doi: 10.1002/bjs5.25
58. Innovation AfC. Chronic wound management. Available from: <https://aci.health.nsw.gov.au/statewide-programs/lbvc/chronic-wound-management> [November 8].
59. Lohr KN, Aaronson NK, Alonso J, et al. Evaluating quality-of-life and health status instruments: Development of scientific review criteria. *Clin Ther* 1996; 18(5):979–992; doi: 10.1016/s0149-2918(96)80054-3
60. Henrica CWdV, Terwee CB, Mokkink LB, et al. *Measurement in Medicine*. Cambridge University Press: 2011.
61. Terwee CB, Roorda LD, Knol DL, et al. Linking measurement error to minimal important change of patient-reported outcomes. *J Clin Epidemiol* 2009;62(10):1062–1067; doi: 10.1016/j.jclinepi .2008.10.011
62. van Kampen DA, Willems WJ, van Beers LW, et al. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *J Orthop Surg Res* 2013;8:40; doi: 10.1186/1749-799X-8-40
63. Jayadevappa R, Malkowicz SB, Wittink M, et al. Comparison of distribution- and anchor-based approaches to infer changes in health-related quality of life of prostate cancer survivors. *Health Serv Res* 2012;47(5):1902–1925; doi: 10.1111/j .1475-6773.2012.01395.x
64. In: Diabetic foot problems: prevention and management. London, National Institute for Health and Care Excellence (NICE); 2023.
65. Bus SA, Monteiro-Soares M, Game F, et al. Standards for the development and methodology of the 2023 IWGDF guidelines. *Diabetes Metab Res Rev* 2023;40(3):e3656; doi: 10.1002/dmrr.3656
66. Ong EKM, Murray C, Hillier S, et al. Decision-making processes for non-emergency diabetes-related lower extremity amputations: A scoping review. *Diabetes Res Clin Pract* 2023;206:111015; doi: 10.1016/j.diabres.2023.111015
67. Chandler J, Shapiro D. Conducting clinical research using crowdsourced convenience samples. *Annu Rev Clin Psychol* 2016;12:53–81; doi: 10.1146/annurev-clinpsy-021815-093623
68. Kelly JD, Comstock B, Kowalewski TM, et al. Crowd-sourced reliability of an assessment of lower facial aging using a validated visual scale. *Plast Reconstr Surg Glob Open* 2021; 9(1):e3315; doi: 10.1097/GOX.0000000000003315
69. Strickland JC, Stoops WW. The use of crowdsourcing in addiction science research: Amazon mechanical turk. *Exp Clin Psychopharmacol* 2019; 27(1):1–18; doi: 10.1037/pha0000235
70. Douglas BD, Ewell PJ, Brauer M. Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS One* 2023;18(3):e0279720; doi: 10.1371/journal.pone.0279720
71. Brühlmann F, Petralito S, Aeschbach LF, et al. The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology* 2020;2:100022; doi: 10.1016/j.metip.2020.100022
72. Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol Bull* 1999;125(6):627–668; discussion 692-700; doi: 10.1037/0033-2909.125.6.627

ABBREVIATIONS AND ACRONYMS

α	= Cronbach alpha
CI	= confidence interval
COPD	= chronic obstructive pulmonary disease
DF	= degrees of freedom
DIF	= differential item functioning
Extr	= extremes
HRQL	= health-related quality of life
IC	= item characteristic curves
ICC	= intraclass correlation coefficient
MID	= minimal important difference
No.	= number
PAD	= peripheral artery disease
PVD	= peripheral venous disease/peripheral venous disease
PRO	= patient-reported outcome
PROM	= patient-reported outcome measure
PSI	= Person Separation Index
REDCap	= Research Electronic Data Capture
RMT	= Rasch Measurement Theory
SD	= standard deviation
SDC	= smallest detectable change
SEM	= standard error of measurement
T1	= first assessment
T2	= second assessment
TRT	= test–retest reproducibility