

ORIGINAL ARTICLE

International study to develop the WOUND-Q patient-reported outcome measure for all types of chronic wounds

Anne F. Klassen¹ | Emiel L. W. G. van Haren² | Tert C. van Alphen³ |
Stefan Cano⁴ | Karen M. Cross⁵ | Anne-Margreet van Dishoeck⁶ |
Kenneth L. Fan⁷ | Maarten Michael Hoogbergen⁸ | Dennis Orgill⁹ |
Lotte Poulsen¹⁰ | Jens Ahm Sørensen¹¹ | Lee Squitieri^{12,13} | Elena Tsangaris⁹ |
Dalibor Vasilic¹⁴ | Andrea L. Pusic⁹

¹Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada

²Department of Plastic and Reconstructive Surgery, Catharina Ziekenhuis, Eindhoven, The Netherlands

³Department of Plastic and Reconstructive Surgery, Catharina Ziekenhuis and Da Vinci Kliniek, Geldrop, Eindhoven, The Netherlands

⁴Modus Outcomes, Letchworth Garden City, UK

⁵Department of Surgery, Division of Plastic Surgery, St. Michael's Hospital, University of Toronto, Toronto, Ontario, Canada

⁶Department of Plastic and Reconstructive and Hand Surgery, Erasmus MC University Medical Center, Rotterdam, The Netherlands

⁷Department of Plastic Surgery, Georgetown University School of Medicine, Plastic and Reconstructive Surgery, MedStar Health Research Institute, Washington, District of Columbia

⁸Department of Plastic and Reconstructive Surgery, Catharina Ziekenhuis, Eindhoven, and Da Vinci Kliniek, Geldrop, Eindhoven, The Netherlands

⁹Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts

¹⁰Research Unit for Plastic Surgery, Odense University Hospital, University of Southern Denmark, and Odense Explorative Patient Network, Odense, Denmark

¹¹Research Unit for Plastic Surgery, Odense University Hospital and University of Southern Denmark, Odense, Denmark

¹²RAND Corporation, Santa Monica, California

¹³Plastic and Reconstructive Surgery, Adventist Health White Memorial, Los Angeles, California

¹⁴Department of Plastic, Reconstructive and Hand Surgery, Erasmus Medical Center, Rotterdam, The Netherlands

Correspondence

Andrea L. Pusic, MD, MSc, Department of Surgery Brigham and Women's Hospital, 75 Francis S, Boston, MA 02116, Email: apusic@bwh.harvard.edu

Funding information

Plastic Surgery Foundation

Abstract

Patient-reported outcome measures (PROMs) for chronic wounds mainly focus on specific types of wounds. Our team developed the WOUND-Q for use with all types of wounds in any anatomic location. We conducted 60 concept elicitation interviews with patients in Canada, Denmark, the Netherlands, and the United States. Analysis identified concepts of interest to patients and scales were formed and refined through cognitive interviews with 20 patients and input from 26 wound care experts. Scales were translated into Danish and Dutch. An international field-test study collected data from 881 patients (1020 assessments) with chronic wounds. Rasch measurement theory (RMT) analysis was used to refine the scales and examine psychometric properties. RMT analysis supported the reliability and validity of 13 WOUND-Q scales that measure

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *International Wound Journal* published by Medicalhelplines.com Inc (3M) and John Wiley & Sons Ltd.

wound characteristics (assessment, discharge, and smell), health-related quality of life (life impact, psychological, sleep impact, and social), experience of care (information, home care nurses, medical team, and office staff), and wound treatment (dressing and suction device). The WOUND-Q can be used to measure outcomes in research and clinical practice from the perspective of patients with any type of wound.

KEYWORDS

patient-reported outcome, quality of life, questionnaire, Rasch, wounds

1 | INTRODUCTION

Chronic wounds, defined as wounds taking from 4 weeks to 3 months or longer to heal,¹ represent an underappreciated quality of life (QoL) and economic burden on individuals.² In the United States, over \$25 billion are spent treating 6.5 million patients with active wounds³ and in the United Kingdom £2.1 billion are spent on 2.2 million patients.⁴ Wound-related symptoms (eg, pain, discharge, and odour) can disrupt people's lives by interfering with the ability to maintain their usual activities, work, and social life.

Protocols for interventions to treat and prevent wounds have typically overlooked the patient perspective of outcomes and focused instead on objective measures (eg, time taken to heal, wound size).⁵⁻⁹ Although objective measures have been shown to be important to patients, they are not a comprehensive method of evaluating patient outcomes or experience of care.¹⁰ To address these gaps in chronic wound assessment, many patients, researchers, payers, and regulatory agencies have expressed increasing interest in developing measurement tools that adequately capture processes (experience) and outcomes of interest. Measuring outcomes from the patient perspective can be achieved using validated patient-reported outcome measures (PROMs).

There are currently a range of wound-specific PROMs available. Most PROMs focus on a specific type of wound, including venous leg ulcers,¹¹⁻¹⁷ foot ulcers,^{18,19} and pressure ulcers.^{20,21} The Wound-QoL represents an exception in that it was designed for all types of wounds.²² A limitation of the 17-item Wound-QoL is the lack of patient input into the content development and the limited range of outcomes (body, everyday life, and psyche) measured by its three scales. Each scale measures severity (not at all to very much) with a number of concepts that are added together to obtain a score. The body scale, for example, has four items that ask about symptoms (ie, pain, smell, exudate, and sleep impact), with a fifth item that asks whether treatment of the wound(s) has been a burden. The everyday scale combines items that ask

Key Messages

- The WOUND-Q is a patient-reported outcome measure for patients with all types of chronic wounds in any anatomic location
- WOUND-Q scales measure wound characteristics, health-related quality of life, experience of care, and treatments
- Rasch measurement theory was used to ensure each scale has interval-level measurement properties
- WOUND-Q scales, which are independently functioning, can be used in patient care, research, and quality improvement initiatives

about physical mobility (eg, moving about and climbing scales) with items that ask about social impact (eg, limitations in leisure activities and felt dependent on other people for help). The psyche scale asks about wound-related sadness, worry, frustration, and fear. A single item asks about the financial burden. In addition to scores for each scale, a global score is obtained from the 17 items. This approach to measurement can be unhelpful in clinical trials as it can mask the effect of treatment if patients are improving on some scales or items and not on others. It is also difficult to pinpoint the concepts that are improved by a treatment using scales and total scores that are multidimensional.

Given the tremendous impact that chronic wounds may have on multiple aspects of a patient's life, a comprehensive PROM for chronic wounds developed using a modern psychometric approach is needed.²³ Such a PROM should measure concerns specific to wounds (ie, not generic) and be designed to measure each important outcome separately so that changes in scores for interventions are easy to interpret. The aim of this international collaborative study was to develop and validate a

comprehensive set of independently functioning scales to measure outcomes important to patients with any type of chronic wound, as well as scales to measure patient's experience of wound care. The use of a modern psychometric approach (Rasch measurement theory [RMT] analysis²⁴) was used to ensure scales have interval (rather than ordinal) level measurement properties.

2 | METHODS

2.1 | Approach

We previously published a protocol paper that described our approach to develop the WOUND-Q.²⁵ Our multi-phase mixed methods approach adheres to published guidelines for the development and translation of a PROM.²⁶⁻³⁰ In phase 1 (qualitative), an interpretive description qualitative approach was followed, which built on existing clinical and research knowledge.³¹ The aim of phase 1 was to elicit concepts of interest to people seeking treatment for any type of chronic wound located in any anatomic location, and to use the findings to develop and refine a conceptual framework and a set of independently functioning scales. In phase 2 (quantitative), a modern psychometric approach was followed.^{24,32} The aim of phase 2 was to examine reliability and validity of each WOUND-Q scale and remove items and scales that were redundant or exhibited poor psychometric performance.

2.2 | Research ethics

Research ethics board approval was obtained from the coordinating site, Partners Human Research Committee, Brigham and Women's Hospital (Boston, Massachusetts), as well as sites in Canada (St. Michael's Hospital and the University of Toronto) and the United States (University of California, Los Angeles Berkley East Nursing Home; University of California, Los Angeles Medical Center; University of Southern California Keck Hospital, University of Southern California Verdugo Hills Hospital; MedStar Georgetown University Hospital and MedStar Washington Hospital Center). In the Netherlands, approval was obtained from the Dutch Medical Research Ethics Committee United (MEC-U) for the sites: Catharina Hospital (Eindhoven), DaVinci Wound Clinic (Geldrop), and Erasmus MC University Medical Center (Rotterdam). Finally, in Denmark, permission from the data protection agency was obtained for data collection at Odense University Hospital.

2.3 | Phase 1: Qualitative

2.3.1 | Sample and recruitment

A qualitative sample was recruited from wound clinics in Canada, Denmark, the Netherlands and the United States. Health care providers in each collaborating site were instructed to recruit participants who varied by age, gender, wound type, wound location, phase in healing, and risk of poor outcomes (eg, people who smoked or had comorbid conditions such as diabetes). Patients who expressed interest in the study were contacted by a qualitative interviewer with experience in qualitative methods to elicit concepts for PROM development. The interviewer set up and conducted the semi-structured interviews. Written and audio-recorded consent were obtained from each participant. Interviews took place by phone or in person, depending on participant's preference and logistics. Participants from the United States and Canada were provided with a small gift card to thank them for their time.

2.3.2 | Concept elicitation

We previously published our interview guide.²⁵ The guide included open-ended questions that aimed to elicit concepts of interest from patients about outcomes (eg, physical, psychological, and social life) and experiences (eg, information and wound care team) of wound care. As interviews progressed and participants discussed topics not on the interview guide, the guide was revised to include these new concepts for probing in subsequent interviews. Interviews were audio-recorded and transcribed verbatim. Danish and Dutch interviews were translated into English by professional translators and checked by the person who performed the qualitative interview.

The transcripts were coded using a line-by-line approach to label each concept of interest. The coded text was transferred into Excel alongside patient characteristics, including age, gender, wound type, and wound location. All codes were categorised into conceptual top-level domains and major/minor themes using constant comparison.³³ The top-level domains and major/minor themes were refined to form a conceptual framework.

An item pool was generated from the coded material and used to develop a set of scales that corresponded with the major concepts of interest identified in the data. Each scale was designed to map out a unidimensional construct composed of a set of items measuring a clinical hierarchy in line with the RMT approach.^{24,32} Each scale

was provided with instructions, a time frame for responding, and set of 4 response options.

Steps to ensure rigour in the qualitative phase included the following: concepts elicited during the interviews were confirmed in subsequent interviews; coding was performed by one member of the team and checked by a second team member; and regular team meetings were held to review the coding. In addition, at the half-way point, a one-day face-to-face research team meeting took place to identify any gaps in the make-up of the sample and to review codes, the item pool, and draft scales. After the meeting, interviews and analysis continued until redundancy of concepts elicited through the interviews was achieved.³⁴

2.3.3 | Scale development and refinement

Scales that were developed from the qualitative data were refined through cognitive interviews conducted with participants from the initial interviews. Participants were invited to review the instructions, response options and items of each scale. Interviews were conducted by a skilled qualitative interviewer. Each interview was audio-recorded, transcribed, and analysed to identify any changes needed. The think-aloud approach with probing was used to establish content validity of the WOUND-Q. The aim was to ensure that the instructions, response options, and items were in line with the consensus-based standards for the selection of health measurement instruments (COSMIN) criteria for developing PROM content that is comprehensible, comprehensive, and relevant.³⁵

Scales were also shown to wound care experts for feedback. A REDCap survey³⁶ was designed to collect feedback from experts known to the research team. Invitations to potential participant were sent via email in October 2017, with 1 reminder sent after 10 days. We also obtained face-to-face feedback from wound experts attending the Chronic Wound and Limb Salvage Conference in Washington in April 2018.

Elsewhere, we describe the methods and results for linguistic validation studies conducted to translate the scales into Danish and Dutch. The translation process also provided evidence of content validity of the WOUND-Q.³⁷

2.4 | Phase 2: Field-test study

2.4.1 | Sample and recruitment

The scales were field-tested in an international study that included sites in Canada, Denmark, the Netherlands, and

the United States. Participants were aged 18 years and older with at least 1 wound that had lasted 3 months or longer and were able to complete the WOUND-Q independently in either Danish, Dutch, or English. In Canada, the Netherlands, and the United States, recruitment took place in hospital inpatient and outpatient clinics using tablets or paper booklets. The United States and Canadian data were entered into a REDCap survey that was hosted at Brigham and Women's Hospital in Boston. The Dutch data were entered into a Castor database hosted at Catharina Hospital, Eindhoven. In Denmark, all patients from the wound clinic at Odense University Hospital were emailed a link to complete the WOUND-Q online in REDCap. Those that did not complete the survey prior to their wound care visit were invited to do so using a tablet in the outpatient clinic. Consent was obtained from participants within REDCap to take part in the study, and in the Netherlands and Denmark, participants were also consented to share their data outside of their respective countries.

Information about the wound (ie, size [length, width and depth], location, and type) was provided by Danish participants and by participants and health care providers for all other sites. Participants were asked their age, gender, weight, height, smoking status, and type and number of wounds. We also asked participants if in the past week their wound(s): (a) had drainage (yes, no); (b) a smell (none, faint, moderately strong, very strong); and (c) had interfered with sleep any nights in the week (none, 1-2, 3-4, 5-7). We asked if in the past 3 months participants: (d) had worked in a paid or volunteer job; (e) had a homecare nurse visit; (f) had visited a wound care clinic; and (g) had used a dressing on their wound. Finally, we asked if in the past 6 months participants: (h) had used a suction device to treat their wound. Branching logic was used with these 8 questions so that anyone who endorsed them was asked to complete associated scales. In addition to the WOUND-Q, participants completed the EQ-5D, a generic health utility measure that we scored using the US normative values.³⁸

2.4.2 | Analysis

Data from each site were merged in SPSS® version 26.0 (IBM Corporation, Armonk, New York for Windows®/Apple Mac®). RMT analysis to examine reliability and validity was performed using RUMM2030 software and the unrestricted Rasch model for polytomous data (RUMM version 2030, RUMM Laboratory Pty Ltd., Duncraig, Western Australia, Australia, 1998-2021). Scales that work have items that line up to map out a single continuum (ie, unidimensional construct). A range of

evidence was examined to determine how well the observed data for each scale fit the Rasch model. Specifically, the following statistical and graphical tests were performed to identify items and scales that did/did not work as hypothesized:

Item fit

Item response options were examined to determine if the item thresholds were properly ordered.³⁹ Three fit indicators were inspected and interpreted together and included graphical (item characteristic curves), statistical (log residuals (item-person interaction), and Chi-square (item-trait interaction) indicators of fit. Ideal fit residuals fall between -2.5 and $+2.5$ with non-significant Chi-square values after Bonferroni adjustment.⁴⁰ For tests of item fit, we amended the sample to 500.³²

Targeting

Items that form a scale should provide information for all levels of the concept experienced by the sample.³² Targeting was examined graphically (person-item threshold distribution) and statistically (proportion of the sample to score outside the range of each scale's measurement) to determine if the items were evenly spread over a reasonable range for each scale.

Differential item function (DIF)

DIF was examined for the following characteristics: age group (≤ 64 and ≥ 65 years), gender (male vs female), and language (English vs other). DIF was computed for scales where there were at least 150 completed scales in each subgroup to aim for 50 participants in each of 3 class intervals. We repeated the DIF analysis 3 times after selecting a random sample of equal size groups. Any item that evidenced significant DIF in any of the 3 analyses was split on the sample characteristic. We then conducted a Pearson correlation between the original and new person locations to determine the impact of DIF on scoring.⁴⁰

Reliability

We computed person separation index (PSI) and Cronbach alpha.⁴¹ Reliability coefficients ≥ 0.70 were considered acceptable.⁴²

Local independence

We examined item residual correlations to identify correlations above 0.30. Such values indicate that answers to an item may depend on answers to another item.⁴³⁻⁴⁵ We performed a substest to determine any drop in PSI value for correlations above 0.30.

Unidimensionality

The final item set for each scale was included in a principal component analysis in SPSS® version 26.0. We hypothesized that items in each scale would load onto a single factor and that factor loadings would be >0.70 to show that each item represents part of the latent variable measured by the scale.⁴⁶

Construct validity

The Rasch logits were used to transform participant scores for each scale from 0 (worse) to 100 (better). Normality was assessed using Kurtosis and Skewness and non-parametric statistics were applied if distributions were considered non-normal (outside of -2 to 2).⁴⁷ The scores were used to test hypotheses. First, correlations between WOUND-Q scales were computed. We hypothesized that WOUND-Q scale scores would correlate more strongly within their top-level domain (eg, HRQOL) than with scales in other top-level domains. Second, we expected that wound characteristics (ie, length [continuous], width [continuous], depth [continuous], discharge [yes/no], and smell [yes/no]) would correlate more strongly with wound, HRQOL, and treatment scale scores than with the experience of care scales. Third, in terms of convergent validity, we correlated EQ-5D and WOUND-Q scores and used the COSMIN criteria that correlations should be ≥ 0.50 for scales measuring similar constructs, 0.30-0.50 between scales measuring related but dissimilar constructs, and <0.30 between scales measuring unrelated constructs.⁴⁸ We expected that EQ-5D scores, which measure HRQOL, would correlate >0.50 with HRQOL scales, 0.30-0.50 with wound scales, and <0.30 with treatment and experience scales. Finally, we expected that scores on the outcome scales would be lower for participants recruited as inpatients compared with outpatients.

3 | RESULTS

3.1 | Phase 1: Qualitative findings

The qualitative phase took place between January 2016 and March 2017. We conducted 60 patient interviews. Table 1 shows the characteristics of the sample. Participants had a range of wound types in different anatomic locations that had lasted from 3 months to 25 years.

Analysis led to the development of a conceptual framework based on 2776 codes that covered 4 top-level domains: wound (726 codes); HRQOL (510 physical, 329 social, and 257 psychological codes); experience of care (572 codes); and treatment (382 codes). Appendix A

TABLE 1 Sample characteristics for the 60 qualitative and 881 field-test participants

	Phase 1		Phase 2	
	N	%	N	%
Country				
Canada	21	35	128	14.5
Denmark	12	20	299	33.9
The Netherlands	15	25	221	25.0
USA	12	20	233	26.4
Gender n, (%)				
Male	35	58	519	58.9
Female	25	42	357	40.6
Other	0	0	2	0.2
Missing	0	0	3	0.3
Age in years				
18-49	17	28	145	16.4
50-59	13	22	183	20.8
60-69	15	25	243	27.6
70-79	15	25	207	23.5
80-95	0	0	102	11.6
Missing	0	0	1	0.1
BMI				
Under/normal weight	15	25	263	29.9
Overweight	24	40	258	29.3
Obese	14	23	328	37.2
Missing	7	12	32	3.6
Number of wounds				
Healed	0	0	11	1.2
1	40	67	553	62.8
2	10	17	162	18.4
3+	10	17	155	17.6
Wound type or cause				
Diabetic foot ulcer	8	14	152	17.2
Surgical wound	9	15	142	16.1
Pressure ulcer	15	25	130	14.8
Venous Ulcer	12	20	111	12.6
Trauma/injury	1	2	96	10.9
Arterial ulcer	0	0	31	3.5
Hidradenitis	5	8	23	2.6
Radiation necrosis	2	3	14	1.6
Infection	0	0	9	1.0
Cancer	0	0	7	0.8
Pyoderma gangrenosum	1	2	6	0.7
Pilonidal abscess	2	3	6	0.7
Other /unknown	3	5	136	15.5
Multiple	2	3	10	1.1
Missing	0	0	8	0.9
Wound location				
Leg or knee	17	28	270	30.6
Foot, ankle	14	24	227	25.8

TABLE 1 (Continued)

	Phase 1		Phase 2	
	N	%	N	%
Toe(s)	0	0	70	7.9
Buttocks	12	20	62	7.0
Abdomen, genitals, chest or back	4	8	59	6.7
Arm, shoulder, armpit, hand	0	0	18	2.1
Face or neck	0	0	6	0.7
Other	0	0	5	0.6
Multiple	0	0	121	13.7
Missing	0	0	43	4.9
Wound age n, (%)				
3-6 months	14	23	318	36.1
7-12 months	13	22	166	18.8
1-2 years	16	27	181	20.5
3-4 years	6	10	88	9.9
5-10 years	5	8	49	5.6
More than 10 years	6	10	36	4.1
Missing	0	0	43	5.0
Wound size (length × width), cm				
<0.1	7	12	206	23.4
1-2.4	15	25	146	16.6
2.5-4.9	10	17	98	11.1
5-9.9	14	24	94	10.7
10-24.9	10	17	136	15.4
≥25	0	0	138	15.7
Missing	3	5	63	7.2

shows the saturation table for the conceptual model for codes that were endorsed by 10 or more participants.

We generated an item for each code related to themes forming the conceptual framework and developed 16 independently functioning scales. Each scale was assigned instructions, a time frame for responding, and 4 response options that measured either severity (assessment, smell, drainage, life impact), frequency (sleep, psychological), agreement (social, work, wound acceptance, home nurses, team, clinic, office staff), or satisfaction (information, bandage, suction device).

We performed 20 patient cognitive interviews with participants from the initial interviews, 15 in round 1 (September 2017) and 5 in round 2 (March 2018). Only minor changes were needed to response options, wording, and items within the scales to ensure that they were comprehensible, comprehensive, and relevant from the patient perspective.

Between rounds of cognitive interviews, the RED-Cap survey of experts was completed by 12 plastic surgeons, 4 vascular surgeons, 2 general surgeons, and

3 nurse practitioners. Experts were based in Australia, Belgium, Canada, Denmark, Germany, the Netherlands, the United States, and the United Kingdom. An additional 5 plastic surgeons who attended a wound conference took part in one-on-one interviews. Based on expert input, 4 items were removed and the t2 treatment scales were added (ie, dressing and suction device). The scales were judged easy to understand, relevant, and comprehensive.

The finalised WOUND-Q was provided to translation teams in Denmark and the Netherlands who performed translation and cultural adaptations.³⁷ The translation process included input from 12 experts (6 Danish and 6 Dutch) and 38 patients (22 Danish and 16 Dutch) with 11 different types of wounds. The experts included 5 translators and 7 health care providers (2 nurses, 4 surgeons, and 1 physician) with clinical expertise in wound care. Expert input and patient cognitive debriefing interviews provided further evidence of content validity. Feedback led to 6 items difficult to translate into either Danish or Dutch being dropped from the WOUND-Q.

Further translation and cultural adaptation results are published elsewhere.³⁷

3.2 | Phase 2: Psychometric findings

The field-test version of the WOUND-Q included 16 scales with 221 items. The field-test study took place between August 2018 and May 2020. Characteristics for the 881 participants who provided 1020 assessments are shown in Table 1. Recruitment was highest in Denmark and lowest in Canada. The mean age of the sample was 62.8 (SD = 14.5) and 58.9% were male. Most participants were a current ($n = 114$, 12.9%) or former (367, 41.7%) smoker. Most participants (719, 81.6%) reported that they had one or more health conditions. The most common health conditions were diabetes ($N = 348$, 39.5%), peripheral vascular disease ($N = 202$, 22.9%), a function disability such as foot/leg amputation, paraplegia or quadriplegia ($N = 127$, 14.4%), high blood pressure ($N = 62$, 7.0%), and arthritis ($N = 43$, 4.9%).

Most participants ($N = 553$, 62.8%) had one wound, and most ($N = 567$, 64.4%) had a wound on their leg, foot, ankle or toe(s). The majority of participants ($N = 632$, 71.7%) reported having drainage from their wound in the past week. Of the 2581 (29.3%) to report that their wound smelled in the past week, 33 (3.7%) indicated the smell was strong, 59 (6.7%) moderately strong, and 166 (18.8%) faint. Participants reported their wound interfered with sleep 1-2 nights a week ($N = 209$, 23.7%), 3-4 nights for ($N = 100$, 11.4%), and 5-7 nights ($N = 136$, 15.4%). In the past 3 months, 247 (28.0%) participants had worked in a paid or volunteer job, 811 (92.1%) had used dressing on their wound, 677 (76.8%) had been to a wound care clinic at least once (not their first visit), and 449 (51.0%) had a home care nurse visit. Use of a suction device on the wound in the past 6 months was reported by 190 (21.6%) participants.

RMT analysis provided evidence of reliability and validity for 13 scales. The 3 exceptions were 2 HRQOL scales (work, wound acceptance) and 1 experience scale (wound clinic). All 3 of these scales had multiple items with disordered thresholds. After we rescored each scale to reduce response options by 1 threshold and dropped items with poor fit, scale reliability was low ($PSI < 0.70$). These scales were dropped from the WOUND-Q.

RMT analysis reduced items in the remaining 13 scales from 190 to 111. Scale level findings for the 13 scales are shown in Table 2 and item fit statistics are in Appendix B. Of the 13 scales, 4 were completed by all participants (assessment, life impact, psychological, social), with the remainder by the subgroup for whom each scale was relevant. To improve targeting and fit of

the data to the Rasch model, for 3 scales (social, dressing, and information), the RMT analysis was performed on participants with wounds ≥ 2.5 cm in length.

All 111 items had ordered thresholds and non-significant Chi-square P -values after Bonferroni adjustment. Item fit was within $+2.5$ for 100/111 items. Data fit the Rasch model for 9 scales (nonsignificant P -values). The other 4 scales showed slight misfit to the Rasch model. Appendix C shows the distribution of person measurement (top histogram) and item locations (bottom histogram) for 1 scale for each domain to illustrate targeting, that is, wound (assessment), HRQOL (life impact), experience of care (information), and treatment (suction device). The proportion of participants to score within the measurement range provided by each of the 4 scales ranged from 88.6% (assessment) to 70.9% (information). Those that scored outside the range tended to score at the high end of the scale (to the right in each figure). Table 2 shows other targeting statistics (floor and ceiling effects) and the proportion of eligible participants who did not complete the scale (missing data).

Appendix B shows the DIF results. The sample size was less than 150 in 1 or both subgroups in 3 scales for gender and language and in 2 scales for age. DIF was detected for 14 items, including 3 of 94 items tested for age, 3 of 84 items tested for gender, and 9 of 84 items tested for language. Pearson correlations between person locations for items before and after splitting the items for DIF indicated marginal impact on scoring (all correlations ≥ 0.996).

PSI values for the wound, HRQOL and treatment scales were ≥ 0.72 (see Table 2). PSI values without extreme values were higher (≥ 0.81) for the experience scales than with extreme values (≥ 0.52) because of their ceiling effects. Cronbach alpha values for the 13 scales were ≥ 0.89 (with extremes) to ≥ 0.80 (without extremes). Residuals in 1 or 2 item pairs in 8 scales were correlated above 0.30. The subtest performed to examine the impact of correlations on the PSI values represented a maximum drop of reliability of 0.04 (office staff scale). The principal component analysis results for each scale supported a single factor with factor loadings that ranged from 0.69 to 0.95.

Table 3 shows correlations among WOUND-Q scales, clinical variables, and the EQ-5D for the sample of 881 participants. With some exceptions, WOUND-Q scales tended to correlate higher with scales in their own top-level domain than with scales in other domains. As predicted, wound, HRQOL, and treatment scale scores were more strongly correlated with the clinical characteristics than were the experience scales. Lower scores on the wound and HRQOL scales were generally associated with wounds that were longer, wider, deeper, and had

TABLE 2 Rasch measurement theory scale level statistics

Concept	Scale	N items tested	N final items	N completed in RMT scale	% scored on RMT scale	Chi-square	DF	P-value	PSI + extr	PSI - extr	Cronbach alpha + ext	Cronbach alpha - ext	% floor	% ceiling	Missing data
Wound	Assessment	16	11	1008	893	88.6	91.9	88	.37	0.84	0.83	0.89	1.0	9.7	1.2
	Drainage	10	8	658	574	87.2	74.9	56	.05	0.78	0.76	0.86	0.9	11.6	5.6
	Smell	9	8	259	211	81.5	22.6	16	.12	0.86	0.84	0.91	5.0	13.1	7.8
HR-QOL	Life impact	13	8	998	824	82.6	53.7	72	.95	0.86	0.84	0.89	3.3	13.2	2.2
	Psychological	17	10	1001	822	82.1	57.0	90	.99	0.85	0.84	0.95	1.3	15.8	1.9
	Sleep	8	5	461	397	86.1	29.7	20	.07	0.88	0.84	0.88	9.3	3.9	6.7
	Social	11	5	451	302	67.0	32.5	20	.04	0.80	0.72	0.80	6.8	27.3	3.8
Treatment	Dressing	9	9	388	262	67.5	49.4	27	.01	0.72	0.75	0.82	0.2	34.4	11.3
	Suction device	15	9	175	126	72.0	18.7	18	.41	0.87	0.87	0.92	1.7	21.1	23.9
Experience of care	Information	19	10	337	239	70.9	43.0	30	.06	0.80	0.84	0.91	0.3	27.5	8.8
	Homecare	21	10	491	226	46.0	52.3	30	.01	0.71	0.83	0.90	0.6	52.7	6.5
	Wound team	28	10	707	241	34.1	43.1	30	.06	0.52	0.81	0.91	0.1	62.6	10.0
	Office staff	14	8	662	180	27.2	37.9	16	.00	0.61	0.82	0.90	0.8	68.4	15.7

Abbreviations: N, number; RMT, Rasch Measurement Theory; DF, degrees of freedom; PSI, person separation index; ext, extremes.

TABLE 3 Spearman correlations between WOUND-Q scales, EQ-5D, and clinical variables

	Assessment	Drainage	Smell	Life Impact	Psychological	Sleep	Social	Dressing	Suction Device	Information	Homecare	Wound Team	Clinic Staff
Drainage	.66 ^a												
Smell	.50 ^a	.64 ^a											
Life Impact	.58 ^a	.49 ^a	.38 ^a										
Psychological	.67 ^a	.52 ^a	.44 ^a	.66 ^a									
Sleep	.41 ^a	.34 ^a	.36 ^a	.42 ^a	.43 ^a								
Social	.49 ^a	.48 ^a	.37 ^a	.78 ^a	.63 ^a	.38 ^a							
Dressing	.32 ^a	.36 ^a	.29 ^a	.29 ^a	.33 ^a	.20 ^a	.27 ^a						
Suction Device	.23 ^a	.26 ^a	.20	.28 ^a	.32 ^a	.32 ^a	.26 ^a	.56 ^a					
Information	.18 ^a	.23 ^a	.11	.18 ^a	.24 ^a	.16 ^a	.13 ^a	.46 ^a	.28 ^a				
Homecare	.12 ^b	.10	-.02	.11 ^b	.22 ^a	.04	.14 ^a	.34 ^a	.29 ^a	.30 ^a			
Wound Team	.11 ^a	.12 ^b	.05	.05	.15 ^a	.10	.05	.42 ^a	.27 ^a	.57 ^a	.35 ^a		
Clinic Staff	.11 ^a	.12 ^b	.12	.08	.14 ^a	.19 ^a	.06	.38 ^a	.29 ^a	.49 ^a	.34 ^a	.59 ^a	
EQ-5D	.44 ^a	.38 ^a	.36 ^a	.57 ^a	.50 ^a	.46 ^a	.54 ^a	.31 ^a	.15 ^a	.17 ^a	.15 ^a	.16 ^a	.18 ^a
Length	-.18 ^a	-.24 ^a	-.17 ^a	-.21 ^a	-.17 ^a	-.06	-.22 ^a	-.11 ^a	-.03	-.03	-.03	.06	.08
Width	-.19 ^a	-.21 ^a	-.18 ^a	-.18 ^a	-.16 ^a	-.09	-.18 ^a	-.10 ^a	.06	-.04	.02	.08	.07
Depth	-.11 ^a	-.12 ^a	-.07	-.14 ^a	-.06	.00	-.17 ^a	-.07	-.08	-.03	-.00	-.02	-.02
Smell	-.39 ^a	-.43 ^a	NA	-.26 ^a	-.28 ^a	-.13 ^b	-.25 ^a	-.22 ^a	-.01	-.08	-.09	-.05	-.05
Drainage ^c	-.36 ^a	NA	-.13 ^b	-.22 ^a	-.23 ^a	-.07	-.21 ^a	-.17 ^a	-.28 ^a	-.01	-.09	-.02	-.02

^aCorrelation is significant at the .01 level (2-tailed).

^bCorrelation is significant at the .05 level (2-tailed).

^cExcluded participants who answered "not sure."

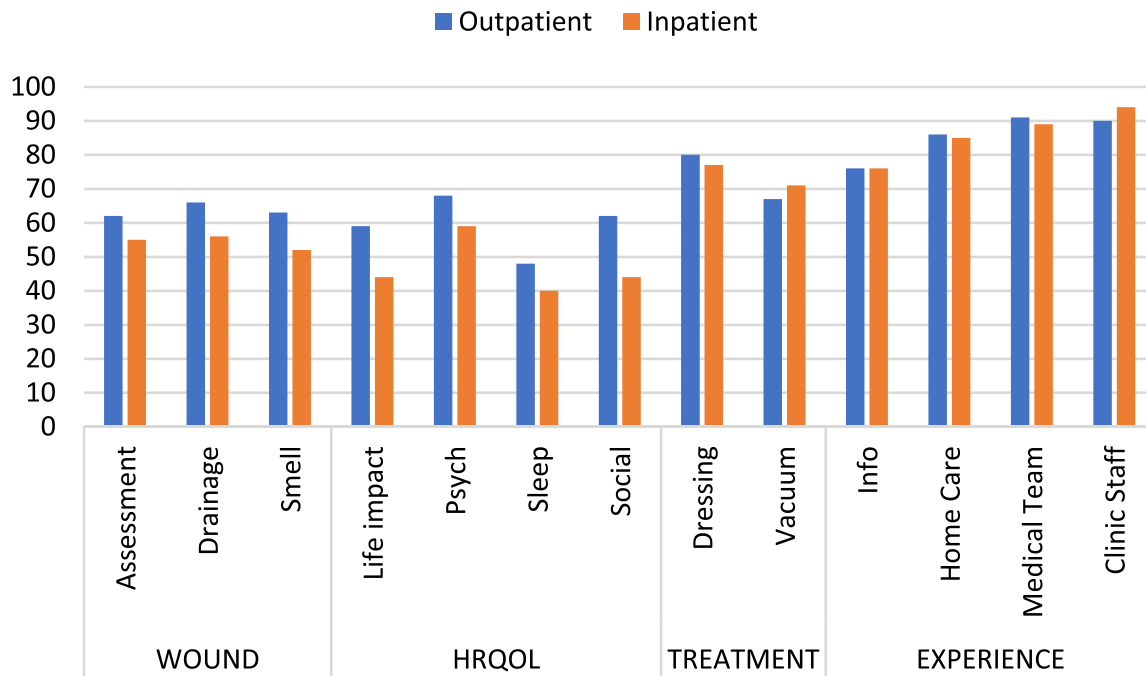


FIGURE 1 Mean scores on WOUND-Q scales comparing inpatients and outpatients

discharge or a smell. In the treatment domain, lower scores for dressing correlated with worse outcomes for most of the wound characteristics and low scores for the suction device scale correlated with having drainage. Our expectation that correlations between the WOUND-Q and EQ-5D scores would be >0.50 for the HRQOL scales, 0.30 - 0.50 for wound scales, and <0.30 with treatment and experience scales also received support with 11 of 13 correlations meeting the COSMIN criteria.

Figure 1 shows the WOUND-Q scale scores comparing inpatients ($N = 117$, 13.3%) with outpatients ($N = 761$, 86.4%). Mean scores differed significantly for all outcome scales ($P < .05$ on independent samples t tests). For the experience and treatment scales, no significant differences were found between the 2 groups. The characteristics of the subgroups for these tests of construct validity can be found in Appendix D.

4 | DISCUSSION

Each year, millions of individuals around the world require treatment for acute (eg, surgical and traumatic wounds, abrasions, or superficial burns) and chronic (eg, venous ulcers, diabetic, and pressure ulcers) wounds.⁴⁹ Such wounds, which have many causes and numerous treatments, can have a tremendous impact on HRQOL of patients and have a huge economic impact on health care systems.⁴⁹⁻⁵¹ The development of the WOUND-Q fills an important unmet need and can be expected to

significantly improve our ability to measure treatment outcomes from the patient perspective.

A key strength of the WOUND-Q development was the rigorous qualitative methods used with extensive input from patients and wound care experts to establish content validity. Patients with chronic wounds provided rich descriptive detail in the qualitative accounts of their concerns about their wound(s) and its impact on their HRQOL and experience of wound care. We created 16 scales covering outcomes and experiences of care identified in the qualitative data. An additional strength is the large international field-test which facilitated robust psychometric analysis of the preliminary scales. An advantage of RMT analysis using RUMM 2030 software is the ability to diagnose problems in the functioning of scales. We found that three scales did not work psychometrically and were therefore dropped from the WOUND-Q. Evidence from the RMT analysis provided strong support for the reliability and validity of the remaining 13 scales. These scales covered a range of topics important to patients. Our finding that DIF by age, gender, and language did not impact scoring provides support that the WOUND-Q scales can be used internationally in research, clinical practice, and quality improvement efforts to supplement objective outcome measures.

The WOUND-Q is the only wound-specific PROM with a modular structure providing a comprehensive set of independently functioning scales that cover important concepts from the patient perspective. Given that many patients with chronic wounds are elderly, it is critical to

reduce survey burden of PROM completion. WOUND-Q scales range in length from 5 to 11 items. The modular approach of the WOUND-Q means that patients, clinicians, researchers, and other stakeholders do not have to use them all, but instead can select a subset of scales using only those that are important to a specific context of use. For example, of the wound and HRQOL scales, 4 (assessment, life impact, psychological, and social) are relevant to all patients and three (drainage, smell, and sleep) are relevant to the subset of patients who report a problem in these areas.

Although the scales measuring outcomes evidenced good targeting and distribution of scores, the patient experience scales evidenced high ceiling effects. This finding highlights the challenge associated with developing scales to measure aspects of quality. Participants were very satisfied with the care provided by the medical team, home care nurses, and office staff at the collaborating centres. The scale measuring satisfaction with information showed the most variability in experience of care scale scores. To test differences in patient experience with providers, we would ideally have included a larger number of different types of wound-care sites (eg, academic and non-academic) with a range of wound care teams.

As noted above, the strength of our study was that we followed rigorous guidelines for PROM development and translation.²⁶⁻³⁰ Both the qualitative and quantitative samples included patients from different countries and health care systems. Both samples also included patients with a wide variety of wounds that varied in size, location, and age of wound. In the qualitative phase, we were able to establish content validity of the scales by patient and expert input from multiple countries prior to finalising the WOUND-Q. Given the large sample in the quantitative phase, we were able to examine and confirm that the scales worked the same (no bias) by language for 10/13 scales. Future translations will need to follow best practice guidelines³⁰ to ensure that additional linguistic validation/cultural adaptations of the WOUND-Q are done appropriately.

Our study had some limitations. Recruitment of the Danish sample through email meant that the clinical information for that country was patient-reported. Although a sizable number of Danish participants were not sure what type of chronic wound they had, the dataset from Denmark was large, which helped to ensure that different types of wounds were well represented. Also, some participants in the Danish sample may have had more than one type of wound, but, unlike other participants, were not provided the opportunity to select multiple answers. A limitation of the DIF analysis was that WOUND-Q scales with a particularly high ceiling effect (homecare, wound team, and office staff) had a

smaller sample size, because extreme scores are not included in Rasch analysis. Future research to explore DIF is warranted. The WOUND-Q field-test was long and some participants did not make it to the end of the survey. It is not possible to know if participants ended the survey prematurely because of fatigue, unfamiliarity with use of a tablet, technical issues, being called into the appointment, or another reason. Some psychometric properties were not examined in our cross-sectional study, including responsiveness and test-retest reliability and should be the focus of future studies.

5 | CONCLUSION

The WOUND-Q was designed for patients with all types of chronic wounds in any anatomic location. This modular PROM measures wound characteristics, HRQOL, satisfaction with treatments, and experience of care. This work builds on our team's previous experience with condition-specific PROM development, including the BREAST-Q,⁵² FACE-Q,⁵³ and BODY-Q.⁵⁴ The strength of our study includes extensive involvement of patients and wound care experts, a large international field-test, and the use of RMT analysis to ensure that each scale was unidimensional with interval-level measurement properties. The WOUND-Q is a powerful new measurement tool that can be used to bring the patient perspective to evidenced-based clinical care, quality metrics, value-based health care, and regulatory decisions in chronic wound care.

ACKNOWLEDGEMENTS

Phases I and II of this study were funded by research grants received from the Plastic Surgery Foundation. The authors have no financial interest to declare in relation to the content of this article. The authors thank the patients and health care staff who helped with the study at Brigham and Women's Hospital, St. Michael's Hospital, and Georgetown University School of Medicine. Lauren Bayer, Philip Brower, Mike Carr, Anouk Claessen, Seray Er Jessica Figueroa, Paul Hunter, Joanna O'Gorman, Lisa Van Den Berg, Anamika Veeramani, and Jonanna Woodman were instrumental in data collection at each site.

CONFLICT OF INTEREST

The WOUND-Q is owned by Memorial Sloan-Kettering Cancer Center and Mass General Brigham. Drs Pusic and Klassen are co-developers of the WOUND-Q and could potentially receive a share of any license revenue on the inventor sharing policies of the institutions that own them. The other authors declared no potential conflicts of interest.

DATA AVAILABILITY STATEMENT

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

REFERENCES

1. https://www.facs.org/-/media/files/education/core-curriculum/nonhealing_wounds.ashx
2. 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.
3. 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.
4. Guest J, Ayoub N, McIlwraith T, et al. Health economic burden that different wound types impose on the UK's National Health Service. *Int Wound J.* 2017;14(2):322-330.
5. Cochrane Wounds. Cochrane Ireland and Cochrane UK symposium 2020. <https://wounds.cochrane.org/>. Accessed September 2020.
6. Norman G, Westby MJ, Rithalia AD, Stubbs N, Soares MO, Dumville JC. Dressings and topical agents for treating venous leg ulcers. *Cochrane Database Syst Rev.* 2018;6:CD012583.
7. Kranke P, Bennett M, Martyn-St James M, Schnabel A, Debus SE, Weibel S. Hyperbaric oxygen therapy for chronic wounds. *Cochrane Database Syst Rev.* 2015;6:CD004123.
8. Dumville JC, Webster J, Evans D, Land L. Negative pressure wound therapy for treating pressure ulcers. *Cochrane Database Syst Rev.* 2015;5:CD011334.
9. Smith F, Dryburgh N, Donaldson J, Mitchell M. Debridement for surgical wounds. *Cochrane Database Syst Rev.* 2013;9:CD006214.
10. Gould LJ, Liu J, Wan R, Carter MJ, Dotson M, Driver VR. 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.
11. Hyland ME, Ley A, Thomson B. Quality of life of leg ulcer patients: questionnaire and preliminary findings. *J Wound Care.* 1994;3:294-298.
12. Hareendran A, Doll H, Wild DJ, et al. The venous leg ulcer quality of life (VLU-QoL) questionnaire: development and psychometric validation. *Wound Repair Regen.* 2007;15:465-473.
13. Smith JJ, Guest MG, Greenhalgh RM, Davies AH. Measuring the quality of life in patients with venous ulcers. *J Vasc Surg.* 2000;31:642-649.
14. Venkatraman PD, Anand SC, Dean C. Pilot study investigating the feasibility of an ulcer-specific quality of life questionnaire. *Phlebology.* 2005;20(1):14-27.
15. Palfreyman S, Michaels J, Brazier J. Development of a tool to examine the effect of venous ulcers on patients' quality of life. *Nurs Stand.* 2007;21:57-58.
16. Bland JM, Dumville JC, Ashby RL, et al. Validation of the VEINES-QOL quality of life instrument in venous leg ulcers: repeatability and validity study embedded in a randomised clinical trial. *BMC Cardiovasc Disord.* 2015;15:85.
17. Brown A, Kendall S, Flanagan M, Cottee M. Encouraging patients to self-care - the preliminary development and validation of the VeLUSSET®, a self-efficacy tool for venous leg ulcer patients, aged 60 years and over. *Int Wound J.* 2014;11(3):326-334.
18. Abetz L, Sutton M, Brady L, McNulty P, Gagnon DD. The Diabetic Foot Ulcer Scale (DFS): a quality of life instrument for use in clinical trials. *Practical Diabetes Int.* 2002;19:167-175.
19. Vileikyte L, Peyrot M, Bundy C, et al. The development and validation of a neuropathy- and foot ulcer-specific quality of life instrument. *Diabetes Care.* 2003;26:2549-2555.
20. Gorecki C, Nixon J, Lamping DL, Alavi Y, Brown JM. Patient-reported outcome measures for chronic wounds with particular reference to pressure ulcer research: a systematic review. *Int J Nurs Stud.* 2014;51:157-165.
21. Kisala PA, Tulskey DS, Choi SW, Kirshblum SC. Development and psychometric characteristics of the SCI-QOL pressure ulcers scale and short form. *J Spinal Cord Med.* 2015;38:303-314.
22. Blome C, Baade K, Debus ES, Price P, Augustin M. The "Wound-QoL": a short questionnaire measuring quality of life in patients with chronic wounds based on three established disease-specific instruments. *Wound Repair Regen.* 2014;22:504-514.
23. Petrillo J, Cano SJ, McLeod LD, Coon CD. Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value Health.* 2015;18:25-34.
24. Rasch G. *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen, Denmark: Danmarks pædagogiske Institut; 1960.
25. Klassen A, van Haren ELWG, 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:e032332.
26. Aaronson N, Alonso J, Burnam A, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. *Qual Life Res.* 2002;11:193-205.
27. US Department of Health and Human Services. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. <https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf>. Accessed May 2019.
28. 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 Health.* 2011;14:967-977.
29. 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 Health.* 2011;14:978-988.
30. 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;8:94-104.

31. Thorne S, Kirkham SR, MacDonald-Emes J. Interpretive description: a noncategorical qualitative alternative for developing nursing knowledge. *Res Nurs Health*. 1997;20:169-177.
32. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13:iii, ix-x, 1-iii, ix-x, 177.
33. Pope C, Ziebland S, Mays N. Qualitative research in health care. Analysing qualitative data. *Br Med J*. 2000;320:114-116.
34. Sandelowski M. Theoretical saturation. In: Given LM, ed. *The Sage Encyclopedia of Qualitative Methods*. Vol 1. Thousand Oaks, CA: Sage; 2008:875-876.
35. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res*. 2018;27:1159-1170.
36. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377-381.
37. 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;42:495-504. <https://doi.org/10.1007/s00238-019-01529-7>.
38. Shaw JW, Johnson JA, Coons SJ. US valuation of the EQ-5D health states: development and testing of the D1 valuation model. *Med Care*. 2005;43:203-220.
39. Wright BD, Masters GN. *Rating Scale Analysis*. Chicago, IL: MESA Press; 1982.
40. Andrich D. *Rasch Models for Measurement*. Sage University Papers Series Quantitative Applications in the social Sciences. Vol 07-068. Thousand Oaks, CA: Sage; 1988.
41. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16:297-334.
42. Nunnally JC. *Psychometric Theory*. 3rd ed. New York, NY: McGraw-Hill; 1994.
43. Christensen KB, Makransky G, Horton M. Critical values for Yen's Q3: identification of local dependence in the Rasch model using residual correlations. *Appl Psychol Measur*. 2017;41:178-194.
44. Marais I. Local dependence. In: Christensen KB, Kreiner S, Mesbah M, eds. *Rasch Models in Health*. London, UK: Wiley-ISTE Ltd; 2013:111-130.
45. Andrich D. An elaboration of Guttman scaling with Rasch models for measurement. *Sociol Methodol*. 1985;15:33-80.
46. Caskin CJ, Happell B. On exploratory factor analysis: a review of recent evidence, an assessment of current practice, and recommendations for future use. *Int J Nurs Stud*. 2014;51:511-521.
47. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restor Dent Endod*. 2013;38(1):52-54.
48. Prinsen CAD, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27(5):1147-1157.
49. 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.
50. Norman RE, Gibb M, Dyer A, et al. Improved wound management at lower cost: a sensible goal for Australia. *Int Wound J*. 2016.13(3):303-316.
51. 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.
52. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg*. 2009;124(2):345-353.
53. Klassen AF, Cano SJ, Scott A, Snell L, Pusic AL. Measuring patient-reported outcomes in facial aesthetic patients: development of the Face-Q. *Facial Plast Surg*. 2010;26(4):303-309.
54. Klassen AF, Cano SJ, Soldin M, et al. The BODY-Q: a patient-reported outcome instrument for weight loss and body contouring treatments. *Plast Reconstr Surg Global*. 2016;4(4):e694.

How to cite this article: Klassen AF, van Haren ELWG, 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;18:487-509. <https://doi.org/10.1111/iwj.13549>

APPENDIX A.

Data saturation table for conceptual framework developed in the phase 1 qualitative study.

Domain	Major	Minor	Interview number												Total		
			1-5	6-10	11-15	16-29	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60			
Wound	Characteristic	Location	4	4	3	5	2	4	4	5	3	3	1	4	4	5	44
Wound	Characteristic	Size	5	4	4	3	2	3	3	4	4	4	2	5	5	5	45
Wound	Characteristic	Exudate	3	5	5	4	4	5	5	3	2	2	4	1	3	3	42
Wound	Characteristic	Colour	5	4	4	5	4	4	4	5	1	1	2	1	1	4	40
Wound	Characteristic	Smell	2	4	4	2	2	5	5	5	2	2	3	3	2	3	37
Wound	Characteristic	Texture	4	3	5	2	3	3	3	3	1	1	0	1	1	2	28
Wound	Characteristic	Depth	2	3	3	2	3	1	1	1	2	2	3	2	1	3	26
Wound	Characteristic	Shape	5	3	2	4	0	3	1	1	1	1	0	0	0	19	
Wound	Characteristic	Qualitative	1	2	2	1	1	4	3	0	0	0	0	1	1	3	19
Wound	Characteristic	Moisture	1	1	0	1	1	4	3	0	0	1	3	1	1	1	17
Wound	Characteristic	Holes/funnels	0	2	3	0	1	3	2	0	0	0	0	3	2	0	16
Wound	Characteristic	Visible	2	2	0	2	2	0	0	0	1	1	2	1	0	0	12
Wound	Healing	Better	2	2	3	2	2	5	3	3	3	3	3	5	5	4	39
Wound	Healing	Worse	4	3	2	1	2	4	2	2	2	2	3	4	4	4	35
Wound	Healing	Duration	1	3	0	1	1	0	4	1	1	1	0	1	4	4	20
Wound	Healing	Not improving	0	2	0	1	0	0	2	1	1	1	0	5	2	2	15
Wound	Healing	Recurring	0	0	0	0	1	1	2	0	0	0	0	1	2	3	10
Physical	Function	Mobility	2	5	2	3	5	5	4	2	2	4	4	3	2	3	40
Physical	Function	Accommodation	4	2	4	3	2	4	4	2	2	2	1	5	2	5	38
Physical	Function	Activity daily living	2	3	4	4	4	5	2	2	2	2	3	2	3	2	36
Physical	Function	Recreation	0	2	2	1	2	2	2	2	3	3	3	3	2	3	25
Physical	Function	Sleep	1	1	3	1	4	1	4	1	2	2	1	3	2	0	23
Physical	Symptom	Pain	4	5	5	3	5	5	5	5	5	5	3	3	2	4	49
Physical	Symptom	Sensations	3	1	4	2	0	4	1	1	1	1	3	1	0	0	20
Physical	Symptom	Itchy	2	1	1	2	1	2	2	1	1	1	0	0	1	0	13
Physical	Symptom	Swelling	1	1	1	2	1	1	3	1	1	1	0	0	0	1	12
Psychological	Distress	Irritation	3	5	4	4	5	4	3	5	4	3	5	1	2	3	41
Psychological	Distress	Worried	0	2	2	3	4	4	3	3	3	3	1	3	2	2	29
Psychological	Distress	Feeling down	1	2	1	2	1	2	4	2	3	3	1	2	3	0	22
Psychological	Distress	Self-conscious	2	2	1	1	2	2	3	2	2	2	1	2	0	0	18

(Continues)

Domain	Major	Minor	Interview number												Total		
			1-5	6-10	11-15	16-29	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60			
Psychological	Distress	Bothered	0	2	0	2	1	0	0	1	0	1	0	4	1	3	14
Psychological	Distress	Angry	0	1	1	0	0	3	3	1	1	0	1	0	1	0	11
Psychological	Positive	Acceptance	1	1	2	1	2	3	2	5	3	2	2	2	2	1	25
Social	Function	Work impact	2	3	3	0	3	2	4	2	4	2	4	1	3	3	30
Social	Function	Recreation/leisure	3	3	1	3	2	1	1	2	4	1	1	1	3	0	24
Social	Isolation	Stay in	2	1	1	1	1	3	2	2	2	2	3	3	1	2	21
Social	Isolation	People notice	1	2	0	0	0	2	3	0	0	1	1	1	0	10	
Social	Relationships	Instrumental	3	4	4	3	5	4	4	4	4	3	2	2	3	4	43
Social	Relationships	Emotional support	4	4	3	5	5	4	4	4	3	1	3	3	3	1	40
Experience	Patient-provider interaction	Rapport	2	5	4	2	4	5	5	3	4	2	2	2	3	4	43
Experience	Patient-provider interaction	Credentials	4	3	0	0	2	3	2	3	2	4	3	4	3	0	26
Experience	Patient-provider interaction	Accountability	0	4	2	0	3	2	1	2	1	2	3	2	2	2	22
Experience	Patient-provider interaction	Continuity	2	2	2	0	3	1	0	2	2	0	2	2	0	16	
Experience	Treatment delivery	Appropriateness	3	3	1	2	2	3	2	1	3	2	3	3	1	3	27
Experience	Treatment delivery	Patient-centred	1	2	0	0	1	0	3	2	1	2	2	4	1	1	17
Experience	Treatment delivery	Expectations	1	0	2	1	2	4	1	2	2	2	0	0	2	1	18
Experience	Treatment delivery	Continuity	1	1	2	2	1	1	0	0	2	3	2	3	2	1	16
Experience	Information delivery	Understandability	2	4	2	2	1	4	2	4	3	2	2	2	4	3	33
Experience	Information delivery	Reciprocity	1	2	3	0	1	2	4	2	3	2	2	2	3	0	23
Experience	Information delivery	Modality	2	2	0	0	0	2	3	1	2	1	1	1	1	2	16
Experience	Establish/obtain care	Availability	1	3	3	2	3	3	4	0	2	2	2	2	2	2	27
Experience	Establish/obtain care	Professionalism	3	1	3	2	1	1	0	3	3	3	3	3	4	2	26
Experience	Establish/obtain care	Specialist referral	1	2	1	1	1	3	0	2	0	2	2	2	0	0	15
Experience	Establish/obtain care	Efficiency	2	1	3	1	1	0	1	2	3	1	1	1	2	1	18
Experience	Care coordination	Interdisciplinary	1	2	1	0	0	2	2	2	1	2	1	2	1	2	16
Treatment	Medication		4	5	3	3	5	5	5	4	3	5	4	5	4	4	50
Treatment	Dressing		3	4	4	3	4	3	4	4	5	4	3	4	3	4	45
Treatment	Debridement		3	4	2	5	3	3	1	2	0	2	2	4	4	31	
Treatment	Surgery		1	3	3	0	1	3	3	2	2	3	2	3	5	3	29
Treatment	Cleaning		0	1	2	1	1	3	4	0	1	3	4	3	4	4	24
Treatment	Suction device		0	1	1	1	2	0	1	3	1	4	4	2	4	2	20
Treatment	Compression stocking		2	1	0	1	2	2	1	1	0	4	4	1	3	18	
Treatment	Hyperbaric oxygen		0	1	0	0	0	0	0	0	0	3	4	2	4	10	

APPENDIX B.

Rasch measurement theory statistics indicators of fit and differential item function (DIF) results for each scale.

Scale	Items	Item fit statistics							DIF		
		Location	SE	Fit res	DF	χ^2	DF	P-value	Gender	Age	Language
Assessment	Bleeding	-0.59	0.05	1.97	785.24	7.10	8	.53	No	No	No
	Smell	-0.56	0.05	0.83	778.93	4.67	8	.79	No	No	No
	Burning	-0.40	0.05	1.18	778.02	6.72	8	.57	No	No	No
	Holes	-0.28	0.05	2.10	764.49	10.14	8	.26	Yes	No	No
	Swelling	-0.12	0.05	1.09	780.73	6.87	8	.55	No	No	No
	Edges	-0.07	0.05	-1.62	756.36	8.72	8	.37	No	No	No
	Colour	-0.03	0.05	-2.42	762.68	10.47	8	.23	No	No	No
	Drainage	0.10	0.05	-2.36	792.47	10.86	8	.21	No	No	No
	Pain	0.40	0.04	3.25	788.86	4.96	8	.76	Yes	No	No
	Deep	0.51	0.05	-1.27	788.86	9.88	8	.27	No	No	No
	Size	1.03	0.05	-2.76	793.37	11.56	8	.17	No	No	Yes
Drainage	Colour	-0.39	0.07	-2.42	472.15	15.75	7	.03	No	No	No
	Thick	-0.35	0.07	-1.96	479.94	12.13	7	.10	No	No	No
	Smell	-0.25	0.06	1.07	484.27	6.95	7	.43	No	No	No
	Noticing	-0.23	0.06	-0.96	484.27	7.87	7	.34	No	No	No
	Clothes	-0.17	0.06	-1.03	485.14	9.43	7	.22	No	No	No
	Enjoy life	0.00	0.06	0.34	482.54	4.59	7	.71	No	No	No
	Amount	0.65	0.06	-1.72	495.54	12.00	7	.10	No	No	No
	Dressing	0.75	0.06	2.74	485.14	6.15	7	.52	No	No	No
Smell	Relationships	-0.73	0.11	-1.16	178.75	2.56	2	.28	-	-	-
	Comments	-0.61	0.11	0.49	179.61	0.48	2	.79	-	-	-
	Social life	-0.51	0.11	-1.68	178.75	4.87	2	.09	-	-	-
	Dressing on	-0.07	0.11	0.78	178.75	0.85	2	.65	-	-	-
	Noticing	0.05	0.10	-1.40	178.75	6.33	2	.04	-	-	-
	Stopping	0.16	0.10	0.47	179.61	2.10	2	.35	-	-	-
	Unpleasant	0.72	0.11	-1.04	178.75	1.99	2	.37	-	-	-
	Dressing off	0.99	0.11	0.03	177.03	3.45	2	.18	-	-	-
Life impact	Relationships	-0.58	0.05	-0.03	696.29	4.88	9	.85	No	No	No
	Relax	-0.57	0.05	1.94	709.32	4.34	9	.89	No	No	No
	Emotional	-0.15	0.05	-0.81	702.37	8.57	9	.48	Yes	No	No
	Social life	-0.11	0.05	-3.17	700.63	14.63	9	.10	No	No	No
	Independence	0.07	0.05	-0.08	698.02	8.95	9	.44	No	No	No
	Move around	0.27	0.05	1.70	712.80	2.00	9	.99	No	No	No
	Activities enjoy	0.43	0.05	-0.06	706.72	4.81	9	.85	No	No	No
	Physically active	0.64	0.05	0.26	705.85	5.55	9	.78	No	No	No
Psychological	Hopeless	-0.69	0.06	-1.56	712.34	8.13	9	.52	No	No	Yes
	Desperate	-0.66	0.06	-1.94	706.08	6.44	9	.70	No	No	Yes
	Overwhelmed	-0.58	0.06	-1.14	710.55	5.09	9	.83	No	No	Yes
	Sorry for self	-0.47	0.06	1.66	716.80	4.47	9	.88	No	No	No

(Continues)

Scale	Items	Item fit statistics							DIF		
		Location	SE	Fit res	DF	χ^2	DF	P-value	Gender	Age	Language
	Depressed	-0.33	0.06	-2.62	718.59	11.48	9	.24	No	No	No
	Self-conscious	-0.31	0.06	2.83	713.23	8.39	9	.50	No	No	Yes
	Anxious	0.12	0.06	-1.33	715.91	4.19	9	.90	No	Yes	No
	Irritated	0.53	0.06	-0.09	721.27	2.14	9	.99	No	No	Yes
	Frustrated	1.17	0.06	-1.08	726.64	2.15	9	.99	No	No	Yes
	Worried	1.23	0.06	-1.34	718.59	4.47	9	.88	No	No	No
Sleep	Falling asleep	-0.71	0.09	1.25	310.72	8.92	4	.06	No	No	No
	Enough sleep	-0.16	0.09	-0.83	308.34	4.21	4	.38	No	No	No
	Staying asleep	0.10	0.09	-3.70	310.72	3.48	4	.48	No	No	Yes
	Position	0.22	0.09	0.17	309.93	3.82	4	.43	No	No	No
	Woken up	0.56	0.10	-2.86	312.30	9.27	4	.05	No	No	No
Social	Isolated	-1.05	0.08	0.43	248.82	3.81	4	.43	No	Yes	Yes
	Meet people	-0.11	0.07	0.95	244.09	2.46	4	.65	No	No	No
	Missed out	0.17	0.07	-1.71	247.24	13.21	4	.01	No	No	No
	Cut down	0.41	0.08	-0.96	246.46	10.15	4	.04	No	No	No
	Enjoy life	0.57	0.08	2.11	250.39	2.88	4	.58	No	No	No
Dressing	Put on	-0.47	0.11	-1.61	222.17	7.54	3	.06	No	No	No
	Looks	-0.31	0.11	-0.49	218.67	4.02	3	.26	No	No	No
	Absorb	-0.26	0.11	0.20	228.29	8.07	3	.04	No	No	No
	Smell	-0.16	0.10	0.61	215.17	0.56	3	.90	No	No	No
	Remove - easy	-0.02	0.10	-2.30	224.79	7.74	3	.05	No	No	No
	Change	0.06	0.10	-2.04	222.17	7.41	3	.06	No	No	No
	Comfortable	0.24	0.10	-0.28	228.29	2.44	3	.49	No	Yes	No
	Remove - felt	0.32	0.09	0.84	227.42	3.53	3	.32	No	No	No
	Active	0.61	0.09	2.68	223.04	8.12	3	.04	No	No	No
Suction device	Drainage	-1.62	0.18	0.40	105.66	5.21	2	.07	-	-	-
	Looks	-1.10	0.17	-0.59	103.08	2.15	2	.34	-	-	-
	Sleep	-0.21	0.15	0.77	106.52	0.62	2	.74	-	-	-
	Noise	-0.12	0.15	1.47	106.52	0.59	2	.75	-	-	-
	Carry	0.06	0.15	-0.05	100.50	3.13	2	.21	-	-	-
	Comfortable	0.24	0.15	1.49	107.37	1.31	2	.52	-	-	-
	Socialize	0.67	0.15	-1.18	99.64	0.50	2	.78	-	-	-
	Enjoy life	0.83	0.16	-1.83	97.93	3.20	2	.20	-	-	-
	Active	1.25	0.16	-1.38	98.78	1.95	2	.38	-	-	-
Medical team	Respect	-1.34	0.15	-1.53	209.82	8.14	3	.04	No	No	No
	Understand	-1.13	0.14	-0.85	210.70	6.68	3	.08	No	No	No
	Professional	-0.51	0.14	0.73	208.93	3.32	3	.34	No	No	No
	Questions	-0.17	0.13	-0.39	209.82	2.31	3	.51	No	No	No
	Knowledgeable	-0.09	0.13	-2.83	206.28	7.90	3	.05	No	No	No
	Right experience	0.07	0.13	-0.49	207.16	1.15	3	.76	No	No	No
	Worked together	0.66	0.11	0.56	208.05	2.21	3	.53	No	No	No
	High level care	0.67	0.11	-2.10	210.70	6.09	3	.11	No	No	No

Scale	Items	Item fit statistics							DIF		
		Location	SE	Fit res	DF	χ^2	DF	P-value	Gender	Age	Language
Home care	Decisions	0.78	0.11	1.79	209.82	3.36	3	.34	No	No	No
	Available	1.06	0.11	1.43	202.73	1.96	3	.58	No	No	No
	Respect	-1.73	0.15	0.77	198.39	8.82	3	.03	No	No	-
	Spent time	-0.31	0.13	1.77	198.39	1.50	3	.68	No	No	-
	Attentive	-0.23	0.13	-1.58	199.27	2.58	3	.46	No	No	-
	Careful	-0.11	0.13	-2.03	194.84	6.16	3	.10	No	No	-
	Decisions	0.01	0.12	0.96	193.96	4.48	3	.21	No	No	-
	Professional	0.09	0.12	0.34	198.39	4.75	3	.19	No	No	-
	Questions	0.09	0.13	-0.36	198.39	3.37	3	.34	No	No	-
	Right experience	0.66	0.13	-0.43	196.61	5.43	3	.14	No	No	-
	Knowledgeable	0.72	0.12	0.25	198.39	8.88	3	.03	No	No	-
Knew what to do	0.82	0.12	-1.39	198.39	6.35	3	.10	No	No	-	
Wound clinic staff	Respect	-0.96	0.16	-0.25	152.13	5.77	2	.06	No	No	No
	Comfortable	-0.92	0.16	-1.85	152.13	2.64	2	.27	No	No	No
	Caring	-0.56	0.16	-1.77	147.01	4.61	2	.10	No	No	No
	Professional	-0.34	0.16	-3.17	152.13	7.54	2	.02	No	No	No
	Thorough	0.19	0.15	-1.75	147.01	1.17	2	.56	No	No	No
	Attentive	0.50	0.14	-0.56	151.28	2.52	2	.28	No	No	No
	Questions	0.71	0.14	0.81	147.01	3.57	2	.17	No	No	No
	Available	1.36	0.14	1.88	145.30	10.12	2	.01	No	No	No
Information	Ask questions	-0.76	0.12	-2.84	201.33	8.35	3	.04	-	No	No
	Team members	-0.59	0.12	0.76	207.51	6.31	3	.10	-	No	No
	Easy to understand	-0.35	0.11	-1.83	203.98	5.75	3	.12	-	No	No
	Time to discuss	-0.27	0.11	-1.73	202.21	6.94	3	.07	-	No	No
	Written information	-0.06	0.11	0.08	201.33	1.80	3	.62	-	No	No
	Promote healing	0.18	0.11	2.56	204.86	2.88	3	.41	-	No	No
	Wound products	0.19	0.10	1.15	203.09	3.07	3	.38	-	No	No
	Timing of information	0.49	0.11	-0.98	196.91	3.95	3	.27	-	No	No
	Consistent	0.56	0.10	-0.25	201.33	0.71	3	.87	-	No	No
	Expectations	0.61	0.10	1.65	200.45	3.27	3	.35	-	No	No

APPENDIX C.

Person-item threshold distributions for the assessment, life impact, information, and suction device scales.

APPENDIX D.

Characteristics of the subgroups for hypotheses testing,
inpatient = 117, outpatients = 761.

	Inpatient		Outpatient	
	N	%	N	%
Country				
Canada	19	16.2	108	14.2
Denmark	10	8.6	289	38.0
The Netherlands	24	20.5	197	25.9
USA	64	54.7	167	21.9
Gender n, (%)				
Male	73	62.4	445	58.5
Female	42	35.9	313	41.1
Other	0	0	2	0.3
Missing	2	1.7	1	0.1
Age in years				
18-49	28	23.9	116	15.2
50-59	32	27.4	151	19.8
60-69	28	23.9	215	28.3
70-79	19	16.2	186	24.4
80-95	9	7.7	93	12.3
Missing	1	0.9	0	0
BMI				
Under/normal weight	37	31.6	225	29.6
Overweight	31	26.5	225	29.6
Obese	43	36.8	285	37.4
Missing	6	5.1	26	3.4
Number of wounds				
Healed	0	0	11	1.4
1	77	65.8	474	62.3
2	19	16.2	142	18.7
3+	21	18.0	134	17.6
Wound type or cause				
Diabetic foot ulcer	22	18.8	129	17.0
Pressure ulcer	27	23.1	103	13.5
Surgical wound	21	17.9	121	15.9
Venous ulcer	8	6.8	102	13.4
Trauma/injury	9	7.7	87	11.4
Arterial ulcer	6	5.1	25	3.3
Hidradenitis	5	4.3	18	2.4
Radiation necrosis	5	4.3	9	1.2
Infection	2	1.7	7	0.9
Pyoderma gangrenosum	0	0	6	0.8
Cancer	0	0	6	0.8

	Inpatient		Outpatient	
	N	%	N	%
Pilonidal abscess	2	1.7	4	0.5
Other/unknown	7	6.0	129	17.0
Multiple	2	1.7	8	1.0
Missing	1	0.9	7	0.9
Wound location				
Leg or knee	24	20.5	244	32.0
Foot, ankle	33	28.2	194	25.5
Toe(s)	5	4.3	65	8.5
Buttocks	17	14.5	45	5.9
Abdomen, genitals, chest, or back	13	11.1	46	6.1
Arm, shoulder, armpit, hand	5	4.3	13	1.7
Face or neck	0	0	6	0.8
Other	0	0	5	0.7
Multiple	18	15.4	102	13.4
Missing	2	1.7	41	5.4
Wound age n, (%)				
3-6 months	45	38.5	272	35.7
7-12 months	21	17.9	143	18.8
1-2 years	21	17.9	160	21.0
3-4 years	10	8.5	78	10.2
5-10 years	12	10.3	37	4.9
More than 10 years	5	4.3	31	4.1
Missing	3	2.6	40	5.3
Wound size (length × width), cm				
<0.1	12	10.3	194	25.5
1-2.4	12	10.3	134	17.6
2.5-4.9	12	10.3	86	11.3
5-9.9	11	9.4	82	10.8
10-24.9	24	20.5	111	14.6
≥25	32	27.3	105	13.8
Missing	14	11.9	49	6.4

