Received: 27 February 2024 | Revised: 19 March 2024 | Accepted: 28 March 2024

DOI: 10.1111/wrr.13179

ORIGINAL ARTICLE-CLINICAL SCIENCE

Wound Repair and Regeneration — WILEY

Further psychometric evaluation of the WOUND-Q: A responsiveness study

Lucas Gallo MD, MSc[1] | Charlene Rae PhD[2] | Sophocles Voineskos MD, MSc[3] | Nina Vestergaard Simonsen MD, PhD[4,5] | Andrea L. Pusic MD, MHS, FACS, FRCSC[6] | Lotte Poulsen MD, PhD[4,5] | Jens Ahm Sørensen MD, PhD[4,5] | Anne F. Klassen BA (Hons), DPhil[2] | Stefan J. Cano BSc (Hons), PhD, CPsychol, AFBPsS[7]

[1]Division of Plastic Surgery, McMaster University, Hamilton, Ontario, Canada

[2]Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada

[3]Division of Plastic, Reconstructive & Aesthetic Surgery, University of Toronto, Toronto, Ontario, Canada

[4]Department of Plastic Surgery, Odense University Hospital, University of Southern Denmark, Odense, Denmark

[5]Research Unit for Plastic Surgery, Odense University Hospital, Odense, Denmark

[6]Division of Plastic and Reconstructive Surgery, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

[7]Modus Outcomes (a Division of Thread), Cheltenham, UK

Correspondence
Anne F. Klassen, Department of Pediatrics, McMaster University, 3N-27, Health Sciences Center, 1280 Main Street W, Hamilton, ON L8S 4L8, Canada.
Email: aklass@mcmaster.ca

Abstract

The WOUND-Q is a modular patient-reported outcome measure (PROM) with 13 scales measuring constructs across 4 domains (i.e., wound characteristics, health related quality of life, experience of care and wound treatment). The psychometrics of the WOUND-Q were previously assessed and the 13 scales evidenced good validity and reliability. However, the responsiveness (i.e., ability to detect clinical change) of the WOUND-Q has yet to be assessed. The objective of this study was to evaluate responsiveness for 9 WOUND-Q scales that assess outcomes, in a sample of people 18 years of age or older with chronic wounds that were present for at least 3 months. This study conducted a 4 month follow-up of 421 participants who completed the WOUND-Q as part of a previous psychometric study. Participants completed an online survey answering questions about their current wound state (e.g., number, type, size, smell, drainage), anchor questions about change, as well as the WOUND-Q scales that they had completed in their initial assessment. Predefined hypotheses were tested with a 75% acceptance threshold indicating sufficient evidence of responsiveness. Minimally important differences (MIDs) were also calculated using both anchor-based and distribution-based methods. Of 390 invited participants, 320 provided responses, ranging in age from 19 to 84 years. Acceptance of hypotheses ranged from 60% to 100%, with only the Symptom scale not meeting the 75% threshold. The findings of this study provide evidence that the WOUND-Q can validly measure clinical change in patients with chronic wounds.

KEYWORDS
chronic wounds, minimally important difference, patient-reported outcome measure, psychometrics, responsiveness

Abbreviations: COSMIN, COnsensus-based Standards for the selection of health Measurement Instruments; ES, effect size; HRQL, health-related quality of life; MID, minimally important difference; PROM, patient-reported outcome measure; REDcap, research electronic data capture; SD, standard deviation; SRM, standardised response mean.

1 | INTRODUCTION

Approximately 2.21 out of 1000 people are affected by chronic wounds.[1] The aetiology of chronic wounds varies, and can include vascular ulcers, diabetic ulcers and pressure ulcers. Treatment for chronic wounds is complex, further complicated by a lack of evidence for the efficacy of many wound care products.[2] Patient-reported outcome measures (PROMs) are useful tools providing the patient perspective on outcomes that can contribute to both clinical care as well as evidence on treatment efficacy.[3–5] PROMs can be used in clinical care for chronic wounds to capture the outcomes that are otherwise important to patients but often not readily observable (e.g., pain, quality of life and burden of disease).[6] To be useful, PROMs must follow rigorous international guidelines, and evidence both reliability and validity to ensure these tools accurately measure the constructs that they intend to.[7–9]

The WOUND-Q is a PROM developed for people aged 18 years of age or older with any type of chronic wounds.[10,11] This instrument is comprised of 13 independent scales measuring 4 domains: wound characteristics, health-related quality of life (HRQL), experience of care, and wound treatment. The validity and reliability of the WOUND-Q was assessed in an international sample of 881 participants demonstrating that the PROM measures the constructs accurately within its intended target population. Further evidence of these properties for the WOUND-Q was evidenced in a study of 421 participants from 22 countries.[12] As part of this later study, 2 new scales measuring Function and Symptoms were shown to be reliable and valid in a subset of 233 participants with chronic lower extremity wounds.[13] New scales added to the WOUND-Q increase its comprehensiveness by addressing important gaps. Figure 1 shows the expanded WOUND-Q conceptual framework.

In a 2023 systematic review of PROMs that have been used to assess patients with chronic wounds, the WOUND-Q was rated as 'very good' for PROM design[14]. The Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) used in the aforementioned review provides guidelines for the assessment of PROMs. During development of the WOUND-Q, most psychometric properties rated by the COSMIN guidelines were assessed including content validity, internal consistency, and construct validation.[10] A subsequent study further assessed the construct validity and test–retest reliability of these scales, as well as reported on their associated smallest detectable change.[12]

To date, there have been no studies assessing the responsiveness (i.e., ability to detect clinical change) of the WOUND-Q. This measurement property can be assessed by testing a priori hypotheses about expected differences or relationships in the change score of a PROM.[9,15] Notably, it is important for a PROM to be able to detect clinical change for its application in clinical trials, and clinical practice where users are interested in monitoring improvement or deterioration of outcomes over time.

The objective of this study was to evaluate responsiveness of 9 WOUND-Q outcome scales ('Assessment', 'Drainage', 'Smell', 'Sleep', 'Life Impact, 'Social', 'Psychological', 'Function' and 'Symptoms') in a sample of people 18 years of age or older with chronic wounds that were present for at least 3 months. The WOUND-Q scales that assess experience of care and wound treatment were not included as part of this study. This study also reported on the magnitude of change and minimally important difference (MID) for each WOUND-Q scale to aid in their interpretation.

2 | METHODS

This study included 421 participants who completed the WOUND-Q as part of a psychometric study conducted in September 2022.[12] This sample was recruited using the online crowd working platform Prolific Academic (www.prolific.com). Data collection was done in REDCap (Research Electronic Data Capture). This study complies with the

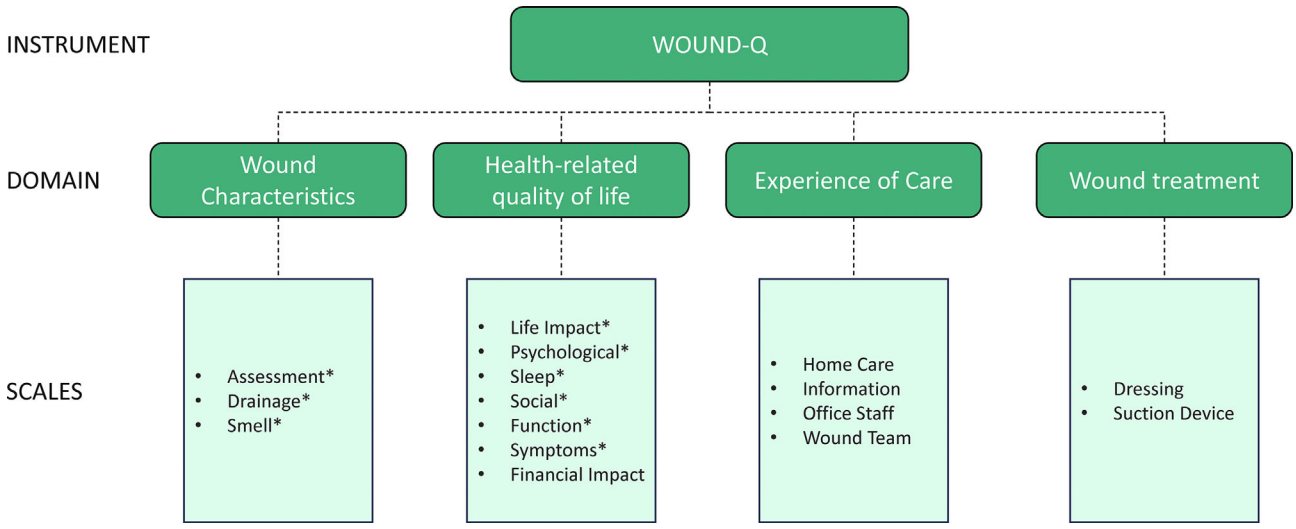

FIGURE 1 WOUND-Q conceptual framework.

1975 Declaration of Helsinki. Ethics approval was obtained from the Hamilton Integrated Research Ethics Board at McMaster University (#14946). Informed consent was obtained from all participants for this portion of the study.

2.1 | Participants

Eligibility criteria for the WOUND-Q psychometric study included: persons aged 18 years or older, with a chronic wound (present for at least 3 months) of any type, in any anatomic area, who were able to read, write and speak English.[12] Participants who completed the psychometric study were eligible to complete this follow-up study, approximately 4 months after the initial survey. A timeframe of 4 months was chosen as this was felt to be the approximate duration required to expect a small but important change in the wound construct(s) being measured based on the authors clinical experience managing various chronic wounds. Participants were compensated at a rate of £11.14/h through Prolific. A total of 390 of the 421 participants from the psychometric study were invited. Notably, 31 participants who responded "not sure" when asked for the type of chronic wound they had were not invited to take part in the responsiveness study. Data were collected using REDCap between December 2022 and January 2023.[16,17]

2.2 | Data collection

Data on demographics, wound characteristics, wound symptoms, and wound treatment were collected as part of the initial psychometric study. This study utilised both distribution-based and anchor-based methods for assessing MIDs in PROMs,[18,19] and adhered to the COSMIN guidelines for assessing responsiveness.[9,20] Data collected during the initial psychometric study were shown to participants at the time of the follow-up survey to remind them of the state of the construct at the time of initial survey.

Clinical variables collected from participants in the follow-up study included: current state of wound compared to 4 months ago (a lot worse, a little worse, about the same, a little better, a lot better, completely healed), wound size (length, width, and depth), amount of drainage (none, a little, moderate, a lot), amount of smell (no smell, faint, moderately strong smell, very strong smell), interference with sleep (never, sometimes (1–2 nights a week), often (3–4 nights a week), very often (5–7 nights a week)), number of old wounds, number of new wounds, type of wound(s) (e.g., Diabetic foot ulcer, Venous ulcer, Arterial ulcer, Pressure ulcer, wound caused by surgery, wound caused by radiation, wound caused by trauma or injury, Hidradenitis suppurativa, Pilonidal cyst/disease), wound location (face or neck, hand, arm, shoulder, chest, abdomen, back, buttocks, genitals, leg, foot, toe(s)), number of visits to a healthcare professional in past month for the chronic wound, number of overnight stays in hospital in past month because of chronic wound, and use of new treatments since the initial survey (yes, no, I am not sure). Before answering each WOUND-Q scale, participants were reminded of their answers to the following questions from their initial survey: wound size, amount of drainage, smell, sleep interference, pain, interference with emotional well-being, social life, and ability to participate fully in life.

2.3 | Outcome measures

2.3.1 | WOUND-Q

The WOUND-Q scales function independently, with respondents answering items on a 4-point scale. Scales are available in multiple languages. This study only used the English version of the scales. All scales scores were converted to a linearized Rasch transformed score from 0 (worst) to 100 (best). Missing data was imputed with the mean of the completed scale items, if less than 50% of the scale's item responses were missing. The recall period for all scales is the past week. 'Assessment' (concern about the wound), 'Drainage' (bothered by drainage), 'Smell' (bothered by the smell), 'Life Impact' (interfere with quality of life) scales all include the response options 'very much', 'quite a bit', 'a little bit', and 'not at all'. For this study, participants completed up to 9 scales from the WOUND-Q ('Assessment', 'Drainage', 'Smell', 'Sleep', 'Life Impact, 'Social', 'Psychological', 'Function' and 'Symptoms').[10] The Function and Symptoms scales were originally developed as part of the LIMB-Q, and these scales have been recently validated in a population of people who have chronic lower extremity wounds.[13] Only participants who reported a wound located on the toe, foot, ankle, or leg completed the Function and Symptoms scales. Also, the 'Drainage', 'Smell' and 'Sleep' scales were only provided to those who completed these scales as part of the initial baseline study.[13] Scales used branching logic so only those individuals who reported an issue with the construct at time of the initial survey completed the follow-up scale.

2.3.2 | EQ-5D-5L

The EQ-5D is a generic measure of HRQL that has been validated for use in populations with chronic wounds.[21–25] This measure is comprised of five questions that address mobility, self-care, usual activities, pain and discomfort, and anxiety and depression. Each dimension has five levels of severity and is rated based on today. Utility scores are calculated for each dimension as well as an overall global score of HRQL.

2.4 | Analysis

2.4.1 | Responsiveness validation

To assess responsiveness, we proposed pre-defined hypotheses (provided in Table 1). COSMIN guidelines consider acceptance of 75% or more of hypotheses by scale sufficient for validation purposes.[9] Differences between categorical variables were assessed using ANOVA

TABLE 1 Responsiveness hypotheses for WOUND-Q scale change scores.

	WOUND-Q scale change scores								
Hypothesis	Assessment	Drainage	Smell	Sleep	Life Impact	Social	Psychological	Function	Symptoms
The change in wound surface area will negatively correlate with change score for scale	Y**	Y**	Y**	NA	N	NA	NA	N	N
Improved wound state from 4 months ago will have positive change scores, change score magnitude will increase with self-reported improvement	Y**	Y**	Y**	Y**	Y**	Y**	Y**	Y**	Y**
EQ-5D overall change score will be positively correlated (0.3–0.5) with scale change score	NA	NA	NA	NA	Y**	Y**	Y**	Y**	Y**
EQ-5D mobility change score will correlate negatively (0.3–0.5) with scale change score	NA	NA	NA	NA	NA	NA	NA	Y**	NA
EQ-5D usual activities change score will correlate negatively (0.3–0.5) with scale change score	NA	NA	NA	NA	NA	NA	NA	Y**	NA
EQ-5D pain and discomfort change score will correlate negatively (0.3–0.5) with scale change score	NA	NA	NA	NA	NA	NA	NA	NA	N
EQ-5D anxiety and depression subscale score will correlate negatively (0.3–0.5) with scale change score	NA	NA	NA	NA	Y**	Y**	Y**	NA	NA
Participants who report improvement from 4 months ago will score higher than those that report no change	Y**	Y**	Y**	Y**	Y**	Y**	Y**	Y**	Y**
Proportion of hypotheses met	3/3	3/3	3/3	2/2	4/5	4/4	4/4	5/6	3/5

Note: Y, yes hypothesis met; N, no hypothesis rejected; NA, not applicable.
**$p < 0.001$.

or independent T-test. Correlations were assessed using Pearson correlation coefficients. Correlation hypotheses are in accordance with COSMIN criteria with correlations expected as follows: ≥0.5 between change scores in instruments measuring similar constructs; 0.3 and 0.5 between changes scores in instruments measuring related but dissimilar constructs; and <0.3 between change scores in instruments measuring unrelated constructs.[20] All EQ-5D dimensions were considered to be dissimilar with an expected correlation between 0.3 and 0.5. Statistical significance was considered $p \leq 0.05$. Descriptive statistics were calculated for each scale by magnitude of change. Categories with fewer than 30 responses were excluded.

2.4.2 | Responsiveness Indicators

Group level

Differences between 4-month and baseline Rasch transformed scores were assessed using paired-T tests for each scale. Two indicators of responsiveness were also calculated: Kazis effect size (ES)[26] and standardised response mean (SRM).[27] The ratio of ES/SRM was examined with larger values indicating greater responsiveness, and interpreted using Cohen's criteria as follows: 0.20, small; 0.50 moderate; >0.80 large.[28]

Individual level

Responsiveness at the individual level was calculated as the significance of each person's change in their own score.[29] This calculation was based on the following formula: (4 month score – baseline score)/SE_{diff} where $SE_{diff} = \sqrt{(SE^2_{baseline} + SE^2_{4\ month})}$. Significant change was then interpreted as: Significant worsening ≤1.96, Non-significant worsening −1.95 to 0, No change =0, Non-significant improvement 0 to +1.95, and Significant improvement ≥+1.96.[29]

2.4.3 | Minimally important difference

Two standard estimates of MID using distribution-based methods were calculated for each scale: 0.5 SD of the change score, and 0.5

SRM.[19] Finally, an anchor-based approach was taken to estimate the MID for each scale based upon the criteria suggested by Devji et al,[30] where a 2-stage anchor question was used. Anchor questions for each WOUND-Q scale were developed by the research team and these questions are provided in Supplement S1 for use by researchers in future research. First, participants were asked to answer about change in the scale construct comparing now to the baseline assessment (i.e., 4 months ago). Response options included: less, equally, or more. If participants answered "less" or "more" they were asked to assess the magnitude of that change using the

TABLE 2 Demographics and clinical characteristics.

***n* = 320**		***n***	**%**
Gender	Woman	160	50.0
	Man	157	49.1
	Other gender	3	0.9
Education	Some high school	6	1.9
	Completed high school	39	12.2
	Some college or trade school or university	61	19.1
	Completed college or trade school or university degree	141	44.1
	Some masters or doctoral degree	27	8.4
	Completed masters or doctoral degree	46	14.4
Race	Black	68	21.3
	East Asian	3	0.9
	Latin American	4	1.3
	Middle Eastern	4	1.3
	South Asian	5	1.6
	Southeast Asian	4	1.3
	White	215	67.2
	Other	5	1.6
	Multiple races	12	3.8
Financial stability	Not at all difficult	71	22.2
	A little difficult	103	32.2
	Somewhat difficult	80	25.0
	Very difficult	29	9.1
	Extremely difficult	34	10.6
	Prefer not to answer	3	0.9
Country	Canada	11	3.4
	Poland	14	4.4
	Portugal	13	4.1
	South Africa	64	20.0
	United Kingdom	111	34.7
	United States	63	19.7
	Other	42	13.0
	Prefer not to answer	2	0.6
Started new treatment in last 4 months	No	252	78.8
	Yes	61	19.1
	Not sure	3	0.9
	Missing	4	1.3
Wound state	A little/A lot worse	19	5.9
	About the same	51	15.9
	A little better	82	25.6
	A lot better	104	32.5
	Completely healed	64	20.0

following response options: small decrease/increase, but it is not important to me, small decrease/increase, and it is important to me, moderate decrease/increase and it is important to me, or large decrease/increase, and it is important to me. The response options of "small decrease/increase, and it is important to me" was considered a MID.

3 | RESULTS

Demographics and clinical statistics are shown in Table 2. Of the 390 participants invited, 320 provided responses (82% response rate). Participants ranged in age from 19 to 84 years with a mean age of 39 years (SD ± 14). Most participants reported that their wound was a little (26%) or a lot better (33%) than 4 months ago, with 6% reporting that their wound had gotten worse.

3.1 | Responsiveness validation

A summary of results for hypothesis testing is shown in Table 1, detailed results are provided in the supplementary file (Tables S2 to S3). The percent change in wound surface negatively correlated (i.e. as wound got smaller, scale change scores increased) with change scores ($p < 0.001$) on the Assessment ($r = -0.3$), Drainage ($r = -0.3$) and Smell ($r = -0.3$) scales. The association between percent change in surface area and change scores for the Life Impact scale was negatively correlated ($p = 0.02$); however, the Pearson correlation value was small (i.e., −0.2) and considered negligible. For the current state of the wound (i.e., a little/a lot worse, about the same, a little/a lot better, completely healed) there was a significant difference between categories ($p \leq 0.002$), with change scores for all scales increasing with self-reported improvement in wound state (Table S2). All correlation hypotheses between change scores for EQ-5D global score, its subscales and the WOUND-Q scales were met (Tables 1 and S3), apart from the relation between change scores for the EQ-5D pain and WOUND-Q Symptoms scale ($r = -0.2$, $p = 0.002$). Although the correlation was significant and in the expected direction, the magnitude of the correlation was less than expected. For the final hypotheses, the mean change scores were compared between the 'equally/same/no change', and improved group for each scale based on the anchor questions for each scale. The groups that indicated they got worse in the scale construct were not included due to small sample sizes in these group. For all scale change scores, the group that reported improvement in the construct had a greater change in scale score ($p < 0.001$) than those that reported no change in the construct (Table 3).

TABLE 3 Mean change scores for participants reporting improvement or no change to anchor questions for each scale.

WOUND-Q Scale change score	Anchor question	Response	N	Mean change[a]	SD	SE
Assessment	How concerned are you about this wound (e.g., size, drainage, smell, pain) now compared to 4 months ago?	Less	230	19	18	1
		Equally	75	4	13	1
Drainage	How bothered are you by the drainage (fluid produced by your wound) from this wound now compared to 4 months ago?	Less	121	32	20	2
		Equally	26	2	23	5
Smell	How bothered are you by the smell from this wound now compared to 4 months ago?	Less	92	31	24	3
		Equally	34	2	20	3
Sleep	How often does your wound(s) affect your sleep now compared to 4 months ago?	Less	165	27	25	2
		Same	57	2	20	3
Life impact	How much does your wound(s) interfere with your life (e.g., close relationships, work/volunteer, social life, doing activities you enjoy) now compared to 4 months ago?	Less	204	24	21	1
		Same	106	2	16	2
Social	With your wound(s) in mind, how is your social wellbeing (e.g., felt isolated, missed out on events) now compared to 4 months ago?	Same	148	4	24	2
		Better	150	27	28	2
Psychological	With your wound(s) in mind, how is your psychological wellbeing (e.g., depressed, self-conscious, anxious, frustrated) now compared to 4 months ago?	Same	139	6	17	1
		Better	152	18	21	2
Function	How difficult has it been to use (e.g., walk, moderate exercise) your lower limb (e.g., foot, ankle, knee, leg) now compared to 4 months ago?	Easier	92	23	19	2
		Same	68	5	21	3
Symptoms	How does your lower limb (e.g., foot, ankle, knee, leg) feel (e.g., pain, swollen, numb, stiff) now compared to 4 months ago?	Same	64	0	17	2
		Better	92	17	18	2

[a]All differences statistically significant $p < 0.001$.

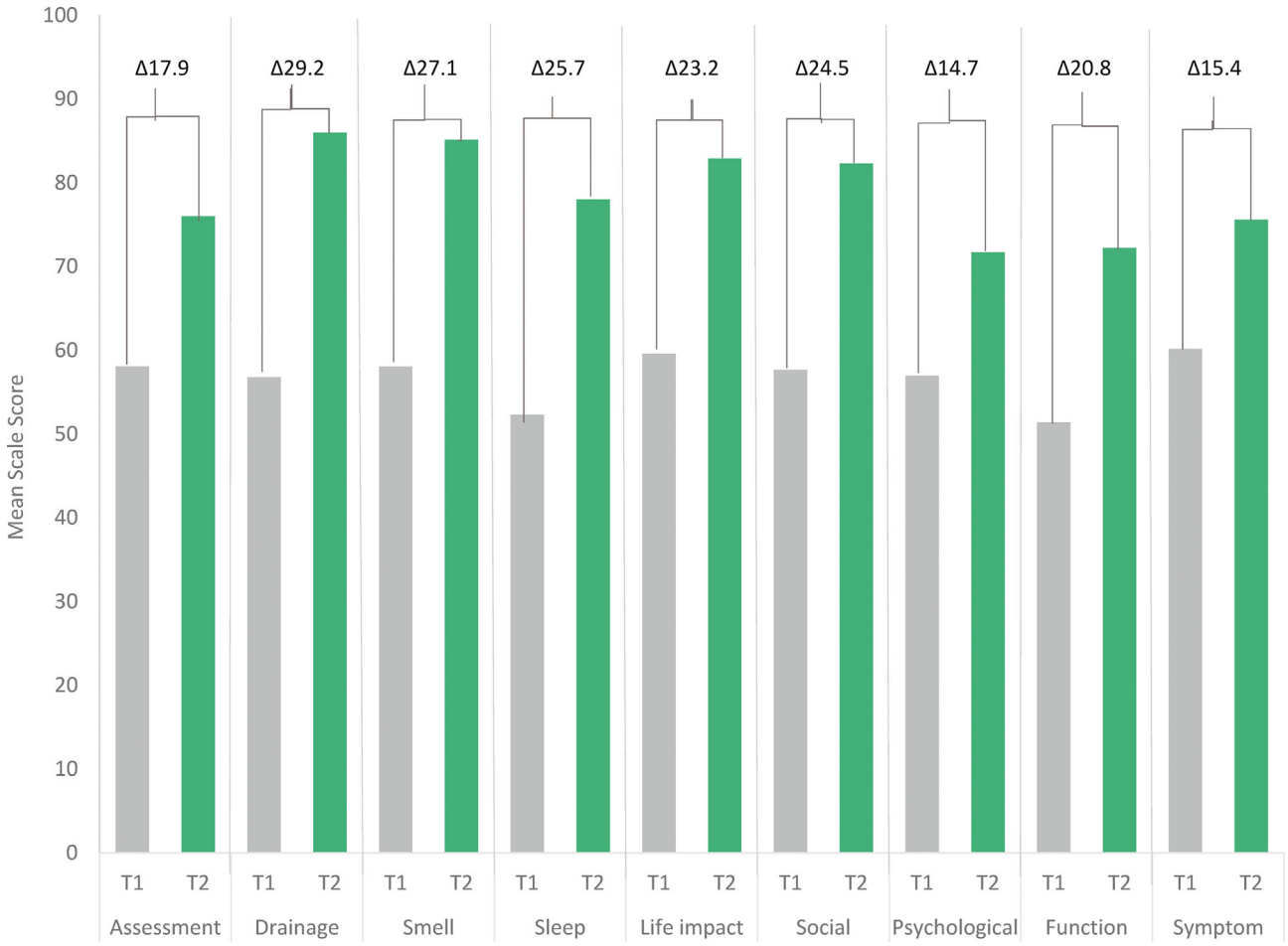

FIGURE 2 Mean score by scale at baseline (T1) and 4-month (T2).

3.2 | Responsiveness indicators

Mean score change over the 4-month period ranged from 15.4 to 28.9 ($p < 0.001$), with most of the sample scoring higher on the scales (Figure 2). Effect sizes associated with significant change scores for the scales were large ranging between 0.80 to 1.79 and 0.68 to 1.10 for ES and SRM, respectively (Table 4). The number of patients at the individual level showing significant change ranged between 79% and 93% (Table 5).

3.3 | Minimally important difference

Distribution based MIDs are shown in Table 3. Table 6 provides the mean change score by the magnitude of change reported by the participants. Those who reported 'a small change, but it wasn't important to them' were excluded from these analyses. Also, participants who reported that they got worse in the scale construct were excluded due to small sample sizes.

4 | DISCUSSION

The WOUND-Q is 1 of only 2 PROMs used in chronic wounds that has been shown to meet COSMIN standards for acceptable psychometric properties.[14] The findings of this current study further support the validity of the WOUND-Q, providing evidence on the responsiveness of this instrument. Indicators of responsiveness and MID values were also presented to aid in the interpretability of the WOUND-Q.

For this study, responsiveness was assessed using a construct rather than criterion approach, since a gold-standard instrument does not exist for all scales in the WOUND-Q. The construct approach involved testing hypotheses about difference between mean change scores between sub-groups, as well as conducting correlations between change scores for WOUND-Q scales and both the EQ-5D sub-scale and global utility score. Eight of the WOUND-Q scales met the COSMIN criteria of acceptance of 75% of the pre-defined hypotheses based on change scores, providing evidence that these scales can measure change over time.[20] Only the Symptoms scale, where 3 of the 5 hypotheses were accepted, did not meet this threshold. The 2 rejected hypotheses for this scale were correlation with the EQ-5D pain and discomfort item, and correlation with wound size change. It is likely the correlation with the EQ-5D pain and discomfort item is not as high as predicted as only 3 of 10 items in the scale focus on pain (i.e., at rest, when touched, weight bearing). In addition, the measurement of wound size in this study was self-reported and may not be accurate. Further assessment of the validity of change scores for the Symptoms scale is warranted.

TABLE 4 Group level indicators of change, and MID values for each scale.

Scale	Time point	Mean	SE	SD	N+	SE diff. for a person	ES (T2-T1)/SD1	Mean change[a]	SD change	ES/SRM	MID			SDC group level[11,12]
											SRM	0.5* SD_{change}	Anchor-based	
Assessment	T1	58.1	0.91	14.4	244	1.4	1.24	17.9	17.8	1.24	1.01	8.9	11	2.13
	T2	76.0	1.00	15.7	244									
Drainage	T1	56.8	1.40	16.3	132	2.2	1.79	29.2	21.4	1.31	1.36	10.7	23	2.23
	T2	86.0	1.68	19.3	132									
Smell	T1	58.1	2.04	20.3	99	3.1	1.33	27.1	27.1	1.33	1.00	13.5	3	3.20
	T2	85.1	2.35	23.4	99									
Sleep	T1	52.3	1.51	19.9	173	2.3	1.29	25.7	26.1	1.31	0.98	13.0	15	2.48
	T2	78.0	1.74	22.8	173									
Life Impact	T1	59.6	1.37	19.9	212	1.9	1.17	23.2	21.1	1.06	1.10	10.5	17	2.82
	T2	82.9	1.30	18.9	212									
Social	T1	57.7	2.12	27.3	166	2.9	0.90	24.5	28.9	1.06	0.85	14.4	17	2.68
	T2	82.3	1.98	25.6	166									
Psychological	T1	57.0	1.38	18.4	178	2.1	0.80	14.7	21.5	1.17	0.68	10.7	10	2.45
	T2	71.7	1.59	21.2	178									
Function	T1	51.4	2.26	22.4	98	3.2	0.93	20.8	21.7	0.97	0.96	10.8	14	2.15
	T2	72.2	2.21	21.8	98									
Symptom	T1	60.2	1.80	18.1	101	2.6	0.85	15.4	18.9	1.04	0.81	9.4	10	2.07
	T2	75.6	1.93	19.4	101									

[a]All differences significant based on paired-*T* test with $p < 0.001$; + excluding those that reported no change.

Abbreviations: ES, effect size; MID, minimally important difference; SDC, smallest detectable change; SRM, standard response mean.

TABLE 5 Individual level change for each scale.

		Significant worsening		Non-significant worsening		No change		Non-significant improvement		Significant improvement		Total significant change	
Scale	*N*	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Assessment	244	20	8	3	1	11	5	5	2	205	84	225	92
Drainage	132	5	4	4	3	10	8	2	2	111	84	116	88
Smell	99	8	8	3	3	9	9	4	4	75	76	83	84
Sleep	173	19	11	0	0	13	8	0	0	141	82	160	93
Life impact	212	17	8	2	1	18	9	6	3	169	80	186	88
Social	166	18	11	2	1	28	17	5	3	113	68	131	79
Psychological	178	25	14	13	7	15	8	10	6	115	65	140	79
Function	98	4	4	7	7	2	2	9	9	76	78	80	82
Symptom	101	12	12	6	6	6	6	9	9	68	67	80	79

TABLE 6 Mean change scores for each scale by magnitude of change.

Scale	Magnitude of important change	*N*	Mean change score	SD	SE
Assessment	Small	54	11	17	2
	Moderate	44	14	14	2
	Large	101	24	16	2
Drainage	Small	19	23	20	5
	Moderate	18	21	16	4
	Large	73	36	19	2
Smell	Small	12	3	17	5
	Moderate	11	19	18	5
	Large	54	39	23	3
Sleep	Small	33	15	21	4
	Moderate	30	22	25	5
	Large	70	36	25	3
Life Impact	Small	33	17	20	4
	Moderate	35	20	19	3
	Large	107	29	22	2
Social	Small	28	17	27	5
	Moderate	27	30	32	6
	Large	83	30	27	3
Psychological	Small	35	10	17	3
	Moderate	40	12	20	3
	Large	69	24	22	3
Function	Small	20	14	12	3
	Moderate	25	20	16	3
	Large	40	30	18	3
Symptoms	Small	16	10	12	3
	Moderate	24	13	16	3
	Large	49	22	19	3

In this article, we present multiple MID values based on both distribution and anchor-based approaches, with MID values varying by method. Applying multiple approaches is supported by Revicki et al[18] who noted that MID values may vary by both population and application. There also has been some arguments to incorporate more of a patient perspective in MIDs.[31] The values in this analysis should also be interpreted with caution as a stronger estimate of MID would be provided when derived from multiple studies that report MIDs.[18] Table 7 provides an overview of the MID and responsive indicator methods used in this article and their limitations.

TABLE 7 Tests and interpretation.

Type of test		Meaning	Interpretation	Limitations and Findings
Validity of change (*Responsiveness*)				
Hypothesis testing		Demonstrates the validity of change	Acceptance of 75% or more of hypotheses suggests scale measures change[9]	Dependent on quality and number of hypotheses tested. All the WOUND-Q scales tested had >75% acceptance of pre-defined hypotheses, except for the Symptom scale where 3 of 5 hypotheses were accepted.
Magnitude of change				
Indicators		Provides estimates of statistical significance and magnitude of change	Dependent on method	Does not provide information on whether the change observed is valid and meaningful.[14] Methods assume error is uniform across scale, which is not true for Rasch developed scales. All the indicators examined in this study showed group level change over the 4 month period.
Indicator calculation methods	Paired T	Group-level t = (Sum differences)/$\sqrt{((n \times (\text{Sum differences})^2) - (\text{Sum differences})^2/(n-1))}$	If $p > 0.05$ mean difference is significantly different than zero, meaning change has occurred between two time points	All scales had mean change scores greater than 0 ($p > 0.001$). Indicating change at the group level over the 4 month period with mean change >14
	Effect size (ES)	Group-level $(\text{Mean}_{T2} - \text{Mean}_{T1})/\text{SD}_{T1}$	The larger the effect size the greater the amount of change can be interpreted as follows:[27] >0.8 large 0.5–0.8 moderate 0.2–<0.5 low	Effects sizes were greater than 0.8 indicating a large change, except for the Psychological scale that had an ES of 0.8 indicating a moderate change.
	Ratio	Group-level ES/SRM	Interpreted using criteria above[27]	All scales had mean change scores greater than 0 ($p > 0.001$). Indicating change at the group level over the 4 month period with mean change >14
	Significant change[28]	Individual-level $(\text{Score}_{T2} - \text{Score}_{T1})/\text{SE}_{\text{diff}}$ where $\text{SE}_{\text{diff}} = \sqrt{(\text{SE}^2_{T1} + \text{SE}^2_{T2})}$	Significant improvement where significant change ≥1.96 Nonsignificant improvement where significant change between 0 and 1.95 No change where significant change = 0 Nonsignificant worsening where significant change between −1.95 and 0 Significant worsening where significant change ≤−1.96	79% or more of participants who reported change were found to have a significant change statistically. Both the assessment and sleep scale had the greatest percentage (>90%) of individuals with significant change, while the psychological, social, and symptom scales had the lowest (79%)

TABLE 7 (Continued)

Type of test		Meaning	Interpretation	Limitations and Findings
Interpretation of change				
Minimally important difference (MID)		Provides a threshold representing the smallest difference that indicates a change in the construct measured. Multiple methods available to calculate based on either anchors or distributions.	To be useful the MID needs to be greater than SDC,[8,9] otherwise one can't distinguish change from measurement error. If a mean change score is greater than the MID than the change can be considered clinically important.	May vary depending on method, population and context of use.[17,18] Methods based on CTT assume that error is constant along the scale. In Rasch analysis precision is highest in the middle of the scale and lowest on the extremes[28] therefore value for MID is dependent on a person's location on the scale. Values for MID can also differ based on method applied. In this study MID values ranged from 0.68 to 23, with the SRM value providing the smallest value for MID.
MID calculation methods	1/2 standard deviation (SD) of the mean change	$½ \times \sqrt{(\sum(\text{change} - \text{mean change})^2/(n-1)}$	Smallest value that represents a change based on statistical distribution	Does not account for the patient's perspective on what difference is important. MID values ranged from 8.9 to 14.4
	SRM	Mean difference/SD_{change}	Smallest value that represents a change based on statistical distribution	MID values ranged from 0.7 to 1.4, and like other distribution-based methods this value does not incorporate the patient perspective
	Anchor-based	Uses an anchor question that is independent of the scale that can be interpreted to find the minimal difference that represents a change	Smallest value that represents a small but important change based on the perspective of the patient	Potentially need to survey a large number of individuals to obtain a precise estimate of MID, only those that respond that they had a small but important difference are included in calculation. This method relies on patient answering an external question that could be impacted by recall bias, or current health status.[18] MID values ranged from 3 to 23. The sample size was small in most cases for participants reporting small but important changes. Estimates based on this method lack precision in this study due to sample size and should be interpreted with caution.

An important consideration when applying traditional classical test theory methods to Rasch developed scales is that error is not uniform across a Rasch scale, with precision greatest at the centre and lowest at the ends of the scale. Therefore, MID values are dependent upon the person location on the scale. MIDs may vary depending on baseline score especially when examining change at the individual-level. However, the main issue with the MID anchor-based calculations presented in this article is precision. To be consider a precise estimate of a MID based on criteria from the credibility assessment tool, the MID must be representative of at least 150 patients or have a precision estimate of a minimum of 25%.[30] Our estimates did not meet either of these criteria, and therefore further work in a larger sample of patients who have smaller changes in the state of their chronic wounds is required for the anchor-based method. The values presented in this article should thus be only considered exploratory and used as a guide to help with interpretation of change scores in the WOUND-Q.

This study has 4 key limitations. First, the wounds reported by participants were smaller in comparison to those typically observed in a clinical setting. This smaller wound size likely made the follow-up time longer than ideal for detecting an MID. Second, wound size was self-reported, and the accuracy of the measurements is unknown. However, participants were reminded of their initial wound dimensions reported in the base survey, and asked to answer a question about whether their wound was bigger or smaller to help ensure the consistency of metric. Third, temporally follow-up time was not considered optimal to determine an MID however due to the complex nature of chronic wounds this timeframe was thought to be clinically appropriate.[30] The 4-month follow-up likely resulted in a smaller group of patients who reported 'a small but important change in their wound', limiting the ability of this study to calculate a precise anchor-based estimate of the MID for the WOUND-Q scales. Further work should examine earlier time points, as small but important differences may occur earlier from the patient perspective. Fourth, this study recruited from an online platform with participants self-reporting that they had a chronic wound therefore diagnosis could not be verified. We excluded anyone who was unsure of the type of wound they had, but future work could include examining responsiveness and MIDs in a clinical setting to verify results.

Ultimately, this work fills an important gap in the psychometrics properties examined so far for the WOUND-Q providing initial evidence of its responsiveness. This evidence provides support for the use of the WOUND-Q examining change in both longitudinal studies and clinical applications. More information about the WOUND-Q can be accessed at qportfolio.org/woundq.

FUNDING INFORMATION

A. Klassen received funds from the Department of Paediatrics, McMaster University, for data collection.

CONFLICT OF INTEREST STATEMENT

L. Mundy is a co-developer of the LIMB-Q and A. Klassen and A. Pusic are co-developers of the WOUND-Q and LIMB-Q and would receive a share of any licence revenue in for-profit studies. A. Klassen provides research consulting services to pharmaceutical companies through EVENTUM Research. S. Cano is the CSO of Modus Outcomes, a Division of Thread. Remaining authors have nothing to disclose.

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.

ORCID

Lucas Gallo https://orcid.org/0000-0001-5152-9471
Charlene Rae https://orcid.org/0000-0003-4309-4636
Nina Vestergaard Simonsen https://orcid.org/0000-0001-6006-9753
Lotte Poulsen https://orcid.org/0000-0003-4205-806X
Jens Ahm Sørensen https://orcid.org/0000-0003-4903-0094
Anne F. Klassen https://orcid.org/0000-0003-4720-0096
Stefan J. Cano https://orcid.org/0000-0001-9390-273X

REFERENCES

1. 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.
2. Frykberg RG. Challenges in the treatment of chronic wounds. *Adv Wound Care*. 2015;4:560-582.
3. Porter I, Gonçalves-Bradley D, Ricci-Cabello I, et al. Framework and guidance for implementing patient-reported outcomes in clinical practice: evidence, challenges and opportunities. *J Comp Eff*. 2016;5(5):507-519.
4. Deshpande PR, Rajan S, Sudeepthi BL, Nazir CA. Patient-reported outcomes: a new era in clinical research. *Perspect Clin Res*. 2011;2(4):137-144.
5. Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. *Patient Relat Outcome Meas*. 2018;1:353-367.
6. Augustin M, Blome C, Goepel L, et al. Patient-reported outcomes as diagnostic tools and clues in chronic wounds: considerations for practice. *Wound Med*. 2015;8:6-14.
7. US Food and Drug Administration. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims (Accessed February 8, 2024). https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf.
8. Terwee CB, Prinsen CA, 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.
9. Prinsen CA, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147-1157.
10. Klassen AF, van Haren EL, 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.
11. 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.
12. Simonsen NV, Klassen AF, Rae C, et al. Further psychometric validation and test–retest reproducibility of the WOUND-Q. *Int Wound J*. 2024;21(1):e14354.

13. Simonsen NV, Klassen AF, Rae C, et al. The WOUND-Q function and symptoms scales for chronic lower extremity wounds: a validation study. *Adv Wound Care*.
14. van Alphen TC, ter Brugge F, van Haren EL, Hoogbergen MM, Rakhorst H. 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*. 2023;11(1):e4723.
15. Mokkink LB, Terwee CB, Knol DL, de Vet HC. The new COSMIN guidelines regarding responsiveness. *BMC Med Res Methodol*. 2011; 11:152.
16. Harris PA. Research electronic data capture (REDCap)-planning, collecting and managing data for clinical and translational research. *BMC Bioinform*. 2012;13(12):1.
17. 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(2):377-381.
18. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2): 102-109.
19. Franceschini M, Boffa A, Pignotti E, Andriolo L, Zaffagnini S, Filardo G. The minimal clinically important difference changes greatly based on the different calculation methods. *Am J Sports Med*. 2023; 51(4):1067-1073.
20. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*. 2010;19:539-549.
21. Herdman M, Gudex C, Lloyd A, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20:1727-1736.
22. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol group. *Ann Med*. 2001;33(5):337-343.
23. Lglesias CP, Birks Y, Nelson EA, Scanlon E, Cullum NA. 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;14:1705-1718.
24. Yang Y, Brazier J, Longworth L. EQ-5D in skin conditions: an assessment of validity and responsiveness. *Eur J Health Econ*. 2015;16:927-939.
25. Palfreyman S, Mulhern B. The psychometric performance of generic preference-based measures for patients with pressure ulcers. *Health Qual Life Outcomes*. 2015;13:1-9.
26. Kazis L, Anderson J, Meenan R. Effect sizes for interpreting changes in health status. *Med Care*. 1989;27:178-189.
27. Liang M, Fossel A, Larson M. Comparisons of five health status instruments for orthopedic evaluation. *Med Care*. 1990;28:10-642.
28. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Academic Press; 2013.
29. Hobart J, Cano S, Thompson A. Effect sizes can be misleading: is it time to change the way we measure change? *J Neurol Neurosurg Psych*. 2010;81:4.
30. Devji T, Carrasco-Labra A, Qasim A, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ*. 2020;4:369.
31. Schünemann HJ, Guyatt GH. Commentary—goodbye M (C) ID! Hello MID, where do you come from? *Health Serv Res*. 2005;40(2):593-597.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gallo L, Rae C, Voineskos S, et al. Further psychometric evaluation of the WOUND-Q: A responsiveness study. *Wound Rep Reg*. 2024;32(4):451-463. doi:10.1111/wrr.13179