Validation of the Dutch Version of the BODY-Q Measuring Appearance, Health-Related Quality of Life, and Experience of Healthcare in Patients Undergoing Bariatric and Body Contouring Surgery

Claire E.E. de Vries, MD, PhD; Elena Tsangaris, PhD; Dennis J.S. Makarawung, MD; Aebele B. Mink van der Molen, MD, PhD; Ruben N. van Veen, MD, PhD; Maarten M. Hoogbergen, MD, PhD; Andrea L. Pusic, MD; Caroline B. Terwee, PhD; Stefan Cano, PhD; and Anne F. Klassen, PhD

Aesthetic Surgery Journal
2023, Vol 43(5) 569–579

https://doi.org/10.1093/asj/sjac311
www.aestheticsurgeryjournal.com

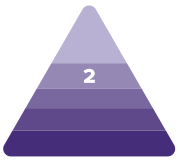
Abstract

Background: The BODY-Q is a patient-reported outcome measure developed for use in bariatric and body contouring surgery.
Objectives: The objective of this study was to examine the validity and reliability of the Dutch version of the BODY-Q.
Methods: The BODY-Q consists of 163 items in 21 independently functioning scales that measure appearance, health-related quality of life, and experience of care. The data used to validate the Dutch BODY-Q were provided by 2 prospective multicenter cohort studies across 3 hospitals in the Netherlands. The BODY-Q was administered before and after surgery at 3 or 4 months and 12 months. Rasch measurement theory (RMT) analysis was used to evaluate the BODY-Q for targeting, category threshold order, Rasch model fit, Person Separation Index, and differential item functioning by language (original English data vs Dutch data).
Results: Data were collected between January 2016 and May 2019. The study included 876 participants, who provided 1614 assessments. Validity was supported by 3 RMT findings: most scales showed good targeting, 160 out of 163 items (98.2%) evidenced ordered thresholds, and 142 out of 163 items (87.1%) fitted the RMT model. Reliability was high with Person Separation Index values >0.70 for 19 out of 21 scales. There was negligible influence of differential item functioning by language on person item locations and the scale scoring.
Conclusions: This study provides evidence for the reliability and validity of the Dutch BODY-Q for use in bariatric and body contouring patients in the Netherlands. The Dutch BODY-Q can be used in (inter)national research and clinical practice.

Dr de Vries is a plastic surgery resident and Dr Tsangaris is a researcher, Patient-Reported Outcomes, Value and Experience (PROVE) Center, Department of Surgery, Brigham and Women's Hospital, Boston, MA, USA. Dr Makarawung is a PhD student and Dr Mink van der Molen is a plastic surgeon, Department of Plastic and Reconstructive Surgery, University Medical Center Utrecht, Utrecht, the Netherlands. Dr van Veen is a surgeon, Department of Surgery, OLVG, Amsterdam, the Netherlands. Dr Hoogbergen is a plastic surgeon, Department of Plastic and Reconstructive Surgery, Catharina Ziekenhuis, Eindhoven, the Netherlands. Dr Pusic is a plastic surgeon, Division of Plastic Surgery, Brigham and Women's Hospital, Boston, MA, USA. Dr Terwee is a professor, Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit, Amsterdam, the Netherlands. Dr Cano is the chief scientific officer, Modus Outcomes, Letchworth Garden City, UK. Dr Klassen is a professor, Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada and is an interspecialty consulting editor for *Aesthetic Surgery Journal* and *ASJ Open Forum*.

Corresponding Author:
Dr Claire E.E. de Vries, PhD, Department of Surgery, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA.
E-mail: devries.cee@gmail.com

Level of Evidence: 2



Editorial Decision date: October 28, 2022; online publish-ahead-of-print December 7, 2022.

Severe and complex obesity is prevalent in most countries worldwide and bariatric surgery is considered to be the most effective treatment.[1,2] An adverse consequence of massive weight loss following bariatric surgery is the development of excess skin.[3] As a result of the increase in the number of bariatric surgery procedures performed annually and the resulting excess skin, postbariatric body contouring surgery is also on the rise.[4–7]

Bariatric and body contouring surgery can be evaluated using clinical endpoints (eg, weight loss) and patient-reported outcomes (PROs) data (eg, health-related quality of life [HRQL]). Quality of life (QOL), as defined by the World Health Organization (WHO), encompasses the emotional, social, and physical well-being of patients, and HRQL concerns those aspects of QOL that are related to health. PROs are best assessed by means of PRO measures (PROMs) that can provide insight into patients' experiences regarding the outcome of surgery and subsequent weight loss. Such information can be useful in clinical decision-making as well as in comparative effectiveness research. Identifying the most appropriate PROM for the target population and outcome(s) of interest is essential.[7] A key element in selecting a PROM is to ensure adequate measurement properties, ie, validity, reliability, and responsiveness.[8]

In 2018 members of our team participated in a systematic literature review to identify articles that described the development or validation of PROMs measuring HRQL in bariatric and body contouring surgery.[9] Of the 34 PROMs reviewed, the BODY-Q showed the strongest measurement properties, demonstrating good validity and reliability.[10,11] The BODY-Q is a PROM designed for patients seeking medical and surgical weight management treatment, and body contouring surgery.[11] The BODY-Q can be used as a primary outcome measure in clinical trials comparing different surgical or weight loss approaches and in prospective cohort studies following patients throughout their weight loss treatment journey.

Contrary to other PROMs in this field of surgery developed using classical test theory, the BODY-Q was developed using a modern psychometric approach, called Rasch measurement theory (RMT).[11] RMT offers interval-level measurement, which is more meaningful for clinical interpretation, measurement of person and item estimates that are not dependent on the distribution of the sample, robustness against missing data, and the ability to develop computer adaptive testing versions.[12] The BODY-Q was field-tested in the United States, Canada, and the United Kingdom.[11] It has been linguistically validated into Italian,[13] Finnish,[14] German,[15] French,[16] Swedish,[17] and Danish.[18]

To assess the validity of the BODY-Q in the Dutch population, we translated and culturally adapted the scales in accordance with recommended guidelines of the International Society for Pharmacoeconomics and Outcomes Research and the WHO.[19–21] The aim of this study was to further examine the validity and reliability of the Dutch version of the BODY-Q by replicating the RMT analysis of the original BODY-Q.

METHODS

Ethics

This study was conducted in accordance with the Handbook for Good Clinical Research Practice of the WHO and the Declaration of Helsinki principles, and was approved by the regional (Medical Research Ethics Committees United) and local IRBs of participating sites.

Translation and Cultural Adaptation Process

The BODY-Q scales were translated into Dutch in accordance with recommended guidelines of the International Society for Pharmacoeconomics and Outcomes Research and the WHO.[20,21] The translation involved 2 independent forward translations by professional translators whose mother tongue was Dutch and were fluent in English, and one backward translation by a professional translator whose mother tongue was English and was fluent in Dutch. The back-translation was compared with the original English version. The developers reviewed the discrepancies between the translation and the original version and gave iterative feedback until a final version was reached. An expert panel gave feedback on the scales' instructions, items, and response options. Cognitive debriefing interviews with patients undergoing bariatric or body contouring surgery were performed to identify words and/or phrases that were difficult to understand. Feedback from experts

and patients was used to revise and finalize the Dutch translation of the scales. The translation is described in detail elsewhere.[19]

Setting

This study was designed as a validation study of the Dutch BODY-Q. The data were provided by 2 prospective, multicenter cohort studies that were conducted in 3 hospitals in the Netherlands: OLVG West Hospital in Amsterdam, Catharina Hospital in Eindhoven, and St Antonius Hospital in Nieuwegein.

Study Population

Potential participants were recruited during their regularly scheduled preoperative clinic visit for bariatric or body contouring surgery. Patients undergoing bariatric surgery were screened for eligibility according to international criteria as follows: (1) having a BMI >40 or >35 kg/m^2 with obesity-associated illness; and (2) aged between 18 and 65 years. The eligibility criteria for body contouring surgery in the Netherlands are as follows: (1) had bariatric surgery >18 months previously, (2) had a stable weight for >12 months, (3) having a BMI <30 or <35 kg/m^2 (depending on the type of procedure), and (4) having excess skin classified as a "disfigurement" according to the Pittsburgh Rating Scale[22] or "demonstrable physical dysfunction." Written informed consent was obtained from all patients. Patients unable to read or understand Dutch, or who could not provide informed consent, were excluded.

Measures

The BODY-Q

The BODY-Q is a comprehensive PROM designed to measure outcomes most important to patients who are obese, and who undergo weight loss treatment and/or body contouring surgery.[11] The development of the BODY-Q involved a multiphase mixed-methods study including a literature review, qualitative and cognitive interviews with patients, and input from experts. Qualitative interviews were conducted to explore the impact that obesity, weight loss treatment, and body contouring surgery had on patients.[10] Sixty-three patient interviews were analysed to elicit concepts important to patients undergoing weight loss treatment and body contouring surgery.[10] This led to the development of a conceptual framework with 3 overarching domains (appearance, HRQL, experience of care) and 21 individually scored scales with a total of 163 items (see Table 1). Most BODY-Q items had a good fit to the Rasch model and ordered thresholds.[11] The internal consistency was adequate with Person Separation Index (PSI) values above 0.70 and Cronbach α values 0.90 or higher for all scales.[11] Test-retest reliability was ≥0.87 for 20 out of 21 scales, calculated with intraclass correlation coefficients.[11] Each BODY-Q scale consists of 4 response options and is scored by summing the responses to each item in a scale (raw score) and converting raw scores to a Rasch-transformed score ranging from 0 (worst) to 100 (best).

The structure of having individually scored scales allows for the addition of new scales as the field evolves. For example, 6 scales were later added that measure satisfaction with chest,[23] nipples,[24] stretchmarks,[25] expectations,[25] work life,[25] and eating-related concerns.[25] The scales measuring expectations, work life, and eating-related concerns were not included in the present study because the Dutch versions were not available at the time of study onset.[25]

Data Collection

Data were collected between January 2016 and May 2019. Participants were invited to complete the BODY-Q on a secure web-based application (Castor EDC).[26] A link to the survey was sent via email after their preoperative clinic visit. A separate link was developed for male participants that included the full BODY-Q along with the chest and nipple scales. Nonrespondents were sent up to 2 email reminders. Clinical and demographic characteristics (age, gender, BMI, ethnicity) were extracted from the patients' electronic medical records. The questionnaire was administered before surgery and at 3 or 4 and 12 months after surgery.

Analysis

The age, gender, BMI, ethnicity, and clinical group of the patients were given as mean [standard deviation, SD] or as percentages. Descriptive statistics were analyzed with SPSS 21.0 for Windows (SPSS Inc., Chicago, IL).[27]

RMT analysis was performed with RUMM2030 software (RUMM v. 2030, RUMM Laboratory Pty Ltd, Duncraig, Australia). To maintain consistency, the following graphical and statistical tests used to develop the original BODY-Q were applied to examine if the observed Dutch data fitted the RMT model and provided valid and reliable measurement:[11]

Targeting

Person-item threshold distribution was assessed to determine whether the items in a scale were appropriate for the target population by assessing their person locations relative to item locations. Ideally, scales should provide information at all levels of the construct to be measured so that the sample scores within the scale's range of measurement.[12]

Table 1. Overview of the Domains and Scales of the BODY-Q

Domain	Scale	Items	Example item	Response option format	Comment
Appearance	Body	10	How your body looks when you are dressed?	Dissatisfied/satisfied	
	Abdomen	7	How your clothes fit your abdomen?	Dissatisfied/satisfied	
	Arms	7	The size of your upper arms?	Dissatisfied/satisfied	
	Back	4	How smooth your back looks?	Dissatisfied/satisfied	
	Buttocks	5	The size of your buttocks?	Dissatisfied/satisfied	
	Hips and Outer Thighs	5	The size of your hips and outer thighs?	Dissatisfied/satisfied	
	Inner Thighs	4	How smooth your inner thighs look?	Dissatisfied/satisfied	
	Chest	10	How your chest (breast area) looks in a loose T-shirt?	Dissatisfied/satisfied	Only developed for men who experience weight loss and/or chest contouring
	Nipples	5	The shape of your nipples?	Dissatisfied/satisfied	Only developed for men who experience weight loss and/or chest contouring
	Stretchmarks	10	Not being able to wear certain clothes because of your stretch marks?	Not at all/extremely bothered	Only developed for patients with stretchmarks
	Skin	7	Your excess skin making you look bigger than you are (ie, overweight)?	Not at all/extremely bothered	Only developed for patients with excess skin
	Scars	10	Having to dress in a way to hide your scars?	Not at all/extremely bothered	Only developed for patients who underwent body contouring surgery
Health-Related Quality of Life	Body Image	7	I feel positive towards my body.	Agree/disagree	
	Physical	7	Getting up from a bed?	All the time/never	
	Psychological	10	I believe in myself.	Agree/disagree	
	Sexual	5	Sex is fulfilling for me.	Agree/disagree	Only developed for patients who are sexually active
	Social	10	I feel at ease at social gatherings with people I know.	Agree/disagree	
Experience of Healthcare	Doctor	10	Acted in a professional manner?	Agree/disagree	
	Office Staff	10	Treated you with respect?	Agree/disagree	
	Medical Team	10	Made sure to protect your privacy?	Agree/disagree	
	Information	10	The amount of written information they gave you to read?	Dissatisfied/satisfied	

Category Threshold Order

We assessed whether thresholds for response options were used in an orderly fashion for each scale. The thresholds values should progress in a logical order.[28]

Item Fit Statistics (Rasch Model Fit)

We examined the fit of the items to the Rasch model, ie, whether items in each scale worked together statistically.[29] For this analysis the sample size was amended to a random

sample (without intraperson dependencies) of 500.[30] Misfitting items were identified by examining χ^2 fit statistics to determine whether each individual item fit the Rasch model. Items with significant χ^2 probabilities after Bonferroni adjustment were considered as misfitting items.[12]

Reliability

Cronbach's α and PSI were used to evaluate scale reliability. PSI is a comparable and alternative measure to Cronbach's α that is specific to the RMT model and is calculated with and without extremes (ie, individuals who achieved a minimum or maximum score). Sufficient reliability was assumed for PSI and Cronbach's α values >0.70.[8]

Local Dependency

Local dependency indicates whether a person's responses to each item in the scale are dependent on each other. We examined local dependency by looking for correlations among the person-item residuals. We assumed a residual correlation of item pairs >0.30 was locally dependent.[31] To investigate whether dependency inflated scale reliability, the PSI values of 2 separate analyses were compared in a subtest analysis.

Differential Item Functioning

Differential item functioning (DIF) analyses were conducted to determine whether Dutch- and English-speaking patients, with equal levels of specific underlying characteristic, answer items of BODY-Q scales differently. To assess DIF by language, data from the original field-test study sample of English-speaking participants were included alongside the Dutch dataset and analyzed in RUMM2030.[11] Items that demonstrated substantial DIF were split to investigate whether there was a significant impact on person locations. To determine if DIF influenced the scoring, we used Pearson correlations to examine the extent to which the person locations correlated with the new (split) person locations.[12]

RESULTS

Translation and Cultural Adaptation Process

The Dutch version of the BODY-Q consisted of 163 items across 21 scales measuring appearance, HRQOL, and experience of care. Forward translation in Dutch revealed some items that were difficult to translate into Dutch. Difficulties were resolved after the consensus and reconciliation meeting for the forward translation. Comparison of the back-translation to the original English version identified that 11 out of 163 items (7%) changed in meaning, and 2 out of 163 items (1%) required retranslation. A total of 30 bariatric and body contouring patients took part in the cognitive debriefing interviews and reported difficulty understanding 9 out of 163 items (6%) from the Dutch version. The cognitive debriefing interviews resulted in minor changes in wording of 2 items and 2 instructions, but this did not change the meaning of these items. The translation and cultural adaptation process led to the development of an equivalent version of the BODY-Q.

Demographic Characteristics

A total of 876 patients completed the BODY-Q and provided 1614 assessments. The overall response rate was 77%, which varied by time of follow-up: preoperative, 84%; 3 to 4 months postoperative, 81%; and 12 months postoperative, 66%. Acceptability was high; 97% of participants completed all BODY-Q scales. The mean [SD] age was 45.7 [10.7] years (range, 18-72 years) and 89.1% of patients were female. Table 2 shows the sample characteristics of the Dutch and the English sample from the original BODY-Q development. The sample characteristics by gender, ethnicity, mean age, and mean BMI were similar across the 2 samples, whereas BMI class and clinical group varied.

Targeting

Figure 1 demonstrates examples of the person-item threshold distributions for the psychological and social scales, respectively. The quality of targeting for the scales is good, which is reflected by the mirrored distribution of the persons (top half of the graph) and item locations (lower half of the graph). The scales measuring experience of care demonstrated a high percentage (30%-60%) of patients who scored at the ceiling.

Category Threshold Order

A total of 160 out of 163 items evidenced ordered thresholds (98.2%). Three scales (Hips and Outer Thighs, Scars, and Office Team) had 1 item each with disordered thresholds (Table 3). The 3 items that showed disordered thresholds were "How your hips and outer thighs look from behind?" (Hips and Outer Thighs), "People seeing your scars?" (Body Contouring Scars) and "Treated you with respect?" (Office Team).

Item Fit Statistics (Rasch Model Fit)

For item fit statistics, 142 out of 163 items (87.1%) fitted to the RMT model, with significant χ^2 probability values after Bonferroni adjustment (Table 3). Supplemental Table 1, available online at www.aestheticsurgeryjournal.com, provides item fit statistics for each individual item.

Table 2. Sample Characteristics of the Dutch Patient Sample and the Field-Test (United States, Canada, and United Kingdom) Patient Sample

		Dutch sample (N = 876)	US, Canada, UK sample (N = 734)
Age (years)	Mean [SD]	45.7 [10.7]	46.5 [10.2]
	Range	18-72	18-75
Gender	Female	1405 (89.1)	644 (88.1)
	Male	172 (10.9)	87 (11.9)
BMI (kg/m^2)	Mean [SD]	34.33 [7.84]	33.1 [10.3]
	Range	19-60	17.8-75.8
BMI class	Normal weight	141 (9.7)	171 (24.0)
	Overweight	382 (26.3)	188 (26.4)
	Class I obesity	317 (21.9)	117 (16.4)
	Class II obesity	211 (14.6)	70 (9.8)
	Class III obesity	399 (27.5)	167 (23.4)
Ethnicity	White	806 (85.0)	609 (83.3)
	Other	142 (15.0)	122 (16.7)
Clinical group	Prebariatric	477 (29.6)	119 (12.3)
	Postbariatric	478 (29.6)	401 (41.6)
	Prebody contouring	399 (24.7)	93 (9.6)
	Postbody contouring	260 (16.1)	267 (27.6)
	Weight loss	0 (0.0)	85 (8.8)

SD, standard deviation.

Reliability

BODY-Q scales demonstrated good reliability, with PSI values for most scales that were ≥0.70 with and without extremes (exception: Office Staff [PSI value = 0.60 with extremes] and Medical Team [PSI value = 0.57 with extremes]). Cronbach's α for all scales was ≥0.89 (Table 3).

Local Dependency

Item residual correlations were above 0.30 (range r = 0.30-0.58) for 15 out of 607 (2.5%) item pairs. The subtest analysis demonstrated that the correlated items had a marginal influence on the reliability of the BODY-Q scales with a drop ≤0.050 in PSI values with/without extremes (Table 3).

Differential Item Functioning

DIF by language was evident for 35 out of 163 items (21.5%) in 18 out of 21 scales (Table 3). When the items were split by DIF, all Pearson correlations were ≥0.99, which means that DIF had negligible influence on person item locations and the scoring of the BODY-Q scales (Table 3).

DISCUSSION

RMT analysis provided evidence for the validity and reliability of the Dutch version of the BODY-Q. The distribution of person measurement and item locations showed that bariatric and body contouring patients scored within the range for which the scales provided measurement. This confirms that the BODY-Q scales were appropriately targeted in Dutch patients who underwent bariatric and/or body contouring surgery. Furthermore, most items showed ordered thresholds, and item fit, and scales evidenced high reliability and negligible impact of DIF by language. This study included a large and heterogeneous sample of patients that varied by stage of treatment to ensure that the BODY-Q can be used throughout a patient's treatment trajectory. Despite the length of the BODY-Q and the mode of

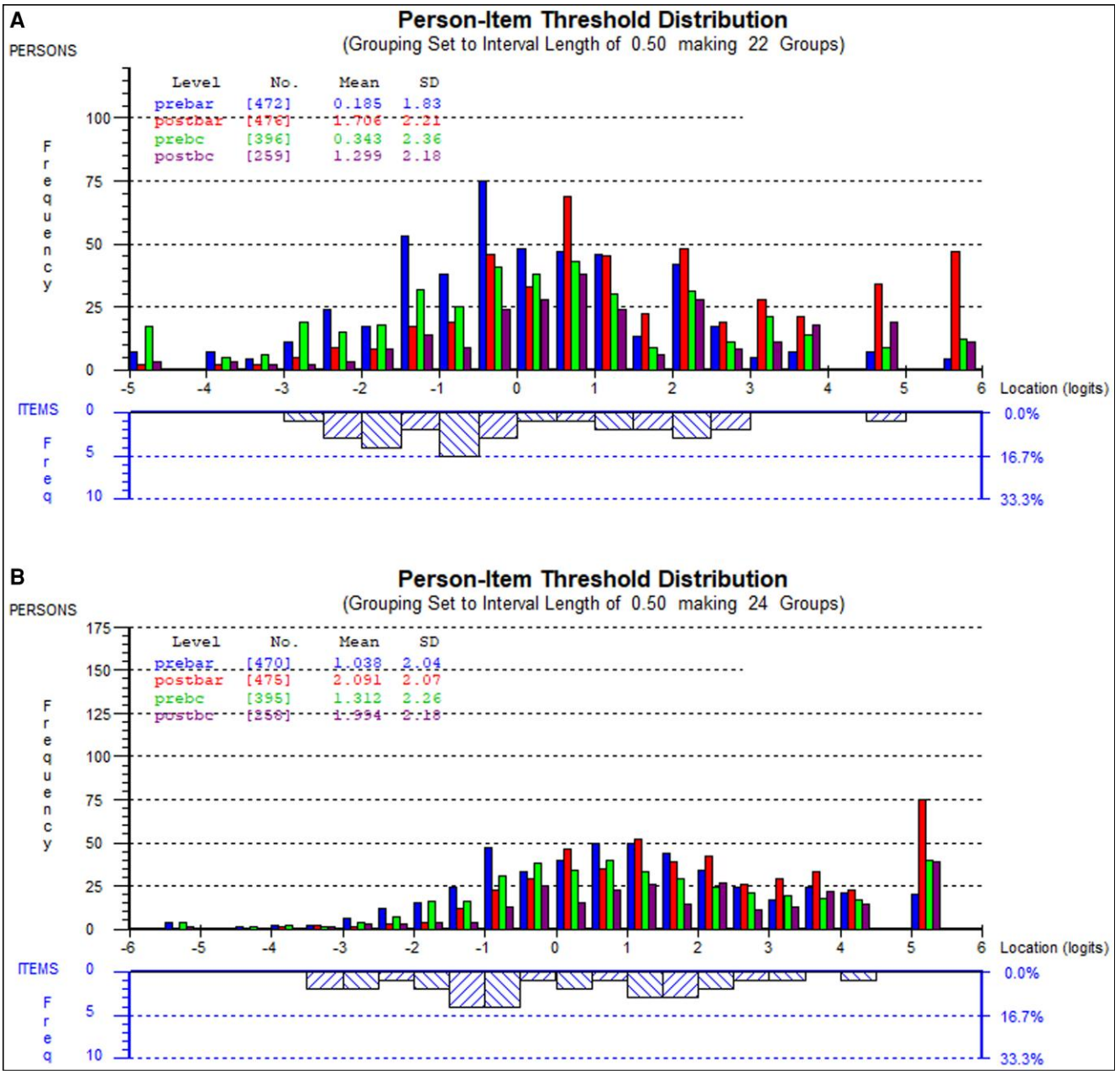


Figure 1. Plots showing the group of patients and their ability levels (upper) and the item locations and their distribution (lower). In the middle part of the plots, the item locations for each scale were evenly distributed to match the range of the construct reported by the patients, while there are patients above and below the range of measurement captured within the items of the scales (the floor and ceiling effects). (A) Person-item threshold distribution map for the psychological scale. (B) Person-item threshold distribution map for the social scale. Prebar, pre-bariatric surgery; Postbar, post-bariatric surgery; Prebc, pre-body contouring surgery; Postbc, post-body contouring surgery; SD, standard deviation.

administration for this study, ie, via email, the response rate was high, demonstrating that the Dutch BODY-Q was highly acceptable.

Concerns related to the fit to the Rasch model were evident because more items from the Dutch version of the BODY-Q showed misfit compared with the original English version.[11] Potential reasons for item misfit include: (1) the quality of the translation, as there were some items that were difficult to translate into Dutch (eg, “size” and “emotionally strong” because “size” can be interpreted as weight or subjective size, and “emotionally strong” is not commonly used in Dutch); (2) the large sample size (1604 assessments)—although larger sample sizes provide better calibration of the items, a sample that is too large

Table 3. Reliability Statistics and Other Indicators of Scale Performance

		Thresholds	Item fit statistics	Reliability			Local dependency	DIF
Scale	Items	Disordered thresholds (no. of items, (%))	Fit χ^2 (no of items, (%))	PSI with extremes	PSI without extremes	Cronbach's α	Residuals <0.30 (no. of correlated items (%), correlated items and correlation value, PSI values after subtest analysis)	DIF (no. of items (%), (Pearson correlation)
Body	10	0	0	0.93	0.93	0.96	2 (4) 1/2 = 0.33 9/10 = 0.37 PSI+ = 0.92 PSI− = 0.92	1 (10) (0.99)
Abdomen	7	0	0	0.91	0.92	0.97	0	1 (14) (0.99)
Arms	7	0	0	0.92	0.91	0.96	0	3 (43) (0.99)
Back	4	0	2 (20)	0.85	0.79	0.98	0	0
Buttocks	5	0	0	0.90	0.84	0.96	2 (20) 1/2 = 0.33 4/5 = 0.42 PSI+ = 0.87 PSI− = 0.81	1 (20) (1.00)
Hips & Outer Thighs	5	1 (20)	1 (20)	0.91	0.84	0.97	2 (20) 1/2 = 0.30 3/4 = 0.35 PSI+ = 0.91 PSI− = 0.85	1 (20) (0.99)
Inner Thighs	4	0	4 (100)	0.86	0.85	0.97	0	0
Chest	10	0	1 (10)	0.96	0.95	0.98	5 (11) 1/2 = 0.43 7/10 = 0.32 8/9 = 0.37 8/10 = 0.40 9/10 = 0.58 PSI+ = 0.95 PSI− = 0.93	0
Nipples	5	0	0	0.90	0.86	0.97	0	1 (20) (0.99)
Stretchmarks	10	0	2 (20)	0.93	0.93	0.98	0	1 (10) (1.00)
Skin	7	0	0	0.93	0.91	0.98	1 (5) 4/5 = 0.32 PSI+ = 0.93 PSI− = 0.90	1 (14) (0.99)
Scars	10	1 (10)	0	0.80	0.82	0.92	0	2 (20) (0.99)
Body Image	7	0	0	0.91	0.91	0.95	0	2 (29) (1.00)
Physical	7	0	0	0.86	0.86	0.94	0	4 (40) (0.99)
Psychological	10	0	1 (10)	0.91	0.91	0.94	1 (2) 1/2 = 0.31	3 (30) (1.00)

Table 3. Continued

		Thresholds	Item fit statistics	Reliability			Local dependency	DIF
Scale	Items	Disordered thresholds (no. of items, (%))	Fit χ^2 (no of items, (%))	PSI with extremes	PSI without extremes	Cronbach's α	Residuals <0.30 (no. of correlated items (%), correlated items and correlation value, PSI values after subtest analysis)	DIF (no. of items (%), (Pearson correlation)
							PSI+ = 0.91 PSI− = 0.90	
Sexual	5	0	0	0.81	0.76	0.89	0	1 (20) (0.99)
Social	10	0	0	0.89	0.89	0.94	2 (4) 3/4 = 0.42 9/10 = 0.41 PSI+ = 0.87 PSI− = 0.88	4 (40) (1.00)
Doctor	10	0	1 (10)	0.70	0.81	0.95	0	1 (10) (0.99)
Office Staff	10	1 (10)	2 (20)	0.69	0.85	0.96	0	1 (10) (0.99)
Medical Team	10	0	1 (10)	0.66	0.81	0.95	0	1 (10) (0.99)
Information	10	0	6 (60)	0.80	0.83	0.93	0	6 (60) (0.99)

DF, degrees of freedom; DIF, differential item functioning; PCA, principal component analysis; PSI, Person Separation Index; PSI+, PSI with extremes; PSI−, PSI without extremes.

may overpower the statistics, leading to more misfitting items;[30] (3) the participants from the Dutch sample differed from the English sample[11] by BMI class, clinical group, and time since surgery; and (4) the Dutch sample included more "extreme" patients. Most prebariatric patients scored lower on the appearance scales, while postbariatric patients evidenced higher scores on the HRQL scales. These results are expected as the goal of bariatric surgery is to improve HRQL. Postbariatric surgery patients were mainly in the "honeymoon period," ie, less than 1 year out from their surgery, whereas weight loss stabilizes after 1 year. The honeymoon period may explain why postbariatric surgery patients scored higher on the HRQL scales. Finally, patients at the extremes for most scales were not adequately represented, which may have further contributed to the individual item misfit. Despite having more items that did not fit the Rasch model, there was minimal influence on the measurement properties of the Dutch BODY-Q.

The experience of care scales showed good Cronbach's α values but lower PSI values, which might be related to the percentage of patients who scored at the ceiling (highest possible scores). Ceiling effects have implications on PSI statistics but not on Cronbach's α statistics. When calculating the Cronbach's α the assumption is that standard errors (SEs) are the same for all patients. Alternatively, for PSIs SEs are computed for each individual patient and are the largest for patients scoring at the ceiling. This results in lower PSI values (inflated by the high SEs) compared with Cronbach's α values.[32] The observed ceiling effects may be due to differences in culture or in how the Dutch sample experience their care. Finally, although our analysis showed DIF by language among 35 items, the impact was negligible and we can assume that scores on the Dutch BODY-Q are comparable to scores on the original English version.

The results of the RMT analysis in the present study were comparable to the results reported in a Danish study by Poulsen et al.[33] Although both studies demonstrated some differences (misfit to the Rasch model and lower PSI values for the experience of care scales), these results do not support the need for changes to the content of the BODY-Q.

Our study has some limitations. First, the overall response rate was 77%, which may have led to bias in the patient population. Second, the population did not include

medical weight loss patients, although the BODY-Q was originally developed to include this population. This limitation may have influenced the targeting of the scales and the generalizability of the Dutch BODY-Q to the medical weight loss population. Currently, the Dutch BODY-Q is implemented in a medical weight loss sample and further research is needed to validate the Dutch BODY-Q in this population. Third, while including pre- and postbariatric or body contouring patients in the RMT analysis has advantages, there is a risk for potential response shift that could influence patient's HRQL evaluation. Future studies should use DIF to examine if measured changes with the Dutch BODY-Q are "real" changes, resulting from a patient's change in health status over time, and not due to item difficulty that has changed over time (ie, response shift). Fourth, the scales assessing satisfaction with chest and nipples were developed for men who experience weight loss or who undergo chest contouring.[23] These scales were therefore administered to a smaller sample of patients. The small sample size (n = 159) may have influenced the results for these scales as the recommended sample size of 200 to 250 patients to assess DIF by language was not achieved.[12] Future validation studies should consider, when performing sample size calculations, that women comprise the majority of patients who undergo bariatric and body contouring surgery.

CONCLUSIONS

This study provides evidence for the use of the Dutch BODY-Q in bariatric and body contouring patients by demonstrating valid and reliable measurement. The Dutch BODY-Q has the potential for (inter)national use in research and clinical practice.

Supplemental Material

This article contains supplemental material located online at www.aestheticsurgeryjournal.com.

Disclosures

The BODY-Q represents intellectual property that is owned by McMaster University (Hamilton, Ontario, Canada) and Memorial Sloan Kettering Cancer Center (New York, NY). As co-developers of the BODY-Q, Dr Pusic, Dr Klassen, and Dr Cano receive royalties when this is used in a for-profit study. The other authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Funding

The development of the BODY-Q study was funded by the Canadian Institute of Health Research (CIHR) (Ottowa, Ontario, Canada) and National Endowment for Plastic Surgery (The Plastic Surgery Foundation, Arlington Heights, IL).

REFERENCES

1. Ng M, Fleming T, Robinson M, et al. Global, regional, and national prevalence of overweight and obesity in children and adults during 1980-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*. 2014; 384(9945):766-781. doi: 10.1016/S0140-6736(14)60460-8
2. Colquitt JL, Pickett K, Loveman E, Frampton GK. Surgery for weight loss in adults (Cochrane Review). *Cochrane Collab*. 2014;2014(8):CD003641. doi: 10.1002/14651858.CD003641.pub4
3. Klassen AF, Cano SJ, Scott A, Johnson J, Pusic AL. Satisfaction and quality-of-life issues in body contouring surgery patients: a qualitative study. *Obes Surg*. 2012; 22(10):1527-1534. doi: 10.1007/S11695-012-0640-1
4. Campos GM, Khoraki J, Browning MG, Pessoa BM, Mazzini GS, Wolfe L. Changes in utilization of bariatric surgery in the United States from 1993 to 2016. *Ann Surg*. 2020;271(2): 201-209. doi: 10.1097/SLA.0000000000003554
5. Kitzinger HB, Abayev S, Pittermann A, et al. The prevalence of body contouring surgery after gastric bypass surgery. *Obes Surg*. 2012;22(1):8-12. doi: 10.1007/s11695-011-0459-1
6. Reiffel AJ, Jimenez N, Burrell WA, et al. Body contouring after bariatric surgery: how much is really being done? *Ann Plast Surg*. 2013;70(3):220-260. doi: 10.1097/SAP.0B013E318236BA85
7. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "core outcome set"—a practical guideline. *Trials*. 2016;17(1):449. doi: 10.1186/s13063-016-1555-2
8. Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60(1):34-42. doi: 10.1016/j.jclinepi.2006.03.012
9. de Vries CEE, Kalff MC, Prinsen CAC, et al. Recommendations on the most suitable quality-of-life measurement instruments for bariatric and body contouring surgery: a systematic review. *Obes Rev*. 2018;19(10): 1395-1411. doi: 10.1111/obr.12710
10. Klassen AF, Cano SJ, Scott A, Tsangaris E, Pusic AL. Assessing outcomes in body contouring. *Clin Plast Surg*. 2014;41(4):645-654. doi: 10.1016/j.cps.2014.06.004
11. Klassen AF, Cano SJ, Alderman A, et al. The BODY-Q: a patient-reported outcome instrument for weight loss and body contouring treatments. *Plast Reconstr Surg Glob Open*. 2016;4(4):e679. doi: 10.1097/GOX.0000000000000665
12. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. *Health Technol Assess*. 2009;13(12):iii, ix-x, 1-177. doi: 10.3310/hta13120
13. Barone M, Cogliandro A, Salzillo R, Persichetti P. Translation and cultural adaptation of the BODY-Q into Italian. *Plast Reconstr Surg*. 2019;144(2):326e. doi: 10.1097/PRS.0000000000005821
14. Repo JP, Homsy P, Uimonen MM, Roine RP, Jahkola T, Popov P. Validation of the Finnish version of the BODY-Q patient-reported outcome instrument among patients who

underwent abdominoplasty. *J Plast Reconstr Aesthet Surg*. 2019;72(6):933-940. doi: 10.1016/j.bjps.2019.02.017
15. Hermann N, Klassen A, Luketina R, Vogt PM, Busch KH. [German linguistic validation of the BODY-Q: standardized PRO instrument after bariatric and bodycontouring surgery]. *Handchir Mikrochir Plast Chir*. 2019;51(4):255-261. doi: 10.1055/a-0824-7116
16. Schettini AV, Rillon P, Pirson G, de Coninck C. Completion of the French translation and linguistic validation of the BODY-Q. *Psychiatr Danub*. 2020; 32(Suppl 1):S150-S152
17. Fagevik Olsén M, Biörserud C, Nouh MA, Staalesen T, Elander A. Translation and validation of a Swedish version of the BODY-Q: a patient-reported outcome instrument for weight loss and body contouring surgery. *J Plast Surg Hand Surg*. 2022;56(6):348-352. doi: 10.1080/2000656X.2021.1956503
18. Poulsen L, Rose M, Klassen A, Roessler KK, Sorensen JA. Danish translation and linguistic validation of the BODY-Q: a description of the process. *Eur J Plast Surg*. 2017;40(1): 29-38. doi: 10.1007/s00238-016-1247-x
19. Nederlands Tijdschrift voor Plastische Chirurgie, 2017;(1): 1-24.
20. World Health Organization. Process of translation and adaptation of instruments. Geneva: World Health Organization; 2016. Accessed February 4, 2022. http://www.who.int/substance_abuse/research_tools/translation/en/
21. 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 Heal*. 2005;8(2):94-104. doi: 10.1111/j.1524-4733.2005.04054.x
22. Song AY, Jean RD, Hurwitz DJ, et al. A classification of contour deformities after bariatric weight loss: the Pittsburgh Rating Scale. *Plast Reconstr Surg*. 2005;116(5): 1535-1544. doi: 10.1097/01.prs.0000182606.92069.13
23. Klassen AF, Kaur M, Poulsen L, et al. Development of the BODY-Q chest module evaluating outcomes following chest contouring surgery. *Plast Reconstr Surg*. 2018;142(6):1600-1608. doi: 10.1097/PRS.0000000000004978
24. Poulsen L, Pusic A, Robson S, et al. The BODY-Q stretch marks scale: a development and validation study. *Aesthet Surg J*. 2018;38(9):990-997. doi: 10.1093/asj/sjy081
25. De Vries CEE, Mou D, Poulsen L, et al. Development and validation of new BODY-Q scales measuring expectations, eating behavior, distress, symptoms, and work life in 4004 adults from 4 countries. *Obes Surg*. 2021;31(8):3637-3645. doi: 10.1007/s11695-021-05462-2
26. Castor EDC. (2019). Castor Electronic Data Capture. [online] Available at: https://castoredc.com
27. IBM Corp. Released 2013. *IBM SPSS Statistics for Windows, Version 22.0*. Armonk, NY: IBM Corp.
28. Wright BD, Masters GN. *Rating Scale Analysis*. MESA Press; 1982.
29. Tennant A, Pallant JF. *Introduction to Rasch Analysis*. Psychometric Laboratory for Health Sciences, Department of Rehabilitation Medicine, The University of Leeds, 2007.
30. Hagell P, Westergren A. Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch unidimensional measurement model (RUMM) program in health outcome measurement. *J Appl Meas*. 2016;17(4):416-431.
31. Wilson M. *Constructing Measures*. Lawrence Erlbaum; 2005.
32. Wright BD. Separation, reliability and skewed distributions: statistically different levels of performance. *Rasch Meas Trans*. 2001;2001:786.
33. Poulsen L, Klassen A, Rose M, et al. Psychometric validation of the BODY-Q in Danish patients undergoing weight loss and body contouring surgery. *Plast Reconstr Surg Glob Open*. 2017;5(10):e1529. doi: 10.1097/GOX.0000000000001529