OPEN

PRS GLOBAL OPEN

ORIGINAL ARTICLE
Cosmetic

# BODY-Q Normative Scores: Psychometric Validation of the BODY-Q in the General Population in Europe and North America

Farima Dalaei, MD*†‡
Claire E.E. de Vries, MD, PhD§¶
Stefan J. Cano, PhD, CSO‖
Manraj N. Kaur, PT, PhD**
Lotte Poulsen, MD, PhD*†
André Pfob, MD††
Danny Mou, MD, MPH‡‡
Jussi P. Repo, MD, PhD§§
Rosa Salzillo, MD¶¶
Jakub Opyrchal, MD‖‖
Jens A. Sørensen, MD, PhD*†
Andrea L. Pusic, MD***
Anne F. Klassen, DPhil, Professor†††

**Background:** BODY-Q is a rigorously developed patient-reported outcome measure designed to measure outcomes of weight loss and body contouring patients. To allow interpretation and comparison of BODY-Q scores across studies, normative BODY-Q values were generated from the general population. The aim of this study was to examine the psychometric properties of BODY-Q in the normative population.
**Methods:** Data were collected using two crowdsourcing platforms (Prolific and Amazon Mechanical Turk) in 12 European and North American countries. Rasch measurement theory (RMT) was used to examine reliability and validity of BODY-Q scales.
**Results:** RMT analysis supported the psychometric properties of BODY-Q in the normative sample with ordered thresholds in all items and nonsignificant chi-square values for 167 of 176 items. Reliability was high with person separation index of greater than or equal to 0.70 in 20 of 22 scales and Cronbach alpha values of greater than or equal to 0.90 in 17 of 22 scales. Mean scale scores measuring appearance, health-related quality of life, and eating-related concerns scales varied as predicted across subgroups with higher scores reported by participants who were more satisfied with their weight. Analysis to explore differential item functioning by sample (normative versus field-test) flagged some potential issues, but subsequent comparison of adjusted and unadjusted person estimates provided evidence that the scoring algorithm worked equivalently for the normative sample as in the field-test samples.
**Conclusions:** The BODY-Q scales showed acceptable reliability and validity in the normative sample. The normative values can be used as reference in research and clinical practice in combination with local estimates for parallel analysis and comparison.
*(Plast Reconstr Surg Glob Open 2023; 11:e5401; doi: 10.1097/GOX.0000000000005401; Published online 16 November 2023.)*

## INTRODUCTION

Obesity is an increasing public health concern that has reached epidemic proportions. The World Health Organization has acknowledged obesity as the largest global chronic health problem in adults.[1] Given the rising obesity trends, there has been an exponential growth in the demand for obesity treatments, including diet and lifestyle changes, pharmacological therapy, and surgical procedures to manage the obesity-related comorbidities, reduce the risk of preterm death and improve quality of life and well-being.[2] However, irrespective of the choice of weight loss therapy, massive weight loss usually leads to

*From the *Research Unit of Plastic Surgery, Odense University Hospital, Denmark; †University of Southern Denmark, Odense, Denmark; ‡OPEN, Odense Explorative Patient Network, Odense, Denmark; §Department of Surgery, OLVG, Amsterdam, the Netherlands; ¶Department of Surgery, Brigham and Women's Hospital, Harvard University, Boston, Mass.; ‖Modus Outcomes Ltd, Cheltenham, United Kingdom; **Brigham and Women's Hospital, Harvard Medical School, Boston, Mass.; ††University Breast Unit, Department of Obstetrics and Gynecology, Heidelberg University Hospital, Heidelberg, Baden-Württemberg, Germany; ‡‡Department of General Surgery, Brigham and Women's Hospital, Harvard University, Boston, Mass.; §§Unit of Musculoskeletal Surgery, Department of Orthopedics and Traumatology, Tampere University Hospital and University of Tampere, Tampere, Finland; ¶¶Department of Plastic, Reconstructive and Aesthetic Surgery, Campus Bio-Medico University of Rome, Rome, Italy; ‖‖Department of Oncologic and Reconstructive Surgery, Maria Sklodowska-Curie Memorial National Cancer Center, Gliwice, Poland; ***Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, Mass.; and †††Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada.*

*Received for publication July 12, 2023; accepted September 20, 2023.*


DOI: 10.1097/GOX.0000000000005401

differing amounts of excess skin where subsequent body contouring surgery (BC) may be needed.[3,4] To determine the comparative effectiveness of different weight loss treatments and to understand the impact of weight loss and/or subsequent BC on patients' lives, rigorous patient-reported outcome measures are needed.[5]

In 2016, the BODY-Q, a patient-reported outcomes measure (PROM) for weight loss and BC, was developed and field-tested in the United States, Canada, and the United Kingdom.[6] The BODY-Q is composed of a set of independently functioning scales that measure four domains: appearance, health-related quality of life (HRQL), eating-related concerns, and experience of care (Fig. 1).[7] Adding to its modular design, additional scales (ie, cellulite, stretch marks, chest, expectations, work life) and the eating-related concerns domain (ie, eating-related distress, eating-related symptoms, and eating behavior) have been developed and validated.[8–12] The BODY-Q is increasingly used worldwide and has a growing number of translations.[13–16] Recent literature showed that the BODY-Q stood out as the PROM with the highest level of validation evidence for use in weight loss and BC patients.[17–19] Recently, our team published general population normative values from 12 European and North American countries to enable interpretation of the BODY-Q scores for clinical research and care.[20] Scores from the general population enables a better understanding of HRQL of obesity and change through the weight loss trajectory.[20] The current article supplements the normative values by providing the psychometric validation information for the normative BODY-Q sample. Psychometric validation is crucial to ensure that the scores obtained from the general population are reliable, valid, consistent over time and across different groups, and accurately reflect the construct being measured.[6,21,22] It is essential to ensure that the scores used are accurate and meaningful for research or clinical purposes.[6,8,23] The aims of this study were to examine the psychometric properties of the BODY-Q scales in the general population sample, including differential item functioning (DIF) that compares the normative study sample with the original field-test samples, and to determine support for a common scoring algorithm for international use.

## METHODS

### Normative Data

We previously published normative scores from the general population to interpret the BODY-Q from a sample of 4051 participants from 12 European and North American countries (Belgium, Canada, Denmark, England, Finland, France, Germany, Italy, the Netherlands, Poland, Sweden, and United States).[20] Participants (18 years or older) were recruited through the crowdsourcing platforms Prolific (www.prolific.co)[24] and Amazon Mechanical Turk (MTurk) (www.MTurk.com).[25] Participants were provided an information letter describing the study, and invited to complete the BODY-Q scales in their respective language through a URL link provided within Prolific and MTurk.

### Takeaways

**Question:** The aim of this study was to examine the psychometric properties of the BODY-Q in the normative population.

**Findings:** The Rasch measurement theory analysis supported the psychometric properties of the BODY-Q in the normative sample. The BODY-Q scales showed high reliability and validity in the normative sample.

**Meaning:** The normative values serve as reference points in both research and clinical practice, allowing for comparison of patients' BODY-Q scores with those of the general population, aiding in the interpretation of patients' scores.

The normative scores for 22 scales from three domains (appearance, HRQL, and eating-related concerns) are published elsewhere.[20]

### Analysis

Data were analyzed using SPSS Software (IBM Corp.; IBM SPSS Statistics for Mac, version 28.0, Armonk N.Y.). The psychometric properties of the international normative BODY-Q data were analyzed using the Rasch measurement theory (RMT) analysis approach using RUMM2030 software (RUMM version 2030, RUMM Laboratory Pty Ltd., Duncraig, Western Australia). In this approach, the analyses assess differences between observed and predicted responses to the items to determine if data from a sample fit the Rasch model.[12] Data that fit the Rasch model generate reliable and valid measurements. We repeated the set of RMT analyses that were performed in the original BODY-Q development and psychometric validation publication for the different scales to compare the findings of the normative population with the original sample.[6,8–12,23]

The RMT analyses involved the following statistical and graphical tests:

1. **Category threshold order**: for each scale, the thresholds between item response options (eg, definitely disagree to definitely agree) were examined. A scale's response categories should be scored with successive integers.
2. **Reliability:** two reliability coefficients were examined, ie, person separation index (PSI) and Cronbach alpha. PSI measures the error associated with the measurement of people in a sample. Cronbach alpha measures how closely a set of items in a scale are related. According to COSMIN criteria, acceptable PSI and Cronbach alpha values should be 0.70 or more.[26]

Disclosure statements are at the end of this article, following the correspondence information.

Related Digital Media are available in the full-text version of the article on www.PRSGlobalOpen.com.

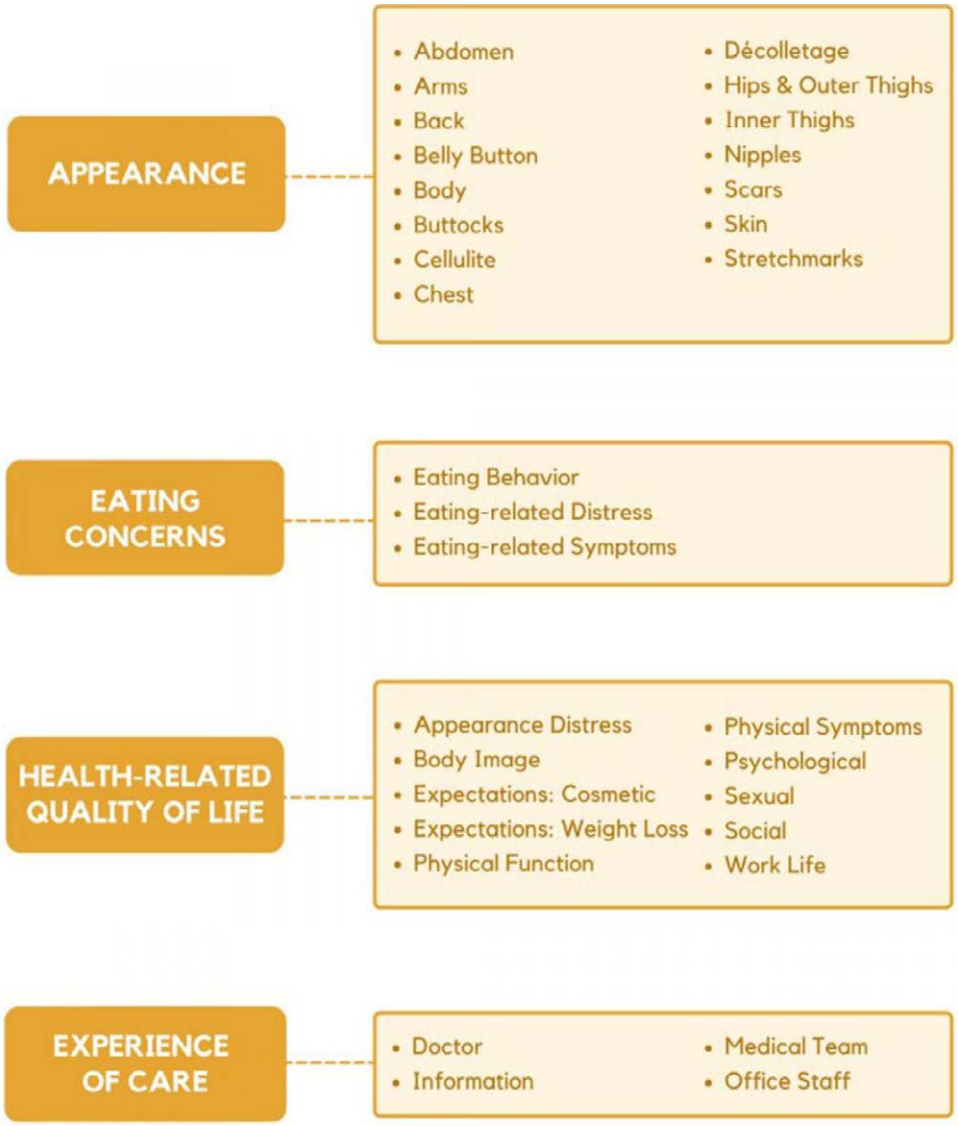


**Fig. 1.** BODY-Q Framework.[7]

3. **Item fit statistics:** three indicators of item fit to the Rasch model were examined to determine whether the items that form a scale worked together to map out a clinically important construct in the form of a hierarchy: (1) log residuals (item-person interaction),[2] chi-squared values (item-trait interaction), and (3) item characteristic curves. Fit residuals should be between -2.5 and +2.5, and chi-square values should be nonsignificant after Bonferroni adjustment.
4. **Targeting:** the person and item locations were examined to determine the extent to which the items that form a scale measured the construct that was experienced by the population. The percentage score on the scale was computed.
5. **Dependency:** the residual correlations between items were examined. Correlation between pairs of items should ideally be less than 0.30. High residual correlations can artificially inflate scale reliability. If values were 0.30 or more, a subtest analysis was performed to investigate the impact on the PSI.
6. **Stability:** DIF was tested to determine if the items that form a scale worked the same across subgroups within the sample. To examine DIF by sample, the BODY-Q dataset from the original field-test samples was used to compare with the normative sample. We also examined DIF within the normative sample for the following participant characteristics: age group (17–29, 30–39, 40–49, 50 or more years), gender (man and woman), continent (North America and Europe), and language (English and non-English). Chi-square values significant after Bonferroni adjustment were used to identify items with potential DIF. To determine if DIF influenced the scoring, Pearson correlations were used to examine the extent to which the unadjusted person locations (estimates) correlated with the new adjusted (split) person location.

7. **Correlation to original scoring:** To examine if the original BODY-Q scoring key could be used for the normative data, we correlated the logit scores for each scale's set of items for the normative sample and the original study sample. (**See figure, Supplemental Digital Content 1,** which displays field-test samples used for the Rasch measurement theory analysis. **http://links.lww.com/PRSGO/C860**.)

The Rasch logit scores were transformed to scores of 0–100 using the BODY-Q conversion table.[7] These scores were used to conduct the following tests of construct validity: First, correlations between the scales measuring similar, related but dissimilar, and unrelated constructs were investigated. According to the COSMIN guidelines for construct validity, correlations should be 0.50 or more for similar constructs, 0.30–0.50 for related but dissimilar constructs, and less than 0.30 for unrelated constructs.[27] Second, participants reported how satisfied they were with their current weight with the following response options: (1) extremely dissatisfied, (2) very dissatisfied, (3) somewhat dissatisfied, (4) somewhat satisfied, (5) very satisfied, and (6) extremely satisfied. Data were recorded info four groups for the analysis: (1) extremely and very dissatisfied, (2) somewhat dissatisfied, (3) somewhat satisfied, and (4) extremely and very satisfied. For each scale, the BODY-Q scores of the participants were reported as mean Rasch scores ± SD. We hypothesized that the BODY-Q appearance, HRQL, and eating-related concern scale scores would be incrementally lower with greater dissatisfaction with current weight. We also hypothesized that the BODY-Q scores for the appearance distress scale would be incrementally higher (more distress) for those who report greater dissatisfaction with their current weight.

## RESULTS

A total of 4051 participants from the general population completed the BODY-Q in their respective languages. The sample included 2052 North Americans and 1999 Europeans with a mean age of 36 (±14.7 SD) and mean body mass index of 26.4 kg per $m^2$ (±6.7 SD). Participant characteristics are outlined in Supplemental Digital Content 2, and Supplemental Digital Content 3 summarizes the mean scores for the normative participants as whole and by continent (ie, North Americans and Europeans).[20] (**See figure, Supplemental Digital Content 2**, which displays the participant characteristics.[20] **http://links.lww.com/PRSGO/C861**.) (**See figure, Supplemental Digital Content 3**, which displays the normative scores.[20] **http://links.lww.com/PRSGO/C862**.)

For the RMT analysis, all 176 items in the 22 scales had ordered thresholds, indicating that respondents could appropriately discriminate amongst response options. The item fit statistics provided evidence of validity for 167 of 176 items with nonsignificant chi-square *P* values after Bonferroni adjustment. Item fit was outside the criteria of -2.5 to +2.5 for 98 of 176 items; of these, only nine items had a significant chi-square *P* value. **(See figure, Supplemental Digital Content 4**, which displays item fit statistics and differential item functioning. **http://links.lww.com/PRSGO/C863**.)

At the scale level, the proportion of participants to score within the measurement (ie, targeting) ranged from 58.6% (stretch marks) to 99.4% (eating behavior). Reliability was high with PSI and Cronbach alpha with/without extremes of 0.70 or more for 20 of 22 scales, with the majority of values more than 0.80. Residuals in one or two item pairs in eight 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 0.05 (eating behavior). Data fit the Rasch model for 12 scales (nonsignificant *P* values). The remaining scales showed some misfit to the Rasch model (Table 1).

The DIF from the original study sample for the development and validation of each scale was compared with the normative study sample. See Supplemental Digital Content 1 (**http://links.lww.com/PRSGO/C860**) or the included original field-test samples. DIF was detected for 31 of 176 items in the sample analysis. In the normative sample, DIF was detected for three of 176 items for age group, eight of 176 items for gender, nine of 176 items for ethnicity, one of 176 items for continent, and seven of 176 items for language (**Supplemental Digital Content 3, http://links.lww.com/PRSGO/C862**). Pearson correlations between person locations for items before and after splitting the items for DIF showed a negligible impact on scoring (all correlations >0.995). The findings confirm that the original scoring key can be used in the normative sample.

For construct validity, both hypotheses were fully or partially supported. The majority of correlations between scales were in concordance with the COSMIN criteria for construct validity (ie, 0.50 or more for similar constructs, 0.30–0.50 for related but dissimilar constructs, and less than 0.30 for unrelated constructs; Table 2).[27] The BODY-Q scores were higher for participants who reported higher levels of satisfaction with their current weight in 21 of 22 scales; however, for the cellulite scale, participants scored higher in the "somewhat satisfied" group compared with "extremely to very satisfied." In the appearance distress scale, as hypothesized, participants scored incrementally higher (more distress) with greater dissatisfaction with their current weight (Fig. 2).

## DISCUSSION

The BODY-Q is a reliable and valid PROM for weight loss and/or BC patients that has been increasingly used to evaluate treatment outcomes worldwide.[17–19] Recently, we published the general population normative scores for interpreting the BODY-Q.[20] In this study, the psychometric properties of the BODY-Q were investigated in the normative population, to evaluate how data fit the Rasch model and to assess the evidence to support the use of a common scoring algorithm for international use.

Overall, this study provides broad support that the BODY-Q scales were acceptable, reliable, and valid for the international normative sample of the general population. The psychometric properties showed that the scales were appropriately targeted to the general population,

**Table 1. Reliability Statistics and Other Indicators of Scales**

| Scales | # Completed Scale | # Included in RMT | Scores on Scale % | $\chi^2$ Scale | DF | *P* | PSI +extr | PSI −extr | α +extr | α −extr | DT | ±2.5 | Residuals 0.3 | $\chi^2$ Item |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Body | 4050 | 3836 | 94.7 | 51.3 | 90 | 0.99 | 0.93<br>0.88 | 0.92<br>0.82 | 0.95 | 0.93 | 0 | 6 | 0 | 0 |
| Abdomen | 1606 | 1375 | 85.6 | 76.4 | 42 | 0.001 | 0.93 | 0.91 | 0.96 | 0.93 | 0 | 7 | 0 | 1 |
| Arms | 1607 | 1410 | 87.7 | 59.7 | 56 | 0.34 | 0.91 | 0.89 | 0.93 | 0.89 | 0 | 5 | 0 | 1 |
| Back | 1607 | 1341 | 83.4 | 28.8 | 28 | 0.42 | 0.78 | 0.77 | 0.91 | 0.82 | 0 | 2 | 0 | 0 |
| Buttocks | 1607 | 1356 | 83.4 | 52.1 | 30 | 0.007 | 0.86 | 0.79 | 0.92 | 0.84 | 0 | 4 | 0 | 0 |
| Hips, outer thighs | 1607 | 1326 | 82.5 | 20.0 | 20 | 0.456 | 0.88<br>0.88 | 0.82<br>0.82 | 0.94 | 0.86 | 0 | 3 | 0 | 0 |
| Inner thighs | 1607 | 1256 | 78.2 | 26.9 | 20 | 0.139 | 0.83 | 0.80 | 0.94 | 0.85 | 0 | 2 | 0 | 0 |
| Skin | 714 | 584 | 81.8 | 79.0 | 63 | 0.08 | 0.91<br>0.88 | 0.88<br>0.86 | 0.95 | 0.91 | 0 | 2 | 2/3 = 0.35<br>4/5 = 0.30 | 0 |
| Chest | 792 | 711 | 89.8 | 111.1 | 90 | 0.06 | 0.94<br>0.94 | 0.94<br>0.94 | 0.96 | 0.94 | 0 | 4 | 0 | 0 |
| Nipples | 795 | 568 | 71.4 | 109.9 | 20 | <0.001 | 0.83 | 0.77 | 0.93 | 0.84 | 0 | 5 | 0 | 2 |
| Stretch marks | 604 | 354 | 58.6 | 122.1 | 80 | 0.002 | 0.82<br>0.79 | 0.89<br>0.87 | 0.96 | 0.93 | 0 | 1 | 2/4 = 0.32 | 0 |
| Cellulite | 170 | 139 | 81.8 | 20.9 | 22 | 0.528 | 0.94<br>0.92 | 0.94<br>0.92 | 0.97 | 0.95 | 0 | 1 | 1/6 = 0.40<br>2/3 = 0.56<br>4/6 = 0.37<br>10/11 = 0.31 | 0 |
| Psychological | 3655 | 3275 | 89.6 | 79.1 | 90 | 0.79 | 0.93 | 0.93 | 0.96 | 0.94 | 0 | 7 | 0 | 0 |
| Physical | 4046 | 2677 | 66.2 | 137.9 | 42 | <0.001 | 0.61 | 0.67 | 0.89 | 0.84 | 0 | 6 | 0 | 2 |
| Social | 4046 | 3873 | 95.7 | 62.0 | 90 | 0.989 | 0.91<br>0.91 | 0.90<br>0.91 | 0.93 | 0.91 | 0 | 6 | 3/4 = 0.35<br>9/10 = 0.43 | 0 |
| Sexual | 2800 | 2408 | 86.0 | 56.7 | 40 | 0.04 | 0.76 | 0.70 | 0.86 | 0.77 | 0 | 4 | 0 | 0 |
| Body image | 3655 | 3267 | 89.4 | 69.0 | 63 | 0.28 | 0.94 | 0.93 | 0.96 | 0.94 | 0 | 4 | 0 | 0 |
| Appearance Distress | 1215 | 1153 | 94.9 | 126.1 | 64 | <0.001 | 0.84 | 0.83 | 0.89 | 0.87 | 0 | 7 | 0 | 1 |
| Work | 1388 | 1165 | 83.9 | 118.0 | 80 | 0.004 | 0.83<br>0.79 | 0.83<br>0.79 | 0.91 | 0.88 | 0 | 3 | 3/5 = 0.36<br>9/10 = 0.42 | 1 |
| Eating symptoms | 2624 | 2212 | 84.3 | 81.9 | 105 | 0.95 | 0.63<br>0.60 | 0.63<br>0.60 | 0.85 | 0.82 | 0 | 8 | 7/9 = 0.50 | 0 |
| Eating distress | 2624 | 1961 | 74.7 | 68.4 | 90 | 0.96 | 0.85 | 0.88 | 0.94 | 0.92 | 0 | 5 | 0 | 0 |
| Eating behavior | 2625 | 2609 | 99.4 | 123.0 | 81 | 0.002 | 0.78<br>0.74 | 0.77<br>0.72 | 0.80 | 0.78 | 0 | 6 | 5/7 = 0.45 | 1 |

DF, degrees of freedom; α, Cronbach alpha; +/−extr, with/without extremes; DT, disordered thresholds.

**Table 2. Pearson Correlations between Scales**

| | Body | Abdomen | Arms | Back | Buttocks | Hips | Thighs | Chest | Nipples | Stretch M. | Skin | Cellulite | Distress | Psych | Social | Sexual | Physical | Body Image | Work | Eat Distress | Eat Symptoms |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Abdomen | 0.75* | | | | | | | | | | | | | | | | | | | | |
| Arms | 0.58* | 0.42* | | | | | | | | | | | | | | | | | | | |
| Back | 0.55* | 0.41* | 0.58* | | | | | | | | | | | | | | | | | | |
| Buttocks | 0.59* | 0.39* | 0.52* | 0.51* | | | | | | | | | | | | | | | | | |
| Hips | 0.62* | 0.45* | 0.45* | 0.48* | 0.67* | | | | | | | | | | | | | | | | |
| Thighs | 0.63* | 0.50* | 0.56* | 0.49* | 0.61* | 0.78* | | | | | | | | | | | | | | | |
| Chest | 0.69* | 0.52* | 0.52* | 0.48* | 0.51* | 0.53* | 0.53* | | | | | | | | | | | | | | |
| Nipples | 0.34* | 0.27* | 0.37* | 0.34* | 0.38* | 0.43* | 0.42* | 0.44* | | | | | | | | | | | | | |
| Stretch M. | 0.40* | 0.33* | 0.32* | 0.28* | 0.31* | 0.37* | 0.42* | 0.32* | 0.23* | | | | | | | | | | | | |
| Skin | 0.55* | 0.59* | 0.59* | 0.25* | 0.38* | 0.44* | 0.48* | 0.36* | 0.23† | 0.63* | | | | | | | | | | | |
| Cellulite | 0.42* | 0.20 | 0.38* | 0.18 | 0.43* | 0.39* | 0.44* | — | — | 0.08 | 0.58† | | | | | | | | | | |
| Distress | —0.62* | —0.46* | —0.46* | —0.41* | —0.45* | —0.49* | —0.48* | —0.40* | —0.25* | —0.35* | —0.55* | —0.25 | | | | | | | | | |
| Psych | 0.57* | 0.46* | 0.44* | 0.41* | 0.45* | 0.47* | 0.49* | 0.52* | 0.34* | 0.35* | 0.35* | 0.06 | —0.70* | | | | | | | | |
| Social | 0.43* | 0.32* | 0.35* | 0.33* | 0.35* | 0.32* | 0.33* | 0.39* | 0.25* | 0.27* | 0.24* | —0.02 | —0.60* | 0.69* | | | | | | | |
| Sexual | 0.55* | 0.34* | 0.37* | 0.36* | 0.39* | 0.38* | 0.39* | 0.40* | 0.32* | 0.33* | 0.38* | 0.18 | —0.49* | 0.55* | 0.48* | | | | | | |
| Physical | 0.30* | 0.19* | 0.21* | 0.23* | 0.21* | 0.24* | 0.22* | 0.24* | 0.23* | 0.23* | 0.25* | —0.09 | —0.25* | 0.25* | 0.17* | 0.29* | | | | | |
| Body image | 0.83* | 0.68* | 0.54* | 0.52* | 0.56* | 0.61* | 0.62* | 0.68* | 0.39* | 0.39* | 0.57* | 0.30† | —0.67* | 0.68* | 0.49* | 0.61* | 0.32* | | | | |
| Work | 0.55* | 0.41* | 0.46* | 0.39* | 0.44* | 0.46* | 0.42* | 0.54* | 0.33* | 0.47* | 0.37* | —0.04 | —0.63* | 0.63* | 0.66* | 0.51* | 0.27* | 0.59* | | | |
| Eat distress | 0.52* | 0.46* | 0.40* | 0.43* | 0.36* | 0.41* | 0.51* | 0.32* | 0.31* | 0.41* | 0.46* | 0.19 | —0.66* | 0.51* | 0.40* | 0.41* | 0.29* | 0.53* | 0.48* | | |
| Eat symptoms | 0.30* | 0.19* | 0.18* | 0.24* | 0.24* | 0.21* | 0.22* | 0.22* | 0.18* | 0.24* | 0.34* | 0.79 | —0.33* | 0.33* | 0.27* | 0.26* | 0.40* | 0.30* | 0.34* | 0.43* | |
| Eat behavior | 0.41* | 0.40* | 0.33* | 0.32* | 0.32* | 0.34* | 0.36* | 0.42* | 0.30* | 0.28* | 0.14* | 0.08 | —0.54* | 0.46* | 0.42* | 0.30* | 0.20* | 0.42* | 0.43* | 0.49* | 0.30* |

Stretch M, stretch marks; psych, psychological; physical, physical function; eat distress, eating-related distress; eat symptoms, eating-related symptoms; eat behavior, eating behavior.
*Correlation is significant at the 0.01 level (two-tailed).
†Correlation is significant at the 0.05 level (two-tailed).

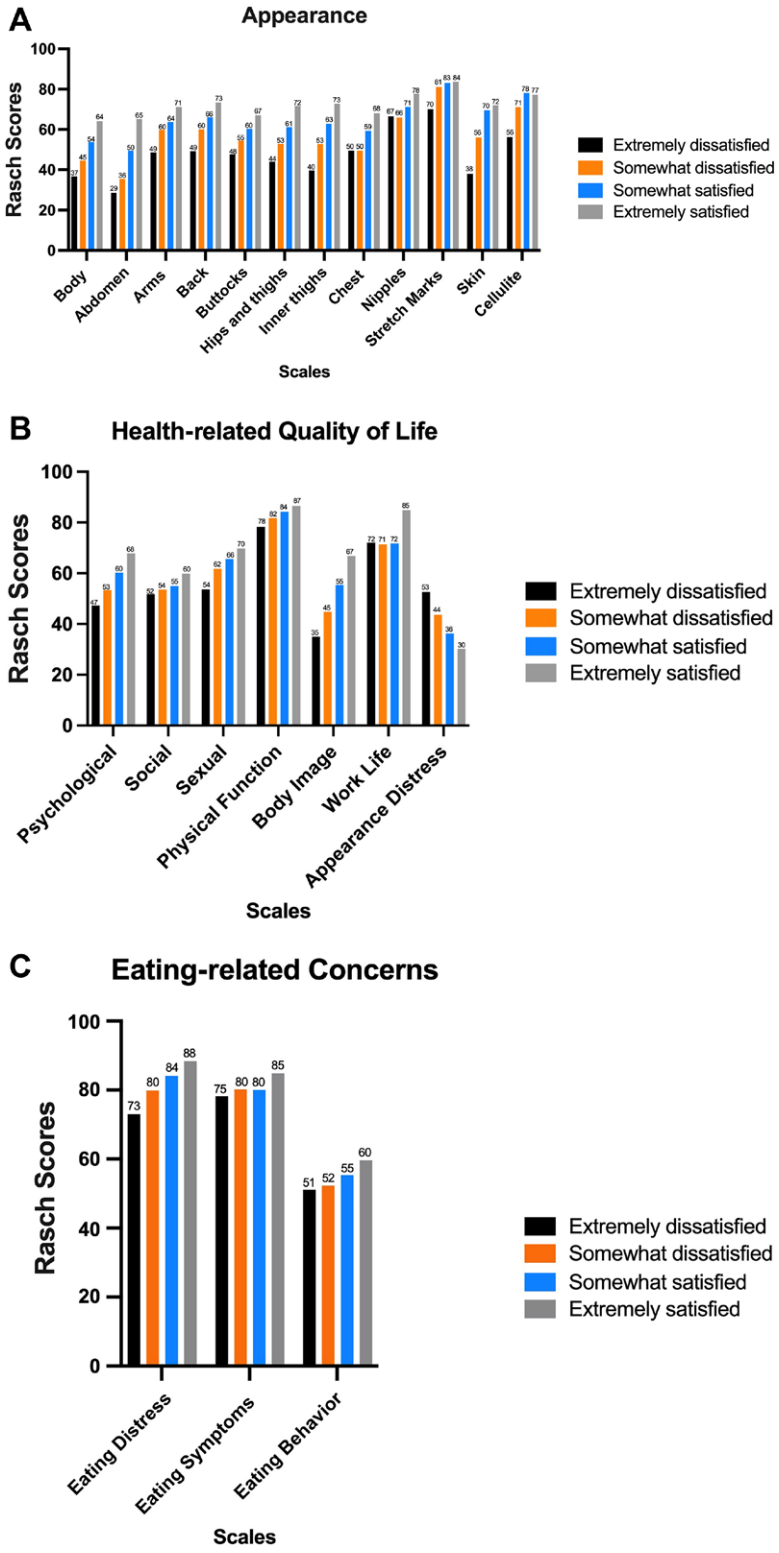


**Fig. 2.** How participants reported their BODY-Q scores based on their satisfaction levels with their current weight for the following BODY-Q domains: A, appearance; B, health-related quality of life; C, eating-related concerns.

who differed by age, gender, continent, and language. All items had ordered thresholds and high reliability, with Cronbach alpha values for all scales of 0.80 or more and PSI values of 0.70 or more for all but two scales (Physical Function and Eating Symptoms). The Cronbach alphas were 0.85 or more for these two scales, providing evidence of reliability.

Both hypotheses of construct validity were fully or partially supported with the majority of correlations between scales meeting COSMIN criteria for construct validity (ie, ≥0.50 for similar constructs, 0.30 to 0.50 for related but dissimilar constructs, and <0.30 for unrelated constructs.[27]) Higher BODY-Q scores were detected with higher levels of satisfaction and lower appearance distress for 21 of 22 scales. The only exception was in the cellulite scale, where patients scored higher when "a little bothered" compared with "not at all bothered." However, the sample size was small in the cellulite group (n = 466), which might explain why this scale only partially supported the expected hypothesis, and participants had to have cellulite to answer this scale.[20]

Within the RMT, we observed some degree of misfit on item and scale level. For the item fit statistics, item fit was outside the criteria of -2.5 to +2.5 for 98 of 176 items; of these, only nine items had a significant chi-square *P* value after Bonferroni adjustment. Additionally, in eight of nine items with significant chi-square (except the arm item), all the items occupied the end of the Rasch ruler. This solely indicated that the observed response to these items did not fit the Rasch model perfectly. The abdomen, nipples, physical function, appearance distress, and eating behavior scales showed more item misfit. This might be due to the heterogeneity between the original field-test sample and the normative sample.[6,8,10–12] The field-test samples included participants who were more likely woman with a higher body mass index than the normative sample. Importantly, the field-test sample was clinical and included people seeking weight loss treatments and treatment to improve appearance of various body parts such as their abdomen, thighs, and upper arms.[6,8,12,23] As the two samples differ, we would not expect to achieve the same results, and misfit of data to the Rasch model might be due to patients scoring differently than the normative population on these specific scales. DIF was significant for 31 items spread over 13 scales. In 11 of 13 scales, there was DIF in less than 30% of the items. However, DIF had a negligible impact on scores after splitting for DIF with Pearson correlations more than 0.995 for all scales with DIF. Based on our findings, we recommend the use of the tables presented in this study and the use of the original scoring key in the normative population.[20]

Prior outcome studies that used the BODY-Q to measure changes in HRQL and satisfaction with appearance in weight loss and BC patients demonstrated improvements in mean BODY-Q scores after weight loss.[3,28] Additionally, patients who underwent subsequent BC after weight loss scored higher compared with those who did not.[15,29–31] However, a limitation in these studies was the inability to juxtapose patients' scores with reference values from the general population. Hence, there has been a crucial need for a reference point for weight loss and BC patients for each scale used in the BODY-Q to enable accurate comparison with the general population. The previously published normative study aided in enhancing the interpretation of BODY-Q data to understand the actual impact of weight loss and BC on different areas of patient's lives.[20] This study is a key supplement when generating BODY-Q normative scores, as it explores the psychometric properties of the BODY-Q in the general population to assure the validity and reliability of the reference scores. The insights derived from this study have significant potential to improve patient treatment, follow-ups, and clinical decisions. It enables healthcare providers to better understand patients' progress and state of recovery in relation to the general population, thereby allowing for more comprehensive, empathetic, and effective patient management. Future research, clinical care, and healthcare policy could potentially benefit from these findings.

The strengths of our study were the large international sample of 4051 participants. Furthermore, using the online crowdsourcing databases Prolific and MTurk has been shown to be a valid and reliable method for recruitment of research participants, facilitating cross-cultural and international research with low costs.[32–35] Our study had some limitations. Recruitment of the normative sample via the crowdsourcing platforms is a potential limitation, as it is unknown whether the sample is a representative sample of the general population of the included countries. Furthermore, online surveys do not reach participants who do not have internet access or a reliable device to access the internet, have low levels of digital literacy, or have other physical or cognitive limitations that prevent participants in online research. The majority of the participants were White (87.9%), whereas only 12.1% identified themselves as another ethnicity. The racial homogeneity should therefore be considered with caution when interpreting the normative results, as it limits its applicability. Finally, participants were paid to participate, which may have impacted incentives of participation in this study and their responses to the questions. Some psychometric properties (eg, responsiveness and test-retest reliability) could not be evaluated in this study due to the cross-sectional nature of data collection, and hence, should be the focus of future studies.

## CONCLUSIONS

The RMT analysis provided broad support of the reliability and validity of the normative values of the BODY-Q scales with high PSI and Cronbach alpha and evidence of construct validity. This study confirms that the international normative sample can be used as reference values for the general population for interpreting clinical and research data in research.

***Farima Dalaei, MD***
Research Unit of Plastic Surgery
Odense University Hospital and
University of Southern Denmark
J.B. Winsløws Vej 4, Entrance 20, Penthouse 2. floor
5000 Odense C, Denmark
E-mail: farima.dalaei@gmail.com

## DISCLOSURES

*The BODY-Q is co-developed by Anne F. Klassen, Stefan J. Cano, and Andrea L. Pusic, and they receive a share of any license revenues based on their institution's inventor sharing policy. Anne F. Klassen is an owner of EVENTUM Research, which provides consulting services to the pharmaceutical industry. Stefan J. Cano is CSO of Modus Outcomes, a Division of Thread. Manraj Kaur is a recipient of the Canadian Institute of Health Research Fellowship Award (2020–23). All the other authors have no financial interest to declare in relation to the content of this article. This study was funded by research grants from Odense University Hospital (A5006), the Region of Southern Denmark (21/17592), and the Jascha Fund (2021-0183).*

## ACKNOWLEDGMENT

*The authors thank all participants who completed the BODY-Q questionnaire for generating the population norms.*

## REFERENCES


1. Yumuk V, Tsigos C, Fried M, et al; Obesity Management Task Force of the European Association for the Study of Obesity. European guidelines for obesity management in adults. *Obes Facts*. 2015;8:402–424.
2. Bray GA, Frühbeck G, Ryan DH, et al. Management of obesity. *Lancet*. 2016;387:1947–1956.
3. Poulsen L, Klassen A, Rose M, et al. Patient-reported outcomes in weight loss and body contouring surgery: a cross-sectional analysis using the BODY-Q. *Plast Reconstr Surg*. 2017;140:491–500.
4. de Vries CEE, Klassen AF, Hoogbergen MM, et al. Measuring outcomes in cosmetic abdominoplasty: the BODY-Q. *Clin Plast Surg*. 2020;47:429–436.
5. Sharma K, Steele K, Birks M, et al. Patient-reported outcome measures in plastic surgery: an introduction and review of clinical applications. *Ann Plast Surg*. 2019;83:247–252.
6. 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:e679.
7. Klassen AF, Cano SJ, Pusic AL. BODY-Q Framework. 2021. Available at https://qportfolio.org/body-q/. Accessed in December 25, 2021.
8. 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;31:3637–3645.
9. 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;142:1600–1608.
10. Poulsen L, Pusic A, Robson S, et al. The BODY-Q stretch marks scale: a development and validation study. *Aesthet Surg J*. 2018;38:990–997.
11. Klassen AF, Kaur MN, de Vries CEE, et al. The BODY-Q cellulite scale: a development and validation study. *Aesthet Surg J*. 2021;41:206–217.
12. Klassen AF, Cano SJ, Alderman A, et al. Self-report scales to measure expectations and appearance-related psychosocial distress in patients seeking cosmetic treatments. *Aesthet Surg J*. 2016;36:1068–1078.
13. Barone M, Cogliandro A, Tsangaris E, et al. Treatment of severe gynecomastia after massive weight loss: analysis of long-term outcomes measured with the Italian version of the BODY-Q. *Aesthetic Plast Surg*. 2018;42:1506–1518.
14. Christopher AN, Morris MP, Patel V, et al. abdominal body contouring: does body mass index affect clinical and patient reported outcomes? *J Surg Res*. 2021;270:348–358.
15. Elfanagely O, Mauch JT, Mellia JA, et al. Quality of life and concurrent procedures in truncal body contouring patients: a single-center retrospective study. *Aesthetic Plast Surg*. 2021;45:1620–1627.
16. Jørgensen MG, Cho BY, Hansen FG, et al. Patients decision-making characteristics affects gynecomastia treatment satisfaction: a multicenter study using the BODY-Q chest module. *Aesthetic Plast Surg*. 2022;46:1053–1062.
17. 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;19:1395–1411.
18. Barone M, Cogliandro A, Salzillo R, et al. Patient-reported satisfaction following post-bariatric surgery: a systematic review. *Aesthetic Plast Surg*. 2018;42:1320–1330.
19. de Vries CEE, Terwee CB, Al Nawas M, et al. Outcomes of the first global multidisciplinary consensus meeting including persons living with obesity to standardize patient-reported outcome measurement in obesity treatment research. *Obes Rev*. 2022;23:e13452.
20. Dalaei F, de Vries CEE, Poulsen L, et al. General population normative scores for interpreting the BODY-Q. *Clin Obes*. 2022;12:e12528.
21. Christensen KB, Comins JD, Krogsgaard MR, et al. Psychometric validation of PROM instruments. *Scand J Med Sci Sports*. 2021;31:1225–1238.
22. Boone WJ. Rasch analysis for instrument development: why, when, and how? *CBE Life Sci Educ*. 2016;15:rm4.
23. 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:e1529.
24. Prolific 2020. April 2004. Available at https://prolific.co. Accessed September 20, 2020.
25. Amazon Mechanical Turk. November 2005. Available at https://www.mturk.com. Accessed February 9, 2021.
26. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol*. 2007;60:34–42.
27. Prinsen CAC, Mokkink LB, Bouter LM, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. 2018;27:1147–1157.
28. Mou D, DeVries CEE, Pater N, et al. BODY-Q patient-reported outcomes measure (PROM) to assess sleeve gastrectomy vs Roux-en-Y gastric bypass: eating behavior, eating-related distress, and eating-related symptoms. *Surg Endosc*. 2021;35:4609–4617.
29. Geerards D, Klassen AF, Hoogbergen MM, et al. Streamlining the assessment of patient-reported outcomes in weight loss and body contouring patients: applying computerized adaptive testing to the BODY-Q. *Plast Reconstr Surg*. 2019;143:946e–955e.
30. Paul MA, Opyrchał J, Knakiewicz M, et al. The long-term effect of body contouring procedures on the quality of life in morbidly obese patients after bariatric surgery. *PLoS One*. 2020;15:e0229138.
31. Uimonen M, Repo JP, Homsy P, et al. Health-related quality of life in patients having undergone abdominoplasty after massive weight loss. *J Plast Reconstr Aesthet Surg*. 2021;74:2296–2302.
32. Strickland JC, Stoops WW. The use of crowdsourcing in addiction science research: Amazon Mechanical Turk. *Exp Clin Psychopharmacol*. 2019;27:1–18.
33. Chandler J, Shapiro D. Conducting clinical research using crowdsourced convenience samples. *Annu Rev Clin Psychol*. 2016;12:53–81.
34. Ibarra JL, Agas JM, Lee M, et al. Comparison of online survey recruitment platforms for hard-to-reach pregnant smoking populations: feasibility study. *JMIR Res Protoc*. 2018;7:e101.
35. Eriksson K, Simpson B. Emotional reactions to losing explain gender differences in entering a risky lottery. *Judgment and Decision Making*. 2010;5:159–163.