Check for updates

Obesity

WILEY | THE OBESITY SOCIETY

ORIGINAL ARTICLE OPEN ACCESS

Clinical Trials and Investigations

BODY-Q-7D Measures Consensus-Based Patient-Reported Outcomes for the Management of Obesity Treatment

Anne F. Klassen[1] | Charlene Rae[1] | Claire E. E. de Vries[2] | Farmina Dalaei[3] | Manraj N. Kaur[4] | Lucas Gallo[5] | Maarten Hoogbergen[6] | Lotte Poulsen[7] | Andrea L. Pusic[4] | Stefan J. Cano[8]

[1]Department of Pediatrics, McMaster University, Hamilton, Ontario, Canada | [2]Department of Plastic Surgery, Erasmus Medical Centre, Amsterdam, the Netherlands | [3]Department of Plastic Surgery, Odense University Hospital, Odense, Denmark | [4]Department of Surgery, Brigham and Women's Hospital, Boston, Massachusetts, USA | [5]Department of Surgery, McMaster University, Hamilton, Ontario, Canada | [6]Department of Plastic Surgery, Catharina Hospital Eindhoven, Eindhoven, the Netherlands | [7]Lindehaven Plastic Surgery, Odense, Denmark | [8]Modus Outcomes (A Division of Thread), Cheltenham, UK

Correspondence: Anne F. Klassen (aklass@mcmaster.ca)

Received: 26 February 2026 | **Revised:** 26 February 2026 | **Accepted:** 25 May 2026

Keywords: BODY-Q | health-related quality of life | preference-based measure | weight loss

ABSTRACT

Objective: A global multidisciplinary consensus initiative identified eight patient-reported outcome domains important in obesity care: self-esteem; physical, mental, and social health; eating; stigma; body image; and excess skin. Several of these domains are not represented in generic preference-based measures typically used in weight loss research. We developed the BODY-Q-7 Dimensions (BODY-Q-7D) to provide a valuation-ready framework for a preference-based instrument.

Methods: Concepts from the BODY-Q were used to inform the development of items that covered the eight consensus initiative domains. Psychometric performance of the BODY-Q-7D was examined using an online international sample (i.e., Prolific).

Results: Content validity was established with 123 participants. The field test included 770 participants, of whom 269 completed a test–retest, and 599 completed a 3-month follow-up. Rasch Measurement Theory analysis reduced the BODY-Q-7D to 7 items. Items fit the Rasch model with ordered thresholds and structural independence. Reliability was >0.80 across three reliability coefficients. Hypothesis-based tests of construct validity and responsiveness were supported with 22/28 (79%) and 12/16 (75%) tests accepted.

Conclusions: The BODY-Q-7D demonstrated strong psychometric performance in a weight loss sample. The next phase of development is to elicit and model preference weights to assign utility values to each defined health state.

1 | Introduction

Obesity is an urgent public health concern associated with high healthcare costs. In 2020, overweight and obesity affected 38% of the global population, a proportion projected to exceed 50% by 2035 without improved prevention and treatment [1]. The global economic burden for this condition is expected to increase from 2.4% to 2.9%, corresponding to an annual cost of US $4.32 trillion by 2035 [1]. These costs underscore the need for healthcare systems to prioritize cost-effective weight loss interventions.

Treatments for obesity can have a substantial impact on health-related quality of life (HRQL) [2]. Effective interventions include metabolic bariatric surgery (MBS), which can result in sustained weight loss and improvement or resolution of many obesity-related comorbidities [3]. Behavioral and pharmacological interventions, including recent glucagon-like peptide-1 (GLP-1) receptor agonists, also produce clinically meaningful weight loss [4–6]. In health economic evaluations of weight loss treatments, HRQL is commonly measured from the patient perspective using preference-based measures (PBMs)

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *Obesity* published by Wiley Periodicals LLC on behalf of The Obesity Society.

Study Importance

- What is already known?
  - A global multidisciplinary consensus initiative identified eight overarching outcome domains considered important by people living with obesity. Generic preference-based measures lack content validity for obesity research.
- What does this study add?
  - BODY-Q-7D was developed with extensive input from people living with obesity and experts.
  - BODY-Q-7D measures outcome domains important in the multidisciplinary management of obesity.
- How might these results change the direction of research or the focus of clinical practice?
  - BODY-Q-7D health states can be valued for use in economic evaluations of treatments for weight loss.

to estimate quality-adjusted life years (QALYs). Most health economic evaluations of weight loss treatments have relied on generic PBMs to estimate QALYs, including the EQ-5D, EQ-5D-5L, and SF-6D [6, 7]. However, generic instruments may lack sensitivity to detect change in obesity-specific aspects of HRQL following multidisciplinary management of obesity as a chronic disease.

Another approach in health economic evaluations has been to use the scores from an obesity-specific patient-reported outcome measure (PROM) to predict utility scores on a generic PBM [8–12]. PROMs that have been used to map content to generic PBMs include the Moorehead-Ardelt II (MA-II) [8], the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) [9, 10], and the Obesity Problems Scale (OP-scale) [11]. The mapping approach has recognized limitations, particularly in accurately reflecting obesity-specific concerns that are not well represented in generic PBMs. Overall, since both approaches rely on generic PBMs to generate QALYs, the scores can underestimate QALY gains, potentially biasing cost-effectiveness estimates and influencing reimbursement and coverage decisions for weight loss interventions.

To address the need for standardized outcome measurement in obesity research, the Standardize Quality of life measurement in Obesity Treatment (S.Q.O.T.) initiative was established [12–14]. This initiative aimed to identify a set of overarching outcome domains most relevant to obesity research and clinical care. A literature review identified 25 candidate patient-reported outcomes (PROs) [15]. These outcomes were prioritized during international multidisciplinary consensus meetings that involved people living with obesity and healthcare professionals [12–14]. Consensus was reached that the following eight overarching outcome domains are important to measure: self-esteem, physical health/function, mental/psychological health, social health, eating, stigma, body image, and excess skin.

Since generic PBMs do not capture all eight outcomes, a weight-loss-specific PBM is needed to support accurate and meaningful health economic evaluations. BODY-Q is a validated, modular obesity-specific PROM that can be used to measure outcomes related to weight loss treatments (e.g., diet, exercise, bariatric surgery/medicine) and body contouring to remove excess skin after weight loss or for cosmetic purposes [16–18]. BODY-Q is composed of multiple independently functioning scales and is therefore not amenable to valuation since it is not possible to combine scores for different scales into a single score.

The aim of this study was to apply Rasch Measurement Theory (RMT) to develop a short-form BODY-Q-7 Dimensions (BODY-Q-7D) that captures the PROs identified by the S.Q.O.T. initiative and forms a robust foundation for a future PBM. The Rasch approach was used to ensure that the dimensions provide invariant, interpretable, and interval-level measurement, key prerequisites for defining health states suitable for valuation. The specific objectives were: (1) to develop the content of the BODY-Q-7D and establish its content validity and (2) to evaluate its psychometric performance in preparation for subsequent health state valuation and utility modeling.

2 | Methods

2.1 | Ethics

Ethical approval for this study was obtained from the Hamilton Integrated Research Ethics Board in Hamilton, Ontario, Canada (Project ID #18651). The study used the Prolific online platform (prolific.com). All surveys were administered using REDCap, a secure platform for electronic data collection, hosted through the Faculty of Health Sciences at McMaster University. Electronic informed consent was obtained from participants prior to completing the study surveys. Compensation was a prorated rate of 12 pounds sterling per hour.

2.2 | Data Collection

2.2.1 | Content Validity

The BODY-Q-7D was developed and validated in accordance with established PROM development guidelines [19–23]. Concepts from the BODY-Q [16–18] were used to create items to cover the eight PROs from the S.Q.O.T. [12–14], forming a draft instrument. The draft was reviewed by S.Q.O.T. initiative members who provided feedback that was used to make revisions. The revised instrument was included in a content validity survey. On May 5, 2025, the Prolific platform was used to screen for residents of Australia, Canada, Ireland, New Zealand, UK, and USA whose BMI was ≥ 20. Those who were trying to lose weight, wanting to lose weight, or had lost weight through MBS or an obesity medication were invited to complete the content validity survey. Table S1 shows the content validity questions.

2.2.2 | Field Test

More Prolific participants were screened on July 10, 2025, to recruit a larger sample for the field test. For data quality,

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | Tests of construct validity.

Hypotheses	Result
Cross-sectional hypotheses	
1. Scores incrementally higher as report being closer to middle or end of weight loss journey	Yes
2. Scores incrementally higher as report being more satisfied with weight	Yes
3. Scores incrementally higher as report having better physical health	Yes
4. Scores incrementally higher as report having better mental health	Yes
5. Scores incrementally higher as report being happier with social life	Yes
6. Scores incrementally higher as report being happier with eating habits	Yes
7. Scores incrementally higher as report being happier with how body looks	Yes
8. Scores incrementally higher as report being more active	Yes
9. Scores incrementally higher as report less interference with ability to do daily activities	Yes
10. Scores higher if do not want body contouring to remove excess skin	Yes
11. Scores higher if had bariatric surgery vs. want to have*	Yes
12. Scores higher if currently taking a GLP-1 treatment*	Yes
13. Scores higher as report more problems due to excess skin	Yes
Correlations	
14. Negative correlation with lower BMI (0.3–0.5)	Yes
15. Negative correlation with lower amount of excess skin on the body (0.3–0.5)	No
16. Negative correlation with lower amount of weight wanting to lose (0.3–0.5)	Yes
17. Positive correlation with BODY-Q Body (0.3–0.5)	No
18. Positive correlation with BODY-Q Body Image (0.3–0.5)	No
19. Positive correlation with BODY-Q Physical Function (0.3–0.5)	No
20. Positive correlation with BODY-Q Psychological (≥0.5)	Yes
21. Positive correlation with BODY-Q Social (0.3–0.5)	No
22. Positive correlation with BODY-Q Sexual (0.3–0.5)	Yes
23. Positive correlation with BODY-Q Eating Behavior (0.3–0.5)	Yes
24. Negative correlation with EQ-5D-5L Self Care (0.3–0.5)	Yes
25. Negative correlation with EQ-5D-5L Mobility (0.3–0.5)	Yes
26. Negative correlation with EQ-5D-5L Usual Activity (0.3–0.5)	Yes
27. Negative correlation with EQ-5D-5L Anxiety and Depression (0.3–0.5)	No
28. Positive correlation with EQ-5D-5L Global score (≥0.5)	Yes
Hypotheses ACCEPTED	22/28 (79%)
Responsiveness hypotheses based on change scores	
1. Change score will be incrementally higher for those who lost weight vs. gained weight or stayed the same	Yes
2. Change score will be higher for those satisfied with change in weight	Yes
3. Those who report weight-related HRQL is worse will have negative change score*	Yes
4. Those who report weight-related HRQL is the same will have no change in score*	No
5. Those who report weight-related HRQL is better will have positive change score*	Yes
6. Change score will be higher for those who lost 5%–10% vs. 0%–4.99%*	Yes
Correlations	
7. Change score will negatively correlate with an increase in BMI (≥0.3)	No
8. Change score will correlate positively with score for composite HRQL variable** (≥0.3)	Yes
9. Change score will correlate positively with BODY-Q Body change score (≥0.3)	Yes
10. Change score will correlate positively with BODY-Q Body Image change score (≥0.3)	Yes
11. Change score will correlate positively with BODY-Q Physical Function change score* (≥0.3)	No
12. Change score will correlate positively with BODY-Q Psychological change score (≥0.3)	Yes

(Continues)

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 | (Continued)

Hypotheses	Result
13. Change score will correlate positively with BODY-Q Social (< 0.3)	Yes
14. Change score will correlate positively with BODY-Q Sexual (< 0.3)	Yes
15. Change score will correlate positively with BODY-Q Eating Behavior (< 0.3)	Yes
16. Change score will correlate positively with BODY-Q with EQ-5D-5L Global change score (≥ 0.5)	No
Hypotheses ACCEPTED	12/16 (75%)

*Analysis was run in a subgroup (see Tables S4 and S6).
***A composite Likert change score was computed by assigning 0 for same score, −1 for worse score, +1 for improved score for the following HRQL questions: happy with how body looks, social life, eating habits, activity level, physical health, mental health, and interference of weight with activities.

participants were asked to repeat the screening survey on July 22, 2025. Those who provided the same answers for type of MBS and/or type, mode of administration, and/or treatment status (current vs. past use) for GLP-1 receptor agonists were invited to complete the field test survey. This survey included demographic and clinical questions, the BODY-Q-7D and other BODY-Q scales (i.e., Body, Body Image, Physical Function, Psychological, Social, Eating Behavior, Sexual) [16–18], and the EQ-5D-5L [24]. The survey remained open until August 9, 2025.

2.2.3 | Test–Retest (TRT)

TRT data were collected between days 7 and 14. Participants who completed the survey within the time frame and reported no change in their weight-related HRQL since the initial assessment were included in the TRT analysis. We aimed to recruit ≥ 100 participants to obtain a "very good" sample size TRT rating according to COSMIN guidelines [22].

2.2.4 | Responsiveness

At least 3 months after the initial survey, participants were invited to complete a follow-up survey. For change in the BODY-Q-7D scale, we asked directionality (i.e., same, worse, or better) and magnitude (e.g., a little, somewhat, quite a bit, a lot). Participants were reminded of their baseline answers for constructs using the piping function in REDCap. Reminders are recommended by Devji et al. [25], who reported that people have difficulty recalling a previous health state after 4 weeks.

2.3 | Analysis

Content validity results were analyzed descriptively. Table S2 shows the psychometric tests performed. For the field test data, RMT analysis was used to examine the response threshold order, item fit, local dependency, targeting, differential item functioning (DIF), and reliability [26–29]. The probabilistic Rasch model compares the relationship between the ability of the people in the sample and the difficulty of the items in a scale. When data for a scale fit the Rasch model, the scale provides invariant, interpretable, and interval-level measurement. RUMM2030 software (RUMM Laboratory) was used with the unrestricted partial credit model for polytomous data. Rasch logits for the final scale were used to convert raw scores for the BODY-Q-7D into scores that ranged from 0 (worst) to 100 (best). Other BODY-Q scales and the EQ-5D-5L were scored according to developers' instructions.

For TRT, single and average intraclass correlation coefficients (ICC) were calculated with and without outliers identified using box plots. Analysis used a two-way mixed effects model for consistency. ICC values ≥ 0.7 were indicative of acceptable reliability [22]. The smallest detectable change (SDC) was computed at the individual [$1.96*\sqrt{(2)}*\text{SEM}$] and group [$\text{SDCind}/\sqrt{n}$] level [30]. The standard error of the mean (SEM) for SDC calculations was determined as follows: $(SD_{T1} + SD_{T2})/2*\sqrt{(1 - \text{ICC})}$ [22].

Table 1 shows 28 and 16 predetermined hypotheses for the initial and responsiveness surveys, respectively. Variables used to examine group differences in the initial survey included: height, weight, and goal weight; location in the weight loss journey (not started, close to the start, about halfway through, near completion, finished); satisfaction with weight (extremely dissatisfied to extremely satisfied); physical and mental health (excellent, very good, good, fair, poor); social life, eating habits, and how the body looks (extremely unhappy to extremely happy); activity level (not at all to very active); how often weight interfered with ability to do usual daily activities (never, rarely, sometimes, often, always); MBS (want vs. had); GLP-1 receptor agonist (want vs. currently having); body contouring (want vs. do not want); a computed score for amount (none, a little, a moderate amount, a lot) of excess skin across seven parts of the body; and severity of excess skin problems in the past week (none, little bad, somewhat bad, very bad). Additional variables in the follow-up survey used to examine construct validity included: directionality of change (same, worse, better); magnitude of change (a little, somewhat, quite a bit, a lot); and satisfaction with change in weight (dissatisfied vs. satisfied).

Hypothesized correlations between BODY-Q-7D scores, clinical variables, and scores for other BODY-Q scales and the EQ-5D-5L were based on the relatedness of constructs using magnitudes that are commonly applied in COSMIN-based hypothesis testing [21]. Specifically, COSMIN considers correlations ≥ 0.5 to indicate similar constructs, 0.3–0.5 to indicate related but dissimilar constructs, and < 0.30 to indicate weak or unrelated constructs. For the field test data, we hypothesized that the strength of correlations between the BODY-Q-7D and the BODY-Q Psychological scale and the EQ-5D-5L would be ≥ 0.5 and with clinical variables (e.g., BMI), other BODY-Q scales, and EQ-5D-5L subscales would be 0.30–0.50. For the change scores, we hypothesized the direction and the relative magnitude of change scores. Associations were expected to be weak ($r < 0.30$), moderate (0.30–0.49), or strong (≥ 0.50) depending on the conceptual similarity between constructs measured and how much each construct was anticipated to change [31]. Construct validity

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

was considered adequate if ≥ 75% of the hypothesis tests were confirmed as per COSMIN criteria [20, 22].

Analysis took place in SPSS version 30. ANOVA and independent *t*-tests were used to assess group differences. A one-sided *t*-test was used for testing differences between two groups when directionality was hypothesized. Paired *t*-tests were used to examine change for the responsiveness analysis. Pearson correlations were used to assess convergent validity. Normality of data was assessed using kurtosis and skewness and visual plots and nonparametric statistics were applied if distributions were considered non- normal [32]. Statistical significance was set at $p < 0.05$. Construct validity was based on the consistency of results for the hypotheses rather than the individual *p* values. As such, adjustment for multiple comparisons was not considered appropriate.

3 | Results

Figure 1 shows the flow of participants and reasons for exclusions at each step of the study.

3.1 | Content Validity

The first draft of the BODY-Q-7D included 11 questions covering the eight S.O.Q.T. outcomes. The scale included four response options and the following instructions: "These questions ask about how OBESITY and WEIGHT LOSS have affected you. Please answer each question based on the PAST WEEK." This version was sent by email to 16 S.Q.O.T. committee members (2 obesity doctors, 1 endocrinologist, 2 psychologists, 2 dietitians, 2 health economists, and 1 nurse practitioner) and organizers of the meetings (3 surgeons and 3 researchers). Feedback led to several changes. The instructions were revised to broaden the scope: "These questions ask about the impact of your WEIGHT on your HEALTH. Please answer each question based on the PAST WEEK." A question deemed outside the scope of the eight outcomes was dropped (i.e., "Did you experience any unpleasant symptoms related to eating [e.g., heartburn, nausea, vomiting]?"). A question was added to provide an alternative for body image (i.e., "How do you feel about your body?"). Minor wording changes were made to better align questions with the response options.

A total of 123 participants completed the content validity survey. Table 2 shows the participant characteristics and Table 3 shows the type of obesity medication participants were currently taking or had in the past. The sample ranged in age from 20 to 75 years (mean = 41.7, SD = 12.7). When asked if they were currently trying to lose weight, 103 (83.7%) said yes, 14 (11.4%) said no but that they wanted to, and 6 (4.9%) said no. All participants indicated that the BODY-Q-7D instructions were easy to understand. Table 4 shows the content validity results. The response options and 11 questions were generally deemed easy to understand and reflective of participant experiences. For questions measuring similar concepts, some participants found "moderate exercise" more relevant than "walk or move around." The question "Did you experience any unpleasant symptoms due to excess skin caused by weight loss (e.g., rash, infections, odor)?" was considered not relevant by 11 participants and comments such as the following were provided: "No excess skin, I've never been worried about this," "Me personally I had no extreme weight loss causing excess skin," "I've never had problems with my skin due to my weight loss in the past ever," "I've never lost enough weight to have much in the way of excess skin."

For recall period, participants were asked whether they could reflect on their experiences over the "past week." Of the 123 participants, 118 (95.9%) said yes. Most participants ($N = 86$, 69.9%) also felt that the past week was an appropriate time frame, although 34 (27.6%) participants considered it too short and 3 (2.4%) too long. Recommendations for a longer time frame ranged from 2 weeks to more than 1 year, while shorter time frame suggestions ranged from 4 to 5 days.

For missing content, 105 (85.4%) said that no relevant aspects of weight impacting their health were missing. Missing concepts mentioned by at least three participants included energy and sleep quality. Feedback about the scale was overall positive, such as the following: "It gave me a sense of trust in the professionalism of the party posing the questions. The questions themselves were very inoffensive and comfortable to answer. I wouldn't feel awkward while answering the questions or bad after I finished."

3.2 | Field Test

All 11 questions were included in the field test, which involved 770 participants. Table 2 shows the participant characteristics and Table 3 shows the reported use of weight loss medication. Participants ranged in age from 19 to 80 years (mean = 43.2, SD = 11.9). When asked if they were currently trying to lose weight: 618 (80.3%) said yes, 142 (18.4%) said no but they wanted to, and 10 (1.3%) said no. Regarding their weight loss journey: 72 (9.4%) had not started, 331 (43%) were close to the start, 251 (32.6%) were about halfway through, 106 (13.8%) were near completion, and 10 (1.3%) were finished.

Table 5 shows the category frequency responses. For two questions, more than half the sample reported no problems: 479 (62.2%) participants reported they did not experience any unpleasant symptoms due to excess skin caused by weight loss; and 417 (54.2%) participants reported that it was not difficult to walk or move around. The excess skin question also evidenced poor fit to the Rasch model ($p < 0.001$). After this question was dropped, the remaining 10 questions fit the Rasch model and had ordered thresholds.

For the three outcomes with two candidate questions, we picked the questions with a more balanced distribution of responses, i.e., "How difficult was it for you to do moderate exercise (e.g., go for a brisk walk)?", "How difficult was it for you to feel in control when you ate?", and "How self-conscious were you about your body?" After reducing the scale, the overall model fit for the 7-item scale was as follows: chi-square = 81.29, df = 56, $p = 0.02$.

Table S3 shows the item fit statistics and DIF results. All seven questions fit the Rasch model with nonsignificant *p* values after Bonferroni correction and had ordered thresholds

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

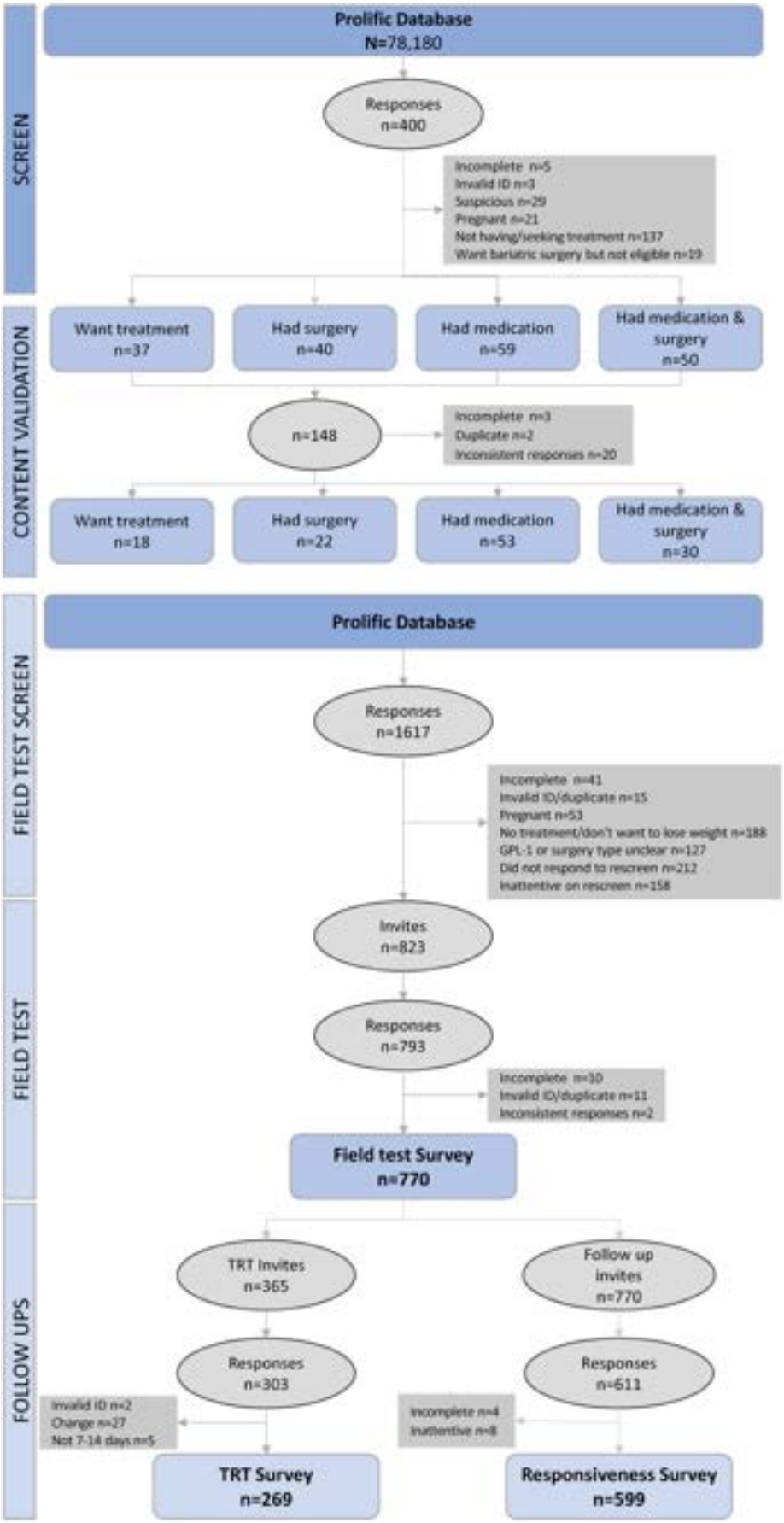

FIGURE 1 | Recruitment for each component of the study. [Color figure can be viewed at wileyonlinelibrary.com]

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | Sample characteristics.

Characteristic	Subgroups	Content validity		Field test		Follow-up	
		N = 123	%	*N* = 770	%	*N* = 599	%
What country do you live in?	USA	80	65.0	384	49.9	299	49.9
	Canada	7	5.7	73	9.5	56	9.3
	UK	35	28.5	275	35.7	217	36.2
	Australia	0	0	23	3.0	18	3.0
	Ireland	0	0	11	1.4	5	0.8
	New Zealand	1	0.8	4	0.5	4	0.7
What is your age?	18–34	41	33.3	206	26.8	149	24.9
	35–44	41	33.3	230	29.9	169	28.2
	45–54	18	14.6	192	24.9	154	25.7
	≥ 55	23	18.7	142	18.4	127	21.2
What is your gender?	Man	65	52.8	223	29.0	182	30.4
	Woman	57	46.3	534	69.4	410	68.4
	Nonbinary	1	0.8	11	1.4	5	0.8
	Other gender	0	0	2	0.3	2	0.3
What racial group(s) do you belong to?	Black	25	20.3	72	9.4	57	9.5
	East Asian	3	2.4	2	0.3	2	0.3
	Southeast Asian	1	0.8	6	0.8	5	0.8
	Latin American	3	2.4	12	1.6	7	1.2
	Middle Eastern	1	0.8	2	0.3	2	0.3
	South Asian	2	1.6	16	2.1	12	2.0
	White	84	68.3	620	80.5	484	80.8
	Multiple races	2	1.6	33	4.3	24	4.0
	Prefer not to answer/ Missing/Other	2	1.6	1	0.1	1	0.2
	Indigenous	0	0	4	0.5	3	0.5
	Pacific Islander	0	0	2	0.3	2	0.3
As of today, what is the highest level of education you have completed?	Some high school	1	0.8	8	1.0	8	1.3
	Completed high school	6	4.8	81	10.5	61	10.2
	Some college or trade school or university	18	14.6	168	21.8	130	21.7
	Completed college or trade school or university	66	53.7	350	45.5	274	45.7
	Some master's or doctoral degree	10	8.1	39	5.1	32	5.3
	Completed master's or doctoral degree	22	17.9	124	16.1	94	15.7

(Continues)

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 | (Continued)

Characteristic	Subgroups	Content validity		Field test		Follow-up	
		N = 123	%	*N* = 770	%	*N* = 599	%
In the past 3 months, how difficult was it for you to cover your household expenses and pay your bills?	Not at all difficult	34	27.6	219	28.4	177	29.5
	A little difficult	46	37.4	245	31.8	192	32.1
	Somewhat difficult	34	27.6	167	21.7	124	20.7
	Very difficult	6	4.9	81	10.5	63	10.5
	Extremely difficult	3	2.4	56	7.3	41	6.8
	Prefer to not answer	0	0	2	0.3	2	0.3
What is your current marital status?	Never married	35	28.5	245	31.8	184	30.7
	Separated	2	1.6	21	2.7	17	2.8
	Divorced	5	4.1	78	10.1	65	10.9
	Widowed	2	1.6	11	1.4	10	1.7
	Living common-law	12	9.8	62	8.1	48	8.0
	Married	66	53.7	346	44.9	270	45.1
	Other	1	0.8	7	0.9	5	0.8
BMI category	< 24.9	17	12.8	68	8.8	51	8.6
	25–29.9	35	28.5	113	14.7	82	13.8
	30–34.9	31	25.2	160	20.8	131	22.0
	35–39.9	15	12.2	158	20.5	123	20.6
	> 39.9	25	20.3	271	35.2	209	35.1
Have you ever had bariatric surgery?	No	71	57.8	668	86.8	526	87.8
	Yes	52	42.3	102	13.2	73	12.2
Ever taken a GLP-1 medication to lose weight?	No	40	32.6	328	42.6	255	42.6
	Yes, I am currently taking	48	39.0	324	42.1	258	43.1
	Yes, I have taken in the past	35	28.5	118	15.3	86	14.4
Have you ever gone to a medically managed weight loss program?	No	—	—	524	68.1	412	68.8
	Yes, I am currently going to one	—	—	70	9.1	52	8.7
	Yes, I have gone in the past	—	—	176	22.9	135	22.5

(Figure 2). On the Rasch "ruler," the concept where participants reported most problems was self-esteem. No DIF was detected for gender. Three questions evidenced DIF, i.e., two questions for age group and two questions for BMI group. However, DIF had minimal impact on scoring, with Pearson correlations between the original and split analyses all ≥ 0.998. No evidence of local dependency was observed (all residual correlations were ≤ 0.20). Reliability was high with PSI values of 0.81 and 0.80 and Cronbach alpha values of 0.83 and 0.82 with and without extremes, respectively. The person-item threshold distribution (Figure 3) demonstrated good targeting; 759 of 770 (98.6%) participants (upper histogram) scored within the scale range (lower histogram), and floor and ceiling effects were below 1%.

Construct validity results are shown in Table 1. Overall, 22 of 28 hypotheses for the initial survey were accepted (79%). Detailed results for group difference are shown in Table S4. Higher scores were reported by participants who were further along in their weight loss journey, more satisfied with their weight (loss), social life, daily functioning, and body appearance, more active, and happier with their eating habits and who had better mental and physical health. Lower scores were observed among participants who wanted MBS, a GLP-1 receptor agonist, and body contouring and among those who reported weight gain or had health problems related to excess skin.

Detailed results for convergent validity are shown in Table S5. As expected, moderate negative correlations were observed

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

between the BODY-Q-7D score and lower BMI ($r = -0.3$; $p < 0.001$) and lower amount of weight someone wanted to lose to reach his or her goal weight ($r = -0.3$; $p < 0.001$). For convergent validity, as expected, most BODY-Q scales correlated 0.3–0.5 with the BODY-Q-7D scale, indicating the new scale measures a related but dissimilar construct. The Psychological scale ($r = 0.6$) correlated ≥ 0.5, indicating the BODY-Q-7D measured a similar construct as hypothesized.

TABLE 3 | Type of weight loss medication for the participants who reported current or past use.

	Content validity ($N = 83$)				Field test ($n = 442$)				Follow-up ($n = 344$)			
	Currently having		Had in past		Currently having		Had in past		Currently having		Had in past	
Type	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**
Ozempic (semaglutide)	20	24.1	26	31.3	89	20.1	97	21.9	70	20.3	74	21.5
Mounjaro (tirzepatide)	14	16.9	14	16.9	140	31.7	45	10.2	112	32.6	34	9.9
Wegovy (semaglutide)	8	9.6	12	14.5	43	9.7	57	12.9	36	10.5	38	11.0
Zepbound (tirzepatide)	7	8.4	10	12.0	37	8.4	7	1.6	26	7.6	5	1.5
Saxenda (liraglutide)	4	4.8	9	10.8	3	0.7	35	7.9	2	0.6	27	7.8
Trullicity (dulaglutide)	5	6.0	3	3.6	7	1.6	15	3.4	5	1.5	13	3.8
Rybelus	0	0.0	0	0.0	3	0.7	5	1.1	3	0.9	5	1.5
Retatrutide	0	0.0	0	0.0	1	0.2	0	0.0	1	0.3	0	0.0
Compounded semiglutide	1	1.2	0	0.0	6	1.4	5	1.1	6	1.7	5	1.5
Other obesity medication												
Phentermine (Adipex-P)	6	7.2	14	16.9	—	—	—	—	—	—	—	—
Contrave (bupropion-naltrexone)	3	3.6	11	13.3	—	—	—	—	—	—	—	—
Orlistat (Xenical, Alli)	3	3.6	17	20.5	—	—	—	—	—	—	—	—
Setmelanotide (Imcivree)	3	3.6	5	6.0	—	—	—	—	—	—	—	—
Qsymia (phentermine/topiramate)	1	1.2	7	8.4	—	—	—	—	—	—	—	—

TABLE 4 | Response option and item results from the 123 content validity participants.

		Response options				Items			
		Easy to understand		Fits experience		Easy to understand		Relevant	
PRO	**Item**	***N***	**%**	***N***	**%**	***N***	**%**	***N***	**%**
Physical	1. Walk or move	123	100	122	99.2	123	100	116	94.3
Physical	2. Moderate exercise	123	100	122	99.2	123	100	122	99.2
Body image	3. Self-conscious	123	100	123	100	123	100	123	100
Psychological	4. Emotional distress	123	100	122	99.2	120	97.6	119	96.7
Body image	5. Feel about body	120	97.6	121	98.4	121	98.4	121	98.4
Self esteem	6. Feel about self	122	99.2	123	100	122	99.2	121	98.4
Stigma	7. Feel accepted	122	99.2	121	98.4	122	99.2	117	95.1
Social	8. Usual activities	123	100	123	100	123	100	121	98.4
Eating	9. Difficult control	123	100	123	100	123	100	123	100
Eating	10. Often control	123	100	123	100	123	100	122	99.2
Excess skin	11. Unpleasant symptoms	123	100	121	98.4	123	100	112	91.1

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The Physical Function ($r = 0.5$) and Social ($r = 0.5$) scales correlated higher than expected (0.3–0.5), indicating that the content of the BODY-Q-7D was more similar than hypothesized. The BODY-Q-7D correlated ≥0.5 with the summary score for the EQ-5D-5L ($r = 0.5$), indicating it measures a similar construct ($p < 0.001$). Finally, correlations with EQ-5D-5L subscales were 0.3–0.5, indicating that they measure a related but dissimilar construct ($p < 0.001$).

3.3 | TRT

Of 365 participants invited, 303 completed the TRT survey. We excluded 27 participants who reported change, 2 who provided an invalid Prolific ID, and 5 who completed the scale outside of 7–14 days. ICC values with and without outliers for the 269 participants were 0.72 ($n = 269$) and 0.81 ($n = 258$), respectively. The SDC was 20.3 and 17.8 at the individual level and 1.2 and 1.1 at the group level with and without outliers, respectively.

TABLE 5 | Category frequencies for the 11 tested items tested in the field test survey.

		Response categories			
PRO	**Item**	**0**	**1**	**2**	**3**
Physical	1. Walk or move	36	101	216	417
Physical	2. Moderate exercise	79	133	262	296
Body image	3. Self-conscious	314	265	135	56
Psychological	4. Emotional distress	94	253	266	157
Body image	5. Feel about body	406	277	76	11
Self esteem	6. Feel about self	214	371	152	33
Stigma	7. Feel accepted	111	308	241	110
Social	8. Usual activities	67	163	235	305
Eating	9. Difficult control	174	191	258	147
Eating	10. Often control	115	336	200	119
Excess skin	11. Unpleasant symptoms	17	94	180	479

Note: Response options varied according to the BODY-Q-7D item: 0 = extremely/severe/never; 1 = moderately/moderate/sometimes; 2 = little/mild/often; 3 = not/no/always.

3.4 | Responsiveness

Responsiveness results are shown in Table 1 with detailed results in Tables S6 and S7. A total of 12 of 16 hypotheses (75%) were accepted. Change scores were positive for those who reported improved weight-related HRQL (mean = 14.7, SD = 14.2; $p < 0.001$). Participants who reported no change had a mean change score that was significantly different from zero ($p < 0.001$), with lower confidence bounds slightly above zero (LB = 1.9). Due to kurtosis of three, small sample size ($n = 49$), and non- normal visual plot, nonparametric statistics were applied to examine the group that reported worsening HRQL. A Wilcoxon signed rank test showed that follow-up scores (median = 39) were lower than baseline scores (median = 42), representing a negative change score ($W = -2.5$, $p = 0.013$). As expected, greater improvement ($p < 0.001$) was observed for participants who lost weight, were satisfied with weight change, and experienced a larger BMI reduction. There was a negative correlation between the change score and change in BMI, which was weaker than expected ($r = -0.2$; $p < 0.001$). The composite Likert change score for weight-related HRQL had a moderate positive Spearman correlation with the BODY-Q-7D change score ($r = 0.3$; $p < 0.001$).

For convergent validity the correlations with the BODY-Q-7D change score and change scores for other BODY-Q scales were as expected ($r \geq 0.3$, $p < 0.001$) with the exception of Physical Function, which was weaker than expected ($r = 0.2$, $p < 0.001$). BODY-Q Social, Sexual, and Eating Behavior change scores, as predicted, had weaker correlations with the BODY-Q-7D change score (≤0.001). The correlation for the BODY-Q-7D and overall EQ-5D-5L change scores was lower than predicted ($r = 0.2$, $p < 0.001$).

4 | Discussion

Building on prior BODY-Q research [16–18] and the S.Q.O.T. initiative to standardize outcomes in multidisciplinary obesity care [12–14], we developed the BODY-Q-7D as a concise, rigorously constructed measure designed to support a future PBM. The instrument captures seven key PRO domains identified as important to people living with obesity and healthcare providers, ensuring that the resulting health states reflect outcomes that matter. Content validity was established in an international sample of 123 individuals living with obesity—more than twice the COSMIN-recommended minimum for survey-based

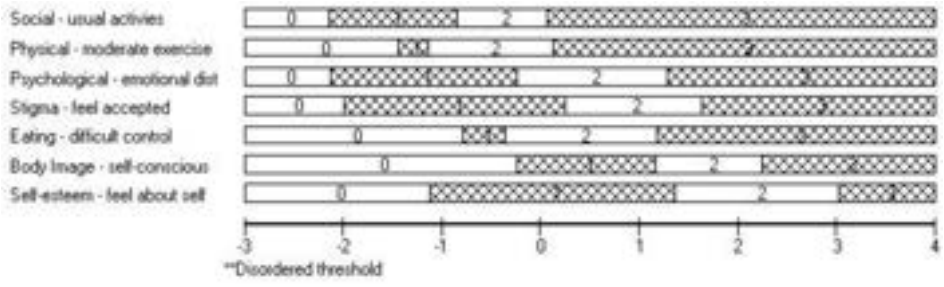

FIGURE 2 | Threshold map.

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

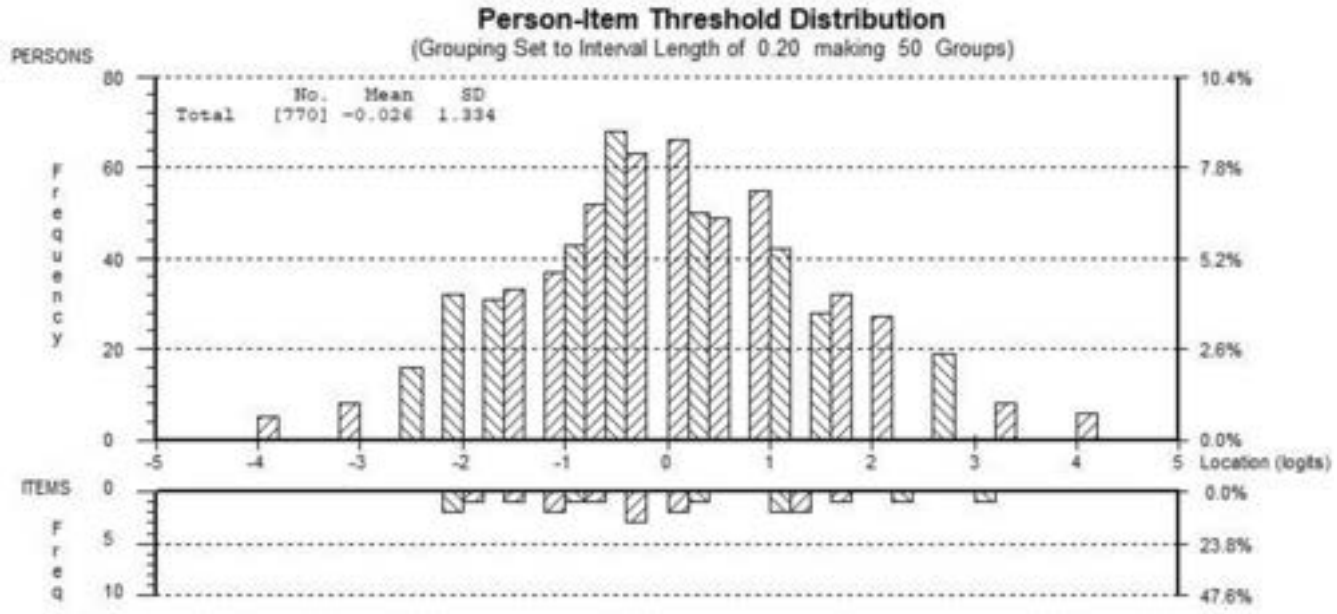

FIGURE 3 | Person-item threshold distribution.

content validity studies [20, 22]. Field test data further demonstrated reliability, construct validity, structural integrity consistent with RMT, and evidence of responsiveness. This staged development approach, consistent with other condition-specific PBM programs [33–36], prioritizes the establishment of a psychometrically robust descriptive system before undertaking valuation. By doing so, it ensures that the defined health states are measurement-sound and suitable for subsequent utility modeling and QALY estimation.

A question that we tested but found that it did not perform adequately within the BODY-Q-7D scale was dropped (i.e., "Did you experience any unpleasant symptoms due to excess skin caused by weight loss [e.g., rash, infections, odor]?"). This question evidenced poor fit to the Rasch model and had a high ceiling effect. Although it was an important concept according to the S.Q.O.T. initiative, having excess skin and experiencing skin-related symptoms are not universally applicable to people who lose weight. Without a specific question about skin, the 7-item scale should indirectly measure concerns for people with excess skin. Participants in the study who wanted body contouring to remove excess skin reported lower scores on the BODY-Q-7D than those who do not want body contouring. Furthermore, BODY-Q-7D scores were lower for participants who reported their excess skin problem to be somewhat or very bad compared with those who reported their skin problem as a little bad or not at all. In addition, research has shown that scores on the BODY-Q Excess Skin scale correlate moderately or strongly with other BODY-Q scales, including Body Image ($r=0.58$), Body Appearance ($r=0.50$), Sexual ($r=0.48$), Psychological ($r=0.47$), and Social ($r=0.40$) [17].

An important finding is that self-esteem was the concept with the highest item location (difficulty) in the RMT analysis. This finding aligns with the S.Q.O.T. initiative finding that self-esteem was the most important PRO for people living with obesity [12]. We also found that the 3 items that ranked lowest in the Rasch hierarchy for the BODY-Q-7D scale (i.e., social, physical, psychological) are concepts included in generic PBMs, such as the EQ-5D-5L [24] and the SF-6D [37]. These findings reinforce the need for an obesity-specific PBM that asks about important concepts to patients, including self-esteem, stigma, eating concerns, and body image.

The next step in the development of the BODY-Q-7D will be to design and implement a valuation study to create preference-based scoring for its health states. Evidence is required to demonstrate that the BODY-Q-7D performs coherently in applied economic evaluations. This research will include the collection of evidence of discriminatory power across clinically meaningful groups, sensitivity to change at the utility level, and plausible behavior of utility decrements across the full severity (e.g., mild, moderate, severe) spectrum. Together, these properties are essential for ensuring that the BODY-Q-7D can support valid estimation of QALYs and inform cost-effectiveness analyses. Given the time and cost associated with valuation studies [38], it was important that we conduct the groundwork to develop a psychometrically robust scale using a modern psychometric method before undertaking valuation. The BODY-Q-7D was shown to perform well across COSMIN criteria [20–22], as it demonstrated content validity, internal consistency, reliability, construct validity, and responsiveness.

Our study has some limitations. We used an online platform to recruit a sample of people living with obesity. We double screened to exclude anyone whose answers changed for data quality purposes, but clinical and demographic data could not be independently verified. While our sample is heterogeneous, the sample was limited to English speaking countries and included more people who identify as White, female, and in obesity Class 3. We had over 100 people who did not respond to our invitation to complete the 3-month

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

follow-up survey, and the reason for nonresponse could not be determined.

5 | Conclusion

The BODY-Q-7D demonstrated strong psychometric properties in a large heterogeneous international sample of people living with obesity. This short, rigorously designed scale functions as a PROM and was structured to allow for preference-based scoring. The next step for this research is to collect preference-based weights for the health states using an established approach. Further information about the BODY-Q PBM can be accessed as https://qportfolio.org/body-q/.

Funding

This study was funded by internal research funding provided to Anne Klassen through the Klassen's Department at McMaster University.

Disclosure

The BODY-Q was authored by Drs. Anne Klassen, Andrea Pusic, and Stefan Cano. The copyright to the work is owned by Memorial Sloan-Kettering Cancer Center, Memorial Hospital for Cancer and Allied Diseases, Sloan-Kettering Institute for Cancer Research (New York City, New York, USA), McMaster University (Hamilton, Ontario, Canada), Brigham and Women's Hospital (Boston, Massachusetts, USA), and Stefan Cano. The developers receive a share of license revenues as royalties for its use in for-profit research based on their institution's inventor sharing policy.

Conflicts of Interest

Stefan Cano reported being president of Modus Outcomes outside the submitted work. Anne Klassen reported research consulting through EVENTUM Research outside the submitted work. The other authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. World Obesity, "World Obesity Atlas 2023," published March 2023, https://www.worldobesityday.org/assets/downloads/World_Obesity_Atlas_2023_Report.pdf.

2. T. Nygaard Flølo, H. Liu, J. R. Andersen, and R. Kolotkin, "Exploring the Relationship Between Obesity, Weight Loss and Health-Related Quality of Life: An Updated Systematic Review of Reviews," *Clinical Obesity* 16 (2025): e70049.

3. J. L. Colquitt, K. Pickett, E. Loveman, and G. K. Frampton, "Surgery for Weight Loss in Adults," *Cochrane Database of Systematic Reviews* 8 (2014): CD003641.

4. L. Peirson, J. Douketis, D. Ciliska, D. Fitzpatrick-Lewis, M. U. Ali, and P. Raina, "Treatment for Overweight and Obesity in Adult Populations: A Systematic Review and Meta-Analysis," *Journal of Evidence-Based Medicine* 2, no. 4 (2014): E306–E317.

5. A. M. Jastreboff, L. J. Aronne, N. N. Ahmad, et al., "Tirzepatide Once Weekly for the Treatment of Obesity," *New England Journal of Medicine* 387, no. 3 (2022): 205–216.

6. J. P. H. Wilding, R. L. Batterham, S. Calanna, et al., "Once-Weekly Semaglutide in Adults With Overweight or Obesity," *New England Journal of Medicine* 384 (2021): 989–1002.

7. Q. Xia, J. A. Campbell, H. Ahmad, et al., "Health State Utilities for Economic Evaluation of Bariatric Surgery: A Comprehensive Systematic Review and Meta-Analysis," *Obesity Reviews* 21, no. 8 (2020): e13028.

8. S. Sauerland, S. Weiner, K. Dolezalova, et al., "Mapping Utility Scores From a Disease-Specific Quality-Of-Life Measure in Bariatric Surgery Patients," *Value in Health* 12, no. 2 (2009): 364–370.

9. W. Guo, S. Xie, D. Wang, and J. Wu, "Mapping the IWQOL-Lite Onto EQ-5D-5L and SF-6Dv2 Among Overweight and Obese Adults," *Quality of Life Research* 33 (2024): 737–749, https://doi.org/10.1007/s11136-023-03598-9.

10. J. E. Brazier, R. L. Kolotkin, R. D. Crosby, and G. R. Williams, "Estimating a Preference-Based Single Index for the Impact of Weight on Quality of Life-Lite (IWQOL-Lite) Instrument From the SF-6D," *Value in Health* 7, no. 4 (2004): 490–498.

11. S. Sun, E. Stenberg, Y. Cao, et al., "Mapping the Obesity Problems Scale to the SF-6D: Results Based on the Scandinavian Obesity Surgery Registry (SOReg)," *European Journal of Health Economics* 24, no. 2 (2023): 279–292.

12. C. E. E. de Vries, C. B. Terwee, M. Al Nawas, 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," *Obesity Reviews* 23, no. 8 (2022): e13452.

13. P. J. Dijkhorst, C. E. E. de Vries, C. B. Terwee, et al., "A Core Set of Patient-Reported Outcome Measures to Measure Quality of Life in Obesity Treatment Research," *Obesity Reviews* 26, no. 2 (2025): e13849.

14. P. J. Dijkhorst, V. M. Monpellier, C. B. Terwee, et al., "Core Set of Patient-Reported Outcome Measures for Measuring Quality of Life in Clinical Obesity Care," *Obesity Surgery* 34, no. 8 (2024): 2980–2990.

15. C. E. E. de Vries, M. C. Kalff, C. A. C. Prinsen, et al., "Recommendations on the Most Suitable Quality-Of-Life Measurement Instruments for Bariatric and Body Contouring Surgery: A Systematic Review," *Obesity Reviews* 19, no. 10 (2018): 1395–1411.

16. A. F. Klassen, S. J. Cano, A. Scott, E. Tsangaris, and A. L. Pusic, "Assessing Outcomes in Body Contouring," *Clinics in Plastic Surgery* 41, no. 4 (2014): 645–654, https://doi.org/10.1016/j.cps.2014.06.004.

17. A. F. Klassen, S. J. Cano, A. Alderman, et al., "The BODY-Q: A Patient-Reported Outcome Instrument for Weight Loss and Body Contouring Treatments," *Plastic and Reconstructive Surgery. Global Open* 4, no. 4 (2016): e679.

18. C. E. E. De Vries, D. Mou, L. Poulsen, 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," *Obesity Surgery* 31, no. 8 (2021): 3637–3645.

19. US Food and Drug Administration, *Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*, (US Department of Health and Human Services, 2009), https://www.fda.gov/media/77832/download.

20. C. B. Terwee, C. Prinsen, A. Chiarotto, et al., "COSMIN Methodology for Assessing the Content Validity of PROMs—User Manual," published February 2018, https://www.cosmin.nl/wp-content/uploads/COSMIN-methodology-for-content-validity-user-manual-v1.pdf.

21. J. J. Gagnier, J. Lai, L. B. Mokkink, and C. B. Terwee, "COSMIN Reporting Guideline for Studies on Measurement Properties of Patient-Reported Outcome Measures," *Quality of Life Research* 30, no. 8 (2021): 2197–2218.

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

22. L. B. Mokkink, C. Prinsen, D. L. Patrick, et al., "COSMIN Study Design Checklist for Patient-Reported Outcome Measurement Instruments," published July 2019, https://gut.bmj.com/content/gutjnl/70/1/139/DC1/embed/inline-supplementary-material-1.pdf.

23. B. B. Reeve, K. W. Wyrwich, A. W. Wu, et al., "ISOQOL Recommends Minimum Standards for Patient-Reported Outcome Measures Used in Patient-Centered Outcomes and Comparative Effectiveness Research," *Quality of Life Research* 22, no. 8 (2013): 1889–1905.

24. M. Herdman, C. Gudex, A. Lloyd, et al., "Development and Preliminary Testing of the New Five-Level Version of EQ-5D (EQ-5D-5L)," *Quality of Life Research* 20, no. 10 (2011): 1727–1736.

25. T. Devji, A. Carrasco-Labra, A. Qasim, et al., "Evaluating the Credibility of Anchor Based Estimates of Minimal Important Differences for Patient Reported Outcomes: Instrument Development and Reliability Study," *BMJ* 369 (2020): m1714.

26. G. Rasch, "A Mathematical Theory of Objectivity and Its Consequences for Model Construction," *European Meeting on Statistics, Econometrics and Management Sciences* (Amsterdam: 1968), https://rasch.org/memo1968.pdf.

27. D. Andrich, *Rasch Models for Measurement* (Sage Publications, 1988).

28. J. Hobart and S. Cano, "Improving the Evaluation of Therapeutic Interventions in Multiple Sclerosis: The Role of New Psychometric Methods," *Health Technology Assessment* 13 (2009): 1–177, https://doi.org/10.3310/hta13120.

29. S. J. Cano and J. C. Hobart, "The Problem With Health Measurement," *Patient Preference and Adherence* 5 (2011): 279–290.

30. A. Geerinck, V. Alekna, C. Beaudart, et al., "Standard Error of Measurement and Smallest Detectable Change of the Sarcopenia Quality of Life (SarQoL) Questionnaire: An Analysis of Subjects From 9 Validation Studies," *PLoS One* 14, no. 4 (2019): e0216065.

31. E. B. Smit, H. Bouwstra, L. D. Roorda, et al., "A Patient-Reported Outcomes Measurement Information System Short Form for Measuring Physical Function During Geriatric Rehabilitation: Test-Retest Reliability, Construct Validity, Responsiveness, and Interpretability," *Journal of the American Medical Directors Association* 22, no. 8 (2021): 1627–1632.

32. H. Y. Kim, "Statistical Notes for Clinical Researchers: Assessing Normal Distribution (2) Using Skewness and Kurtosis," *Restorative Dentistry and Endodontics* 38, no. 1 (2013): 52–54.

33. T. A. Young, D. Rowen, J. Norquist, and J. E. Brazier, "Developing Preference-Based Health Measures: Using Rasch Analysis to Generate Health State Values," *Quality of Life Research* 19, no. 6 (2010): 907–917.

34. I. Mavranezouli, J. E. Brazier, T. A. Young, and M. Barkham, "Using Rasch Analysis to Form Plausible Health States Amenable to Valuation: The Development of CORE-6D From a Measure of Common Mental Health Problems (CORE-OM)," *Quality of Life Research* 20, no. 3 (2011): 321–333.

35. M. Sundaram, M. J. Smith, D. Revicki, B. Elswick, and L. A. Miller, "Rasch Analysis Informed the Development of a Classification System for a Diabetes-Specific Preference-Based Measure of Health," *Journal of Clinical Epidemiology* 62, no. 8 (2009): 845–856.

36. A. D. Keetharuth, D. Rowen, J. B. Bjorner, and J. E. Brazier, "Estimating a Preference-Based Index for Mental Health From the Recovering Quality of Life Measure," *Value in Health* 24, no. 8 (2021): 1160–1172.

37. J. E. Brazier, T. Usherwood, R. Harper, and K. Thomas, "Deriving a Preference-Based Single Index From the UK SF-36 Health Survey," *Journal of Clinical Epidemiology* 51, no. 11 (1998): 1115–1128.

38. J. E. Brazier, D. Rowen, I. Mavranezouli, et al., "Developing and Testing Methods for Deriving Preference-Based Measures of Health From Condition-Specific Measures (And Other Patient-Based Measures of Outcome)," *Health Technology Assessment* 16, no. 32 (2012): 1–114, https://doi.org/10.3310/hta16320.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Table S1:** Questions used to determine content validity. **Table S2:** Psychometric tests performed. **Table S3:** Rasch Measurement Theory (RMT) analysis item fit statistics and differential item functioning (DIF). **Table S4:** Construct validity results. **Table S5:** Convergent validity results. **Table S6:** Construct validity for change scores. **Table S7:** Convergent validity results for change scores.

1930739x, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/oby.70254, Wiley Online Library on [26/07/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table S1: Questions used to determine content validity

PREAMBLE	We would like your help to make sure the BODY-Q questionnaire is EASY TO UNDERSTAND and asks questions that are IMPORTANT to people having weight loss treatment. Below is a snapshot of the BODY-Q questionnaire that has 11 questions (*You do not need to read this NOW, it is just to give you an idea of the questionnaire*). We will ask you to answer each question. After each question, we will ask if it was easy to understand and important. Please answer each question to the BEST OF YOUR ABILITY.
SCALE INSTRUCTIONS - COMPREHENSION	THE BODY-Q QUESTIONNAIRE has these instructions in the white box below:
	These questions ask about the impact of your WEIGHT on your HEALTH. Please answer each question based on the PAST WEEK.
	Are the instructions easy to understand?
	1, YES 2, NO
	Please explain what was HARD to understand about the instructions.
ITEMS - ANSWER	Below is the first of the 11 BODY-Q questions. Please read the instructions and then answer question #1.
	INSTRUCTIONS: This question asks about the impact of your WEIGHT on your HEALTH. Please answer this question based on the PAST WEEK.
	1. How difficult was it for you to walk or move around?
	1, It was NOT difficult. 2, It was a LITTLE difficult. 3, It was MODERATELY difficult. 4, It was EXTREMELY difficult.
ITEM WORDING - COMPREHENSION	Is question #1 EASY to understand?
	1, YES 2, NO
	Please explain what was HARD to understand about question #1.
ITEM RESPONSE OPTION – COMPREHENSION	Are the 4 answers options for question #1 EASY to understand?
	1, YES 2, NO
	Please explain what was HARD to understand about the 4 response options for question #1.
ITEM RESPONSE OPTION - APPROPRIATE	Question #1 gave you 4 options to report how difficult it is for you to walk or move around. Did one of these 4 options describe your experience of walking and moving around?
	1, YES 2, NO
	What option would better describe your experience of walking and moving around?
ITEM - RELEVANT	Thinking about the impact of your weight on your health, a relevant question is one that feels meaningful and asks about something that matters to you. Is question #1 shown above RELEVANT to your experience?
	1, YES - this question is relevant to me NOW or in the PAST. 2, NO - this question has NEVER been relevant to me.
	Please tell us why question #1 is not relevant to your experience.
SCALE - TIME FRAME	All 11 questions asked about your experience in the PAST WEEK. Were you able to think back over the PAST WEEK and remember your experiences to answer the questions?
	1, YES 2, NO
	We asked you to answer each question based on the PAST WEEK. Is the PAST WEEK an appropriate length of time for people to answer the 11 BODY-Q questions?
	1, YES - the PAST WEEK is ENOUGH time (ie, just right) 2, NO - the PAST WEEK is NOT enough time (ie, too short) 3, NO - the PAST WEEK is TOO MUCH time (ie, too long)

	What time frame do you think would be the most appropriate to describe your experience?
	1, Today 2, 2 days 3, 3 days 4, 4 days 5, 5 days 6, 6 days 7, 2 weeks 8, 3 weeks 9, 1 month 10, 2 months 11, 3 months 12, 4 months 13, 5 months 14, 6 months 15, 7 months 16, 8 months 17, 9 months 19, 10 months 20, 11 months 21, 1 year 22, more than 1 year
	Tell us why you answered the PAST WEEK was not the right amount of time.
SCALE - COMPREHENSIVE	Thinking about the impact of your weight on your health, a relevant question is one that feels meaningful and asks about something that matters to you. We asked 11 questions to measure the IMPACT of your WEIGHT on your HEALTH. Is there anything RELEVANT to your experience that is missing?
	1, YES 2, NO
	Please tell us what you think is missing from the BODY-Q Questionnaire
COMMENTS	Do you have any other feedback on the BODY-Q questionnaire?

Table S2: Psychometric tests performed

Test	Description
Response threshold order	To determine if the response categories were properly ordered [28]. We examined thresholds between response options to examine if the response categories for each item are ordered, meaning that a '1' sits lower in the continuum than a '2' and so on.
Item fit	To determine if the data fit the Rasch model [28]. We examined item fit statistics with 3 fit indicators: Chi-square after Bonferroni adjustment, fit residuals and item characteristic curves [28]. The overall fit of the data to the Rasch model will also be determined.
Local dependency	To determine whether items were structurally independent, we examined correlations among item residuals after fitting the Rasch model [2]8]. Residual correlations >0.20 were used to identify potential local dependency between items.
Targeting	To determine how well the scale measures the experience of the sample [28]. We inspected the person-item threshold plots to identify clustering of items and gaps on the scale, computed the proportion of the sample that scores on the scale's range of measurement, and computed the proportion of the sample to score at the floor and ceiling on the scale.
Differential item functioning (DIF)	To determine if items in the scale behaved the same across different characteristics [28]. We computed DIF for age (18-34, 35-44, 45-54, $\geq$55), gender (male, female), and BMI (below class 1, class 1, class 2, class 3). Random samples were drawn to ensure subgroups were the same size. DIF was repeated 3 times for each characteristic to determine if the findings were stable. Any items with DIF were split on the characteristic, and person locations from the original and split analyses were correlated to determine impact of DIF on scoring.
Reliability	We computed person separation index (PSI) [28] and Cronbach alpha values with and without extremes included. For test-retest reliability, intraclass correlation coefficients (ICC) with a two-way random effects model were completed after excluding anyone who reported their health-related quality of life had changed between assessments.
Hypothesis-based construct validity	To determine how well the scale accurately measures weight-related health-related quality of life. We tested 28 a priori hypotheses to ensure unbiased conclusions were drawn. We conducted expected group differences and convergent validity. For the change scores (Time 2 – Time 1), we tested 16 a priori hypotheses. Acceptance of at least 75% of hypotheses was considered sufficient evidence of validity.

Table S3: Rasch Measurement Theory (RMT) analysis item fit statistics and differential item functioning (DIF)

PRO	Item	Item fit							DIF		
		Location	SE	FitResid	DF	ChiSq	DF	Prob	Age	Gender	BMI
									N=568	N=446	N=632
Social	Usual activity	-0.97	0.05	-0.39	648	7.06	8	0.53	no	no	no
Physical	Moderate exercise	-0.81	0.05	4.77	648	13.92	8	0.08	yes	no	yes
Psychological	Emotional distress	-0.36	0.05	-2.85	648	18.30	8	0.02	no	no	no
Stigma	Feel accepted	-0.03	0.05	0.53	648	4.32	8	0.83	yes	no	no
Eating	Difficult control	0.02	0.05	4.55	648	16.16	8	0.04	no	no	yes
Body Image	Self-conscious	1.06	0.05	-1.11	648	8.09	8	0.43	no	no	no
Self-esteem	Feel about self	1.09	0.06	-1.66	648	13.45	8	0.10	no	no	no

Definitions: FitResid = fit residual, ChiSq = Chi square; DF = degrees of freedom; Prob = probability; DIF = differential item function

Table S4: Construct validity results

	N	Mean	SD	SE	95% CI LB	95% CI UB	p-value
1. Scores incrementally higher as report being closer to middle or end of weight loss journey							
Start	403	47	14	1	45.5	48.4	<0.001
Middle	251	50	15	1	48.6	52.3	
End	116	57	20	2	53.5	60.9	
2. Scores incrementally higher as report being more satisfied with weight							
Extremely dissatisfied	294	42	14	1	40.5	43.8	<0.001
Very dissatisfied	223	49	12	1	47.9	51.1	
Somewhat dissatisfied	161	57	14	1	55.2	59.6	
Somewhat satisfied	60	58	19	2	53.3	63.0	
Extremely/Very satisfied	32	65	18	3	58.1	71.1	
3. Scores incrementally higher as report having better physical health							
Excellent/Very good	126	61	17	2	57.9	63.8	<0.001
Good	276	53	14	1	50.8	54.2	
Fair	256	46	13	1	43.9	47.1	
Poor	109	39	15	1	36.4	42.2	
4. Scores incrementally higher as report having better mental health							
Excellent	87	65	17	2	61.0	68.2	<0.001
Very good	141	56	13	1	53.5	58.0	
Good	228	51	14	1	49.4	53.1	
Fair	225	44	13	1	42.4	45.7	
Poor	87	35	14	1	32.5	38.3	
5. Scores incrementally higher as report being happier with social life							
Extremely unhappy	71	38	17	2	34.1	42.0	<0.001
Very unhappy	87	41	15	2	38.2	44.5	
Somewhat unhappy	162	48	12	1	45.6	49.4	
Somewhat happy	266	51	14	1	49.0	52.4	
Very happy	139	57	15	1	54.4	59.6	
Extremely happy	40	65	17	3	60.1	70.7	
6. Scores incrementally higher as report being happier with their eating habits							
Extremely unhappy	65	39	14	2	35.3	42.3	<0.001
Very unhappy	114	43	13	1	41.0	45.7	
Somewhat unhappy	165	50	14	1	47.7	51.9	
Somewhat happy	284	52	14	1	50.4	53.8	
Extremely/Very happy	141	55	20	2	51.5	58.1	
7. Scores incrementally higher as report being happier with how body looks							
Extremely unhappy	197	37	12	1	35.7	39.2	<0.001
Very unhappy	187	49	12	1	47.7	51.1	
Somewhat unhappy	166	55	12	1	53.1	56.8	
Somewhat happy	155	54	15	1	51.3	56.1	
Extremely/Very happy	65	64	22	3	58.5	69.3	
8. Scores incrementally higher as report being more active							

Not at all active	188	43	15	1	41.2	45.6	<0.001
A little active	312	49	14	1	47.6	50.7	
Somewhat active	217	54	16	1	52.2	56.5	
Very active	53	55	20	3	50.0	60.9	
9. Scores incrementally higher as report less interference with ability to do daily activities							
Never	323	58	16	1	56.6	60.1	<0.001
Rarely	196	49	11	1	47.2	50.1	
Sometimes	152	42	12	1	39.8	43.7	
Often	63	37	12	1	33.9	39.7	
Always	33	32	17	3	25.9	37.8	
10. Scores higher if do not want body contouring to remove excess skin							
I don't want this	503	52	16	1	50.2	53.0	<0.001
I want this	253	46	15	1	44.2	48.0	
11. Scores higher as report more problems due to excess skin							
No skin problems	384	52	15	1	50.2	53.3	<0.001
A little bad	149	44	16	1	41.4	46.4	
Somewhat/Very bad	33	40	13	2	35.0	44.3	
1. Subgroup: want to have or had bariatric surgery. Scores higher if had bariatric surgery vs want to have							
No, I have not had bariatric surgery but I want this	76	42	14	2	38.7	44.9	<0.001
Yes, I have had bariatric surgery	102	50	17	2	46.1	52.9	
2. Subgroup: want to have or taking a GLP-1 agonist. Scores higher if currently taking a GPL-1 treatment							
No, I have never taken GLP-1 medication to lose weight but I want to	198	45	14	1	43.5	47.4	<0.001
Yes, I am currently taking GLP-1 medication to lose weight	324	51	17	1	49.2	52.8	

Table S5: Convergent validity results

	r	n	p-value
14. Negative correlation with BMI (0.3-0.5)	-0.3	770	<0.001
15. Negative correlation with lower amount of excess skin on the body (0.3-0.5)*	-0.2	769	<0.001
16. Negative correlation with lower amount of weight wanting to lose (0.3-0.5)	-0.3	769	<0.001
17. Positive correlation with BODY-Q Body (0.3-0.5)	0.5	770	<0.001
18. Positive correlation with BODY-Q Body Image (0.3-0.5)	0.5	769	<0.002
19. Positive correlation with BODY-Q Physical Function (0.3-0.5)	0.5	765	<0.003
20. Positive correlation with BODY-Q Psychological (≥0.5)	0.6	767	<0.004
21. Positive correlation with BODY-Q Social (0.3-0.5)	0.5	764	<0.005
22. Positive correlation with BODY-Q Eating Behavior (0.3-0.5)	0.4	769	<0.006
23. Positive correlation with BODY-Q Sexual (0.3-0.5)	0.4	470	<0.007
24. Negative correlation with EQ-5D-5L Self Care (0.3-0.5)	-0.3	764	<0.001
25. Negative correlation with EQ-5D-5L Mobility (0.3-0.5)	-0.4	764	<0.001
26. Negative correlation with EQ-5D-5L Usual Activity (0.3-0.5)	-0.4	764	<0.001
27. Negative correlation with EQ-5D-5L Anxiety & Depression (0.3-0.5)	-0.5	764	<0.001
28. Positive correlation with EQ-5D-5L (≥0.5)	0.5	764	<0.008

*Spearman correlation

Table S6: Construct validity for change scores

	n	Mean	SD	Mean change	SD change	LB	UB	p-value
1. Change score will be incrementally higher for those who lost weight vs. gained weight or stayed the same								
Gained weight	159	-	-	2.6	11.8	0.7	4.4	<0.001
Stayed the same	106	-	-	1.4	12.6	-1.1	3.8	
Lost weight	325	-	-	7.9	13.2	6.5	9.4	
2. Change scores will be higher for those satisfied with change in weight								
Dissatisfied	359	-	-	2.4	11.7	1.2	3.7	<0.001
Satisfied	240	-	-	9.5	13.8	7.8	11.3	
3. Change score will be higher for those who lost 5-10% vs. 0-4.99% of their body weight								
0-4.99%	263	-	-	4.1	12.1	2.7	5.6	<0.001
5%-10%	114	-	-	9.9	11.9	7.7	12.1	
4. Those who report health is worse will have negative change score (non-normal)								
PBM time 2	49	39 (Med)	-	-	-	-	-	W=-2.5* 0.013
PBM time 1	49	42 (Med)	-					
5. Those who report health is the same will have no change in score								
PBM time 2	400	54	15.8	2.9	10.5	1.9	4.0	<0.001
PBM time 1	400	52	15.8					
6. Those who report health is better will have positive change score								
PBM time 2	150	63	14.2	14.7	14.2	12.4	17.0	<0.001
PBM time 1	150	48	16.1					

*Wilcoxon signed-rank test

Table S7: Convergent validity results for change scores

	r	n	p-value
7. Change score will correlate negatively with change in BMI	-0.2	596	<0.001
8. Change score will correlate positively with score for composite HRQL variable*	0.3	594	<0.001
9. Change score will correlate positively with BODY-Q Body (≥0.3)	0.3	599	<0.001
10. Change score will correlate positively with BODY-Q Body Image (≥0.3)	0.3	598	<0.001
11. Change score will correlate positively with BODY-Q Physical Function (≥0.3)	0.2	594	<0.001
12. Change score will correlate positively with BODY-Q Psychological (≥0.3)	0.3	595	<0.001
13. Change score will correlate positively with BODY-Q Social (<0.3)	0.2	590	<0.001
14. Change score will correlate positively with BODY-Q Sexual (<0.3)	0.2	360	0.001
15. Change score will correlate positively with BODY-Q Eating Behavior (<0.3)	0.2	590	<0.001
16. Change score will correlate positively with BODY-Q with EQ-5D-5L (≥0.5)	0.2	590	<0.001

*A composite Likert change score was computed by assigning 0 for same score, -1 for worse score, +1 for improved score for the following HRQL questions: happy with how body looks, social life, eating habits, activity level, physical health, mental health, and interference of weight with activities