A preliminary report on the development of a validated tool for measuring psychosocial outcomes for massive weight loss patients

Nada Al-Hadithy a,*, Richard Welbourn b, Hosakere Aditya c, Ken Stewart c, Mark Soldin d

a Royal Devon and Exeter Hospital, Barrack Road, Exeter, EX2 5DW, UK
b Musgrove Park Hospital, Taunton, UK
c St. John’s Hospital, Livingston, UK
d St. George’s NHS Trust, London, UK

Received 7 November 2013; accepted 1 July 2014

KEYWORDS
Massive weight loss; Body contouring; Patient report outcome measure; PROM; Validation; Bariatric surgery

Summary  Aim: To validate the newly developed patient report outcome measure (PROM): the Post Bariatric Outcome Tool (PBOT). The tool is designed and developed for massive weight loss patients seeking body contouring procedures.
Method: The PBOT was piloted with three cohorts: massive weight loss patients seeking body contouring; massive weight loss patients who have had body contouring; and healthy, non-obese subjects as controls matched for age and gender. Each cohort completed two PROMS at week one, and then for a second time at week three. The PROMS used were the new Post Bariatric Outcome Tool (PBOT) and the Derriford Appearance Scale 24 (DAS24).
Conclusion: The PBOT was shown to be reliable both in terms of its internal consistency and test-retest reliability. Comparison to the DAS24 demonstrated the PBOT to be valid. However, the cohorts were small and responsiveness was not tested. This needs to be tested in further larger validation studies, ideally, with comparison to functional scales such as the SF-36 or other validated massive weight loss body contouring PROMs; such as the Body Q.

Introduction
Following bariatric surgery, morbidity and mortality decreases, however ptotic redundant skin folds do not contract with the volume loss resulting in intertriginous rash, hygiene issues and functional and psychological impairment. Identifying outcomes in these patients requires an understanding of the complex adjustment they are making to their new body

* Corresponding author.
E-mail address: nadaucl@yahoo.com (N.A-Hadithy).
habituation, the redundant skin, removal of the coping mechanism of food, identity and the functional and psychosocial fall out. Evidence-based health policy emphasizes the importance of using scientifically rigorous patient-based outcome measures to evaluate the impact of disease and treatment.

Ensuring valid, robust data is generated from patient reported outcome measures (PROMs) depends on an appropriate assessment tool, reflecting the population, disease and specific domains relevant to the cohort. Although PROMs have been used widely in chronic illness and cancer, they are still a relatively new concept in the field of surgery. The aim of PROMs is to assess the patient’s perspective of health, illness, and the effects of health care interventions in a reliable, valid, acceptable, and feasible way. Darzi’s “NHS Next Stage Review” indicates that PROMs will be increasingly used in the evaluation and policy making of healthcare technologies and services. The drive to improve quality of care has led to the recognition of the importance of patient perspective and consequently the development of robust PROMs. Currently there is no measure for the massive weight loss body contouring (MWLBC) patient that is psychometrically sound; derived from patient and user experiences; has face validity; and is easy to administer and score.

We have developed a patient report outcome measure for massive weight loss (MWL) patients wishing to undergo body contouring called the Post Bariatric Outcome Tool (PBOT) (Appendix A). This PROM has been designed to fit in with the national guidelines of massive weight loss body contouring published by BAPRAS in 2014.

Utilising this PROM as part of the referral pathway will help identify which patients meet the national criteria and will heighten awareness of psychological disturbance that may warrant early psychological intervention. We anticipate users of the PBOT will come from a range a professional backgrounds including GPs, bariatric surgeons, plastic and reconstructive surgeons, clinical health psychologists and specialist nurses, as well as academics. The PBOT is five pages long. The referrer completes pages 1–2. The patient completes pages 3–5.

The length of time taken to complete the PBOT varies, but is usually between 10 and 15 min for pages 3–5. The completed form (pages 1–5), along with a clinical photograph of the patient is then sent to the MWLBC MDT for analysis and scoring. Figure 1.

In order to measure psychological and functional adjustment to MWLBC it is recommended that the patient completes pages 3–5 of the PBOT for a second time at the final plastic surgery outpatient clinic.

To develop a conceptual model and generate items for the PBOT we followed an established method of: literature review; semi structured patient interviews and expert opinions. This has been described elsewhere and is beyond the remit of this paper. This paper highlights the outcomes of assessment of validity of the PBOT in a prospective study, as per the guidance developed by the Scientific Advisory Committee of the Medical Outcomes Trust.

Methods

Field test and psychometric analysis

The following 3 groups were posted and completed the PBOT and Derriford 24 (DAS24) at week one and week three.

- 10 non-obese, healthy population
- 10 patients following massive weight loss (MWL)
- 10 patients post massive weight loss and body contouring (MWLBC)

Psychometric analysis was then performed on results for conceptual and measurement model, acceptability, responsiveness, reliability and validity.

Conceptual and measurement model

“A PROM should have documentation defining and describing the concept(s) included and the intended population(s) for use.” The PROM is supported by appropriate documentation. Appendices B & C.

Administrative burden/acceptability

The burden of acceptability was assessed by completion percentage of the PBOT. We were willing to accept <10% frequency of missing data from completed scores. Response distributions were examined, focussing on maximum endorsement frequencies, i.e. highest proportion of respondents who endorsed a single category for an item (should be <80%). Reading ease should be assessed. The Flesch/Flesch–Kincaid readability tests are designed to indicate comprehension difficulty when reading a passage of contemporary academic English. There are two tests, the Flesch Reading Ease, and the Flesch–Kincaid Grade Level.

Responsiveness

This is the ability of a PROM to detect change over time or following intervention/surgery.

Reliability

Reliability is a measure of the extent to which a PROM is free from random error. For PROMs, the two most common types of reliability assessed are internal consistency and test-retest reliability. Internal consistency can be measured with Cronbach’s Coefficient Alpha. We judged $r > 0.70$ acceptable. Test-retest reliability is a measure of the reproducibility of the PROM to provide consistent scores over time in a stable population. Test-retest reliability was assessed by estimated Bland and Altman’s method for agreement of repeated scores, where >95% of the mean of the re-test against the difference of the re-test within 2 standard deviations of the bias was considered acceptable.

Validity

Construct validity is the extent to which scores on the PROM relate to other validated measures (for the PBOT we have compared it to the DAS24) in a manner that is consistent with theoretically derived hypotheses concerning the constructs that are being measured. It is calculated using Spearman rank correlation co-efficient for mean scores. Content validity was determined in our previous paper.
Validity hypothesis testing is the ability to measure expected differences between groups within patient population.23

**Results**

**Demographic Information**

30 subjects completed the DAS24 and the PBOT twice, 2 weeks apart. The groups were matched for sex. Table 1. Subjects were reviewed a mean of 27.85 months following bariatric surgery (range 12–60). There were equal numbers of laparoscopic gastric bypass procedures and laparoscopic gastric bands at 7 each. 6 patients had their bands converted to open procedures intra operatively Table 2.

Completed forms for each group were assessed using the mark scheme and aid (Appendices B & C). The scores were reflected in two parts: one for the referral tool component (pages 1–2), and the other for the PROM component (pages 3–5). The DAS24 was scored for comparison Table 3.

**Analysis**

Data was analysed with IBM SPSS V.19 statistical package. The three groups were matched for sex. There was no statistical difference in age with a coefficient of variation

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Relationship status</th>
<th>Sex</th>
<th>Mean age (range)</th>
<th>Mean BMI (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non obese</td>
<td>10</td>
<td>Single = 4</td>
<td>M = 4</td>
<td>47.9</td>
<td>22.59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Married = 5</td>
<td>F = 6</td>
<td>(31–68)</td>
<td>(17.75–26)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Widowed = 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MWL</td>
<td>10</td>
<td>Single = 2</td>
<td>M = 4</td>
<td>45.00</td>
<td>30.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Married = 5</td>
<td>F = 6</td>
<td>(31–67)</td>
<td>(22.55–41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Widowed = 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MWLBC</td>
<td>10</td>
<td>Single = 5</td>
<td>M = 4</td>
<td>48</td>
<td>29.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Married = 4</td>
<td>F = 6</td>
<td>(24–67)</td>
<td>(20.44–49.7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Widowed = 1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
of 28.94%, 22.62% and 26.67% respectively. ANOVA resulted in F value of 0.81, P value 0.45 and R square of 0.05.

**Administrative burden/acceptability**

The burden of acceptability was assessed by percentage completeness of the PBOT. The mean completion percentage was 93.82%, which met our acceptance criteria of frequency of missing data Table 4. The response distributions were examined in Table 4. In the non-obese non-disease group 12–13 items had a >80% maximum endorsement frequency of the 27 items. This represents an 11.1%–44.4% of the range of questions. In the post body contouring group alone, 3–4 items had a >80% maximum endorsement frequency, representing an 11.1%–14.8% range of questions.

The Flesch Reading Ease score for the PBOT was 62.3, indicating a reading age of approximately 13. A key can be seen in Table 5.

**Responsiveness**

This was not measured as the same group of patients did not complete the PBOT before and after surgery or an intervention.

**Reliability**

Internal consistency can be measured with Cronbach’s Coefficient Alpha.9 The data was collated in groups and time (week 1 and 3) Table 6. Test-retest reliability was assessed by estimated Bland and Altman’s method for agreement for

<table>
<thead>
<tr>
<th>Table 2 Surgical history of MWL and BC groups.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Massive weight loss</td>
</tr>
<tr>
<td>2 Lap gastric band</td>
</tr>
<tr>
<td>3 Gastric band</td>
</tr>
<tr>
<td>4 Lap gastric band</td>
</tr>
<tr>
<td>5 Lap gastric bypass</td>
</tr>
<tr>
<td>6 Lap gastric bypass</td>
</tr>
<tr>
<td>7 Lap gastric bypass</td>
</tr>
<tr>
<td>8 Lap gastric band</td>
</tr>
<tr>
<td>9 Lap gastric bypass</td>
</tr>
<tr>
<td>10 Gastric band</td>
</tr>
<tr>
<td>Massive weight loss &amp; body contouring</td>
</tr>
<tr>
<td>2 Lap gastric band</td>
</tr>
<tr>
<td>3 Lap gastric band</td>
</tr>
<tr>
<td>4 Gastric band</td>
</tr>
<tr>
<td>5 Lap gastric band</td>
</tr>
<tr>
<td>6 Gastric band</td>
</tr>
<tr>
<td>7 Gastric band</td>
</tr>
<tr>
<td>8 Gastric band</td>
</tr>
<tr>
<td>9 Lap bypass</td>
</tr>
<tr>
<td>10 lap gastric bypass</td>
</tr>
</tbody>
</table>
repeated scores, where >95% of the mean of the re-test against the difference of the re-test within 2 standard deviations of the bias was considered acceptable Table 7 and Figure 2.

Validity

Construct validity was calculated using Spearman’s Rank correlation coefficients for mean scores between DAS24 and the PROM component of the PBOT Table 8. As \( n = 10 \) for weeks 1 and 3, the \( p \) value was looked up on a table rather than determined by the statistical package. Spearman’s ranks correlation was statistically significant with a \( p \) value of <0.0001 for the three groups.

Validity hypothesis testing

Kruskal Wallis test was used to identify if there was any similarity in the referral tool scores between the three groups. Kruskal–Wallis statistic 17.66. \( P \) value 0.0001. A score above 7 prompts the next step of the referral pathway to be activated, and this was found to be the case for the MWL group, who would be eligible for surgery if they meet the rest of the inclusion criteria. The majority of the respondents in the non-obese and MWL were below the threshold and would not have screened positive for inclusion into the next step of the pathway (Appendix B).

Discussion

The administrative burden was not too great, as the completion was 93.8%. Table 4. If patients do not complete a PROM or omit particular items frequently, this is a potential sign that the questions are difficult to understand, distressing or in some other way unacceptable. However, there are multiple factors which reduce acceptability beyond the nature of questions, including: length of form; time taken to complete; disease burden at time of completion; method of administration; and translation and cultural applicability.

Another study has achieved an acceptability of 95%, however, in that study only 67% of the patients returned the PROM. In our study 100% of PBOTs posted were returned by patients. Furthermore, given this was a postal survey, a completion of 93.8% is good in comparison to the expected completion of 75–80% achievable according to some authors.

Rather than use completion percentages as a proxy for acceptability, other authors suggest direct assessment of patients’ views about a new PROM. This was carried out for the PBOT in the first stage of development. In clinical usage, depending on the dissemination strategy, it may be that forms are not completed satisfactorily and therefore assistance may be required. Further large scale studies are required for a more accurate measure of acceptability.

Maximum endorsement frequencies of >80% occurred in 12 items (at week 1) and 13 items (week 3) for the non-obese non-disease population. The questions with a >80% maximum endorsement in these two groups would not be applicable to the non-obese non-disease group. They included: weight fluctuation; satisfaction with medical
care; satisfaction of most recent surgery; satisfaction with scar and contour. As the PBOT was not devised for this control group, this degree of maximum endorsement was expected.

In the MWL group there were no items with a maximum endorsement frequency of >80%. In the MWLBC group the following items had a >80% maximum endorsement frequency.

1. Have you had any weight fluctuation in the last 6 months?
2. I am satisfied with the medical care I received;
3. I find it difficult to move around;
4. I am unable to independently perform some activities of personal hygiene.

These questions were likely to have similar answers in the post MWLBC group as the inclusion criteria for surgery is stable weight. Furthermore, most patients were satisfied with their surgery and had an improvement in their quality of life. A maximum endorsement frequency of above 80% in a cohort of 10 represents only 2 patients answering differently. It would be worth reviewing the endorsement frequency with a greater number of patients and removing some questions to shorten the PBOT if possible. With a larger cohort, more detailed assessment of item distribution with Rasch analysis may be worthwhile. Some authors believe that if more than 20% of responders score at the maximum level of good or bad health, score distribution general suggests ceiling or floor effects, respectively.32

The Flesch Reading Ease score indicates a reading age of approximately 13, which should be fine for the referrer, and satisfactory for the patient. Administration of the PROM may be assisted if necessary Table 5.

Responsiveness was not assessed, and further work is required here.

Interpretability testing has not been carried out and further work is required here. PROMS must provide scores that are easily interpretable to different stakeholders including patients, researchers, clinicians, and policy makers. In the Supporting Documentation Information on scoring and inclusion and exclusion criteria has been included. Appendix B.

The reliability testing with Cronbach’s Alpha reflected acceptable internal consistency in each group with scores of 0.788, 0.724 and 0.603 for non-obese, MWL and MWLBC groups respectively. When standardised, this became 0.837, 0.763 and 0.811 respectively. Table 6. Cronbach’s

| Table 4 | Administrative Burden: Completion rate of score in each group. |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| Non obese non disease group | Post MWL | Post MWLBC | Total Mean |
| Percentage completions week 1 | 94.4% (91–99%) | 93.2% (90–98%) | 93.5% (90–99%) | 93.7% |
| Percentage completion week 3 | 94.3% (91–99%) | 94.1% (91–99%) | 94.9% |
| Maximum endorsement week 1 > 80% | 13 items >80% | 0 items >80% | 4 items >80% | 63.03% |
| Maximum endorsement week 3 > 80% | 12 items <80% | 0 items >80% | 3 items >80% | 62.39% |

<table>
<thead>
<tr>
<th>Table 5</th>
<th>Flesch reading ease scores.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>Notes</td>
</tr>
<tr>
<td>90.0–100.0</td>
<td>easily understood by an average 11-year-old student</td>
</tr>
<tr>
<td>60.0–70.0</td>
<td>easily understood by 13- to 15-year-old students</td>
</tr>
<tr>
<td>0.0–30.0</td>
<td>best understood by university graduates</td>
</tr>
</tbody>
</table>

| Table 6 | Internal consistency: Cronbach’s coefficient alpha for 3 groups’ score. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| n | Mean | Variance | Std. deviation | Cronbach’s alpha | CA standardised |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Non obese group | 10 Week 1 | 67.28 | 61.895 | 7.87 | 0.788 | 0.837 |
| | 10 Week 3 | 67.27 | 61.895 | 7.87 | 0.788 | 0.837 |
| | 20 Average | 67.27 | 58.64 | 7.66 | 0.788 | 0.837 |
| MWL group | 10 Week 1 | 63.38 | 119.74 | 10.94 | 0.724 | 0.763 |
| | 10 Week 3 | 63.38 | 119.74 | 10.94 | 0.724 | 0.763 |
| | 20 Average | 63.38 | 113.44 | 10.65 | 0.724 | 0.763 |
| MWLBC group | 10 Week 1 | 67.35 | 45.59 | 6.75 | 0.603 | 0.811 |
| | 10 Week 3 | 67.35 | 45.55 | 6.75 | 0.603 | 0.811 |
| | 20 Average | 67.35 | 43.16 | 6.57 | 0.603 | 0.811 |
| All groups | 30 Week 1 | 83.00 | 281.49 | 16.78 | 0.894 | 0.911 |
| | 30 Week 3 | 83.67 | 285.11 | 16.89 | 0.899 | 0.916 |
| | 60 Average | 83.67 | 280.281 | 16.74 | 0.899 | 0.916 |
alpha is grounded in the ‘tau equivalent model’ which assumes that each test item measures the same latent trait on the same measure. When each group was assessed independently the Cronbach’s Alpha underestimated reliability34 because the small size of the group violated the assumption of tau-equivalence. However, heterogeneous test items can also violate the assumptions of the true score equivalence or tau-equivalent model34 which may be the cause of the difference between the standardised item alpha and Cronbach’s alpha. Therefore a further Cronbach’s alpha was carried out on all groups at all time periods (two tailed) resulting in a Cronbach’s alpha of 0.894. This is considered a good to excellent score.

Test-retest reliability was estimated by the Bland and Altman’s method for agreement and all the scores for all periods (n = 60), resulting in a Cronbach’s alpha of 0.894. This is considered a good to excellent score.

Limitations

This preliminary validation study was carried out in 60 people. Further larger studies are required, ideally with
multi centres. This would give a better understanding of the acceptability, maximum endorsement frequencies and reliability testing.

The authors can foresee some problems with the clinical photographs. Funding will be an issue, as will the photographs themselves. Medical photographs can make patients feel vulnerable and could prove to be a barrier to referral. However, recent studies show that patient comfort with full body photography improves quickly as they move through the surgical process and the senior authors have found that these patients are very willing to have pictures taken if it improves their chances of getting funding.

The PBOT has been designed as part of the referral pathway and therefore, in the early stages patients who are keen to be approved for MWLBC may feel coerced to complete the PROM section of the form. Therefore, appropriate consent will need to be carried out before administration of the form. Other studies have examined whether their PROMs can be used to compare new techniques, surgical teams and units. Given the novel stage of the surgical process and the senior authors have found this, it would be worth examining whether the PBOT can achieve this.

It is recommended that comprehensive assessment of outcome should include a combination of generic and specific measures. One limitation of this study is that it has not been compared to other specific measures, as currently there are no other validated MWLBC PROMs published.

Further work required:

1. Development on a clinical population
2. Development on a non-clinical population
3. Maximum endorsement frequencies in a larger group of patients
4. Responsiveness, by asking the same patient cohort to complete the PBOT before and after surgery.
5. Interpretability testing
6. Inter rater reliability
7. Cultural and language translations required.

Conclusion

This new PROM was seen to be reliable both in terms of the internal consistency and test-retest reliability. Comparison to the DAS24 demonstrated it to be valid; however there need to be further larger validation studies, with comparison to functional PROMs such as the SF-36 or Kettering “Body Q”.

Funding

The William Rooney Plastic Surgery and Burns Research Trust.

Conflicts of interest

Nada Al-Hadithy, Ken Stewart and Mark Soldin devised the PBOT.

Ethical approval

All studies conformed to the World Medical Association Declaration of Helsinki (June 1964) and subsequent amendments.

The research protocol was approved by South East Scotland regional ethical committee. Reference number: 10/S1102/2.

R&D Approval: Approval from NHS Lothian. 2010/SJ/PS/01.

Acknowledgements

The British Association of Plastic Reconstructive and Aesthetic Surgeons.

The William Rooney Plastic Surgery and Burns Research Trust and Mr JD Watson for research funding.

The Medical photography department at the Royal Infirmary Hospital, Edinburgh and St John’s Hospital, Livingston, Scotland.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.bjps.2014.07.004.

References

Development of a validated PROM for massive weight loss body contouring patients


