

The Telemark Breast Score: a Valid Method for Evaluation of Outcome after Breast Surgery

Anadi Begic, MD*
Birgit Stark, MD, PhD†

Background: “Telemark Breast Score” (TBS) has been developed at Telemark Hospital in Norway for evaluation of results after breast surgery based on standardized patients’ photographs taken as a part of daily routine. Its reliability has recently been tested and approved. The external validity of the TBS was assessed by matching its data against the internationally recognized Breast-Q (BQ) questionnaire as a further step to study the validity of this new tool.

Methods: The ideal distribution of breast volume is 45% of the total volume above and 55% below the nipple, and a 40° slope line at the upper pole. TBS makes the evaluation of these parameters of breast aesthetics more explicit. The method has been tested on photographs from 31 patients operated on for breast cancer with the Deep Inferior Perforator Flap. The evaluation was done by an independent experienced plastic surgeon earlier participating in the test–retests. The external validity of TBS was investigated against domains 1 and 3 of the BQ reconstruction module. The concordance between ratings was analyzed.

Results: Concordance between TBS items and BQ domain 1 items regarding patient satisfaction, and between TBS items and BQ domain 3 items regarding how the patient experienced the outcome of breast reconstruction was relatively high except for 6 comparisons where we could not statistically ensure that more pairs were concordant than discordant. A total of 178 comparisons appeared to be concordant. This means that for all other comparisons, there was a preponderance of pairs of concordant observations, which indicates that measurements from the 2 instruments follow each other.

Conclusion: The present data indicate that the TBS can be recommended as a valid tool to professionals for assessment of the outcome after breast reconstruction. (*Plast Reconstr Surg Glob Open* 2017;5:e1240; doi: 10.1097/GOX.0000000000001240; Published online 22 February 2017.)

Historically, the evaluation of aesthetic outcome after breast surgery has been highly subjective. Objective evaluation of surgical results after breast surgery is necessary if we are to perform critical analyses and refine reconstructive techniques. Over the last decade,

patient-reported outcome instruments, with greater focus on patient satisfaction and quality-of-life after oncologic treatment and reconstruction of the breast, have been developed and validated in several languages.^{1,2} The use of these questionnaires has become increasingly popular.

The Breast-Q (BQ) is one such patient-based questionnaire developed to evaluate outcome after breast surgery. It has 5 different modules (augmentation, reduction/mastopexy, mastectomy, reconstruction, and breast-conserving therapy), and includes 2 domains: health-related quality-of-life and patient satisfaction.^{3,4}

We recently described a new instrument for outcome assessment: the Telemark Breast Score (TBS) (Fig. 1). This instrument is based on 2D photographs taken from 3 standard views, and assesses surgical outcome in terms of volume (size), shape (upper pole, ptosis and aesthetic proportion), and symmetry. This tool enables the profes-

From the *Department of Plastic Surgery, Telemark Hospital, Skien, Norway; and †Department for Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden.

Received for publication May 25, 2016; accepted December 30, 2016.

Partly funded by an unrestricted grant from the Telemark Hospital Trust.

Copyright © 2017 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

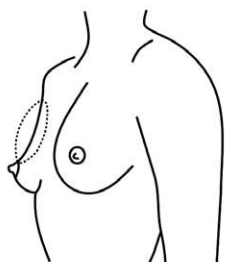
DOI: 10.1097/GOX.0000000000001240

Disclosure: The authors have no financial interest to declare in relation to the content of this article. The Article Processing Charge was paid for by the authors.

Notes :

Size of the Breast

- small ☐
- medium ☐
- large ☐

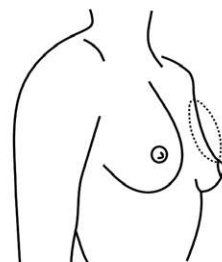


right breast 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

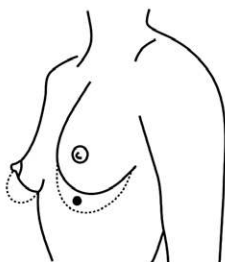
Breast superior pole profile

5 - natural slope

1 - too concave/deformed or too full/convex



left breast 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

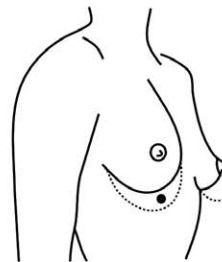


right breast 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

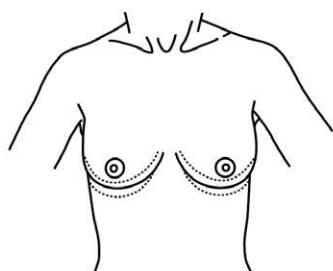
Unsatisfactory form due to ptosis:

5 - no ptosis

1 - big ptosis (gr. IV)



left breast 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐



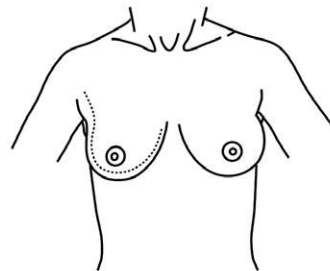
right breast 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

left breast 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Aesthetic proportion

5 - harmonious / proportionate

1 - too big or too small



1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐

Symmetry

Fig. 1. Telemark Breast Score.

Table 1. Demographic Data (y)

No. Patients	Age: Mean (Range)	Follow-up: Mean (Range)	Side of Reconstruction		
			Right	Left	Bilateral
31	54.5 (33–76) y	2 (1–4) y	14	15	2

sional observer to transform a subjective impression into a reproducible objective score. Data from test–retests, previously reported by us, have shown the TBS to be reliable for assessment after breast-conserving surgery and microsurgical reconstruction after mastectomy.⁵ Until now, however, the TBS has not been validated against the patient’s own opinion regarding the surgical result.

The aim of this study was to evaluate the external validity of the TBS for patients who had undergone secondary breast reconstruction after mastectomy, using the deep inferior epigastric perforator (DIEP) flap. The study was based on 2 BQ domains (BQ1a–p and BQ3a–g) specifically addressing patient satisfaction with their breast and the general outcome. The Local Ethics Committee (Norway) approved the study (REK nr: 2013/1107) and it was registered at clinical trials.gov with ID number: NCT02853227.

MATERIALS AND METHODS

Patients

The photographs of 31 consecutive patients operated on and irradiated between 2008 and 2012 for breast cancer with delayed microsurgical breast reconstruction (DIEP flap) at Telemark Hospital, Norway, were eligible for analysis. Primary reconstructions with DIEP are not routine in the Scandinavian health care program for breast cancer patients. For this reason we invited patients with microsurgical

secondary DIEP reconstructions to participate in the study. All patients agreed to participate in the study. Evaluation of the photographs was performed by an independent experienced plastic surgeon who had previously participated in the test–retests mentioned above.⁵ The independent plastic surgeon was blindfolded for the outcome of BQ assessments. Aesthetic assessment of the photographs using TBS was done 2 weeks after taking photographs. Demographic data of the study group are shown in Table 1.

Breast-Q

The validated Norwegian version of the BQ module for reconstructive surgery was used to assess patient satisfaction with outcome during the winter of 2015. Thirty-one patients returned the questionnaire a mean of 2 years after DIEP flap reconstruction (range: 1–4 years). The photographs were taken and BQ was answered by the patients the same day.

We used the subsets of questions included in 2 of the BQ domains in the established BQ module for breast reconstruction, “satisfaction with breasts” (BQ domain 1, BQ1) and “satisfaction with outcome” (BQ domain 3, BQ3), because these were considered most suitable for the purpose of external validity of the TBS (Tables 2, 3).

Incomplete replies to BQ1 were obtained from 3 patients, and to BQ3 from 2 patients.

Statistical Analysis

TBS validity was evaluated by examining the concordance between assessments from TBS and BQ expressed as Svensson’s Measure of Disorder (D) and Monotonic Agreement (MA) presented with 95% jack-knife confidence interval⁶ (Table 4).

The D-value indicates the proportion of disordered paired observations (surgeon and patient) among all pos-

Table 2. BQ1: Question Group Regarding Satisfaction with Breasts

Breast-Q Reconstruction module (postoperative) 2.0 Breast-Q				
The following questions are about your breasts and breast reconstruction surgery. After reading each question, please circle the number in the box that best describes your situation. If you are unsure how to answer a question, choose the answer that comes closest to how you feel. Please answer all questions.				
1. With your breasts in mind, in the past 2 weeks, how <i>satisfied</i> or <i>dissatisfied</i> have you been with:				
	Very Dissatisfied	Somewhat Dissatisfied	Somewhat Satisfied	Very Satisfied
a. How you look in the mirror <i>clothed</i> ?	1	2	3	4
b. The shape of your reconstructed breast(s) when you are wearing a bra	1	2	3	4
c. How normal you feel in your clothes?	1	2	3	4
d. The size of your reconstructed breast(s)	1	2	3	4
e. Being able to wear clothing that is more fitted	1	2	3	4
f. How your breasts are lined up in relation to each other?	1	2	3	4
g. How comfortably your bras fit?	1	2	3	4
h. The softness of your reconstructed breast(s)	1	2	3	4
i. How equal in size your breasts are to each other?	1	2	3	4
j. How natural your reconstructed breast(s) looks?	1	2	3	4
k. How naturally your reconstructed breast(s) sits/hangs?	1	2	3	4
l. How your reconstructed breast(s) feels to touch?	1	2	3	4
m. How much your reconstructed breast(s) feels like a natural part of your body?	1	2	3	4
n. How closely matched your breasts are to each other?	1	2	3	4
o. How your reconstructed breast(s) look now compared to before you had any breast surgery?	1	2	3	4
p. How you look in the mirror <i>unclothed</i> ?	1	2	3	4

Table 3. BQ3: Question Group Regarding Satisfaction with Outcome

	Disagree	Somewhat Agree	Definitely Agree
a. Having reconstruction is much better than the alternative of having no breast(s).	1	2	3
b. I would encourage other women in my situation to have breast reconstruction surgery.	1	2	3
c. I would do it again.	1	2	3
d. I have no regrets about having the surgery.	1	2	3
e. Having this surgery changed my life for the better.	1	2	3
f. The outcome perfectly matched my expectations.	1	2	3
g. It turned out exactly as I had planned.	1	2	3

1. We would like to know how you feel about the *outcome* of your breast reconstruction surgery. Please indicate how much you agree or disagree with each statement:

Table 4. Statistical Project for the Study of the External Validity

TBS	BQ	Purpose of Analysis
The total sum score	The total sum scores from BQ1 and BQ3	Surgeon's overall score using the TBS/patient's total score for BQ1 and BQ3
Median score (item "size" not included in the analyses because it has only 3 increments)	Item (median score for domain 1 and for domain 3) Item (median score for domain 1 and for domain 3) Item scores (questions 1a–1p, respectively, questions 3a–3g)	Surgeon's median score using TBS/patients' median BQ item score, domains 1 and 3 Surgeon's median score for each item in TBS/patients' median BQ item score, domains 1 and 3 Surgeon's median score for each item in TBS/patients' BQ median item score, domains 1 and 3

Table 5. D and MA for Comparison of TBS Sum (Σ) Score/Median Score and BQ Σ Score/Median Score

Comparison	D (95% CI)	MA (95% CI)	n
TBS Σ score/BQ Q1 Σ score	0.48 (0.378–0.581)	0.04 (–0.163 to 0.243)	28
TBS Σ score/BQ Q3 Σ score	0.43 (0.308–0.55)	0.14 (–0.1 to 0.384)	29
TBS median score/BQ Q1 median score	0.25 (0.111–0.392)	0.5 (0.216–0.778)	28
TBS median score/BQ Q3 median score	0.16 (0.047–0.277)	0.68 (0.446–0.905)	29

CI, confidence interval.

Table 6. D and MA with Associated 95% CI, Comparing Median Scores for TBS Items with Those for BQ1 Items

Comparison	D (95% CI)	MA (95% CI)	n
Size/BQ1 median score	0.21 (0.017–0.399)	0.58 (0.202–0.967)	28
Upper pole right breast/BQ1 median score	0.18 (0.059–0.302)	0.64 (0.395–0.881)	28
Upper pole left breast/BQ1 median score	0.16 (0.051–0.270)	0.68 (0.461–0.898)	28
Ptois right breast/BQ1 median score	0.33 (0.176–0.476)	0.35 (0.047–0.648)	28
Ptois left breast/BQ1 median score	0.34 (0.179–0.494)	0.33 (0.013–0.642)	28
Overall aesthetics of right breast/BQ1 median score	0.29 (0.154–0.428)	0.42 (0.144–0.691)	28
Overall aesthetics of left breast/BQ1 median score	0.24 (0.112–0.374)	0.51 (0.251–0.775)	28
Symmetry/BQ1 median score	0.20 (0.073–0.328)	0.60 (0.344–0.854)	28

CI, confidence interval.

sible combinations of pairs and can assume values between 0 and 1. If D = 0, all pairs are concordant, and if D = 1 all pairs are discordant. MA, which is a function of the measure D ($MA = 1 - 2D$), indicates the difference between the proportion of ordered pairs and the proportion of disor-

dered pairs, ranging between –1 and 1. If MA = –1 all couples are disordered, and if MA = 1 all couples are ordered.

There are no thresholds designed for the dimensions D and MA categorizing the degree of validity. Optimum values are very close to 0 (D) and 1 (MA). For comparisons where the analysts made estimates on the 2 instruments with very similar operationalizations, one expects the D-values to be very close to 0 and MA close to 1, which means that almost all couples are concordant.

All data are analyzed using R for Mac version 3.0.1.⁷

RESULTS

Total Scores and Median Scores

The level of concordance between the sum score of TBS and sum score of BQ1/BQ3 is shown in Table 5. When comparing the TBS Σ score with BQ1 Σ score and with BQ3 Σ score, the proportions of pairs that were disordered were D = 0.48 and D = 0.43, respectively. MA values for these comparisons were 0.04 and 0.14, respectively. The fact that the confidence interval overrides zero means that it is not possible to say with statistically certainty that the proportion of ordered pairs was higher than the proportion of disordered pairs. A lower proportion of disor-

dered pairs were seen when comparing median scores than when comparing sum scores.

TBS Item Median Score versus BQ1 and BQ3 Median Scores

The proportion of disordered pairs (D) ranges from D = 0.16 to 0.34 when comparing TBS scores with the median score for domain 1 of BQ (Table 6). Lowest proportion of disordered pairs was seen when comparing items relating to “upper pole” in the TBS and median score of BQ item “patient satisfaction with their breast in general” in BQ1. Highest discordance was seen when comparing items related to ptosis.

The proportion of disordered pairs ranged from D = 0.09 to 0.21 when comparing median scores for TBS items with those for BQ3 questions (Table 7). Lowest discordance was seen when comparing items relating to overall aesthetics of the left breast and how the patient experienced the results of breast reconstruction in the BQ. Highest discordance was seen when comparing items related to ptosis.

Overall Concordance between All TBS Items and Items in BQ1 and BQ3

Concordance between TBS items and items in BQ1 and BQ3, respectively, was analyzed (data not shown). Table 6 shows each TBS item compared with the BQ1 or BQ3 items that gave the lowest proportion of discordant pairs (D-value).

Concordance between TBS items and BQ1 items regarding patient satisfaction with outcome showed the lowest pro-

portion of discordant pairs when comparing item size with BQ1h (satisfaction with softness of breast), item size with BQ1l (satisfaction with how the breast feels to touch), followed by “upper pole right breast” with BQ1i (satisfaction with how similar the breasts are in size), and “upper pole left breast” with BQ1h (satisfaction with softness of breast). For comparisons with the other 5 TBS items, the D-values varied between 0.15 and 0.29 for the BQ item showing the lowest proportion of discordant pairs (Table 8).

Comparison of TBS items with BQ3 items regarding how the patient experienced the outcome of breast reconstruction revealed that for all TBS items, the lowest proportion of discordant pairs was seen when compared with the item BQ3a (reconstruction is a much better option than to not have a breast). When compared with “ptosis-left breast,” BQ3d showed the same proportion of discordance as with BQ3a. In these comparisons D-values varied from 0.01 to 0.11, that is, relatively small proportion of disorder pairs. Highest proportion of disordered pairs was seen when comparing TBS items with items BQ3e, BQ3f, and BQ3g describing patient expectation (Table 8).

DISCUSSION

This study investigates the external validity of the TBS by comparison of answers from TBS questionnaire with those from 2 selected groups of questions from the BQ questionnaire. Data were analyzed at 3 levels: sum score, median score, and separate items (TBS)/items BQ. The lowest degree of concordance was seen when comparing TBS sum score estimates with the sum scores of BQ items followed by comparison of the median scores of the 2 instruments.

Table 7. Order of Consistency between TBS Items and Items in BQ1 and BQ3, D-values (95% CI)

TBS Item	BQ1 Item Showing Best Concordance	D-value (95% CI)	BQ3 Item Showing Best Concordance	D-value (95% CI)
Size	BQ1h and BQ1l	0.08 (0.000–0.160) respective		
		0.08 (0.000–0.172)	BQ3a	0.04 (0.000–0.102)
Upper pole right breast	BQi	0.10 (0.021–0.186)	BQ3a	0.04 (0.000–0.115)
Upper pole left breast	BQh	0.09 (0.012–0.178)	BQ3a	0.04 (0.000–0.107)
Ptosis right breast	BQn	0.29 (0.153–0.428)	BQ3a	0.09 (0.000–0.253)
Ptosis left breast	BQh	0.28 (0.103–0.455)	BQ3a and BQ3D	0.11 (0.000–0.275) respective
				0.11 (0.000–0.275)
Overall aesthetics of right breast	BQc, BQh, and BQl	0.23 (0.095–0.362)	BQ3a	0.03 (0.000–0.071)
		0.23 (0.098–0.372) respective		
		0.23 (0.116–0.346)		
Overall aesthetics of left breast	BQc	0.15 (0.046–0.263)	BQ3a	0.01 (0.000–0.024)
Symmetry	BQl	0.15 (0.051–0.254)	BQ3a	0.01 (0.000–0.024)

CI, confidence interval.

Table 8. D and MA with Associated 95% CI, Comparing Median Scores for TBS Items with Those for BQ3 Items

Comparison	D (95% CI)	MA (95% CI)	n
Size/BQ3 median score	0.19 (0–0.39)	0.62 (0.221–1.000)	29
Upper pole right breast/BQ3 median score	0.18 (0.023–0.336)	0.64 (0.328–0.955)	29
Upper pole left breast/BQ3 median score	0.12 (0.017–0.214)	0.77 (0.573–0.965)	29
Ptois right breast/BQ3 median score	0.21 (0.038–0.375)	0.59 (0.251–0.924)	29
Ptois left breast/BQ3 median score	0.21 (0.046–0.382)	0.57 (0.235–0.909)	29
Overall aesthetics of right breast/BQ3 median score	0.15 (0.035–0.273)	0.69 (0.455–0.93)	29
Overall aesthetics of left breast/BQ3 median score	0.09 (0.002–0.174)	0.82 (0.653–0.996)	29
Symmetry/BQ3 median score	0.13 (0.021–0.242)	0.74 (0.515–0.958)	29

CI, confidence interval.

The lowest degree of concordance seen in the sum score comparisons level could be expected because the instruments are not primarily designed to measure the same variables. This may explain why sum scores are more divergent than median scores at the item level. When comparing median scores, it was seen that the proportion of disordered pairs was lower for BQ3 than for BQ1. Because the sample size was small, it was not statistically possible to confirm differences in D-values between items from BQ1 and BQ3.

In general, there was a lower proportion of disordered pairs for BQ3 than for BQ1 in comparison with TBS. A possible explanation for this is that the patient and the surgeon regard the importance of certain items differently in their subjective assessment of surgical outcome. Should the patient receive improved information on outcome, even higher concordance might be obtained regarding answers to questions regarding expectation (BQ3e–g). The definition of ptosis remains a matter of discussion and even BQ3 D questions on ptosis are difficult to compare. It is not surprising, therefore, that less concordance was seen for this item.

The reason for the present relatively low concordance for some items in this study compared with results reported from other validity studies may be explained by the fact that even though 2 instruments investigate, to some extent, the same underlying variables, they may be operationalized differently from both the professional and non-professional point of view.

Although the surgeon and the patient probably take different factors into account when they answer the TBS and BQ questionnaires, both instruments seem to provide valuable information when evaluating surgical outcome in a consistent manner.

After comparison of all TBS items and BQ domain items, we found only 6 pairs where we could not statistically ensure that the pairs were more concordant than discordant.

This means that for all other comparisons there was a preponderance of concordant pairs, even though the degree varied, indicating that assessments from the 2 instruments follow each other.

As stated previously,^{1,8–10} the main disadvantage of a clinical assessment tool is that what is considered to be a successful result may not concur with the patients' opinion. Furthermore, when measuring subjective parameters bias may be considerable. This may be seen as large interobserver variability not only between clinical observers, but also between clinicians and patients.

BQ is a validated instrument from the patient's point of view and TBS is a clinical tool in which questions are asked and answered through professional approach. Even though BQ and TBS are not exactly comparable in the formulations of questions, the results follow each other.

In conclusion, the results of this study would suggest that the TBS can be recommended as a valid tool for the assessment of outcome after breast reconstruction.

Anadi Begic, MD

Department of Plastic Surgery
Telemark Hospital
3724 Skien, Norway
E-mail: anadi.begic@sthf.no

ACKNOWLEDGMENTS

The authors wish to acknowledge Anna Maria Kling for help with the statistical analyses and the Telemark Hospital for financial support in conducting this study.

REFERENCES

1. Pusic AL, Lemaine V, Klassen AF, et al. Patient-reported outcome measures in plastic surgery: use and interpretation in evidence-based medicine. *Plast Reconstr Surg*. 2011;127:1361–1367.
2. Chen CM, Cano SJ, Klassen AF, et al. Measuring quality of life in oncologic breast surgery: a systematic review of patient-reported outcome measures. *Breast J*. 2010;16:587–597.
3. Pusic AL, Klassen AF, Scott AM, et al. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. *Plast Reconstr Surg*. 2009;124:345–353.
4. Cano SJ, Klassen AF, Scott AM, et al. The BREAST-Q: further validation in independent clinical samples. *Plast Reconstr Surg*. 2012;129:293–302.
5. Begic A, Stark B. The Telemark Breast Score: a reliable method for the evaluation of results after breast surgery. *Plast Reconstr Surg*. 2016;138:390e–400e.
6. Svensson E. Concordance between ratings using different scales for the same variable. *Stat Med*. 2000;19:3483–3496.
7. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2015. <https://www.R-project.org/>
8. Alderman A, Chung K. Measuring outcomes in aesthetic surgery. *Clin Plastic Surg*. 2013;40:297–304.
9. Eriksen K, Nordstrand Lindgren E, Olivecrona H, et al. Evaluation of volume and shape of breasts: comparison between traditional and three-dimensional techniques. *J Plast Surg Hand Surg*. 2011;45:14–22.
10. Ching S, Thoma A, McCabe RE, et al. Measuring outcomes in aesthetic surgery: a comprehensive review of the literature. *Plast Reconstr Surg*. 2003;111:469–80; discussion 481.