Evaluation of cosmetic outcome following breast-conserving therapy in trials: panel versus digitalized analysis and the role of PROMs

Mirelle Lagendijk MD1 | Elvira L. Vos MD, Msc1 | Daan Nieboer MSc, PhD2 | Cornelis Verhoef MD, PhD1 | Evelien M.L. Corten MD, PhD3 | Linetta B. Koppert MD, MSc, PhD1

1Department of Surgery, Erasmus MC Cancer Institute, Rotterdam, The Netherlands
2Department of Public Health, Erasmus MC, Rotterdam, The Netherlands
3Department of Plastic and Reconstructive Surgery, Erasmus MC, Rotterdam, The Netherlands

Correspondence
Mirelle Lagendijk, MD, Department of Oncological Surgery, Erasmus MC Cancer Institute, Rotterdam, The Netherlands. Email: mirelle.lagendijk@erasmusmc.nl

Abstract
Cosmetic outcome is an important quality of life-related end point following breast-conserving therapy (BCT). To advise on a gold standard, we compare cosmetic outcome evaluated by panel and an objective evaluation (BCCT.core software). Second, patient-reported outcome measures (PROMs) are compared to cosmetic outcome evaluation by panel and BCCT.core. Sixty-eight breast cancer patients were included following BCT between 2007 and 2012. Two independent 6-member panels and two observers using the BCCT.core evaluated cosmetic outcome. First, reproducibility, repeatability, and relatedness of panel and BCCT.core were analyzed using the interclass correlation coefficient (ICC). Second, the association between panel/BCCT.core with PROMs (EORTC-QLQ-C30/BR23, EQ-5D-5L, and BREAST-Q) was analyzed with a linear regression and the goodness of fit by the $R^2$. Both panel and BCCT.core evaluations showed “excellent” intraobserver agreement (ICC 0.93 [95% CI: 0.83; 0.97] and 0.93 [95% CI: 0.84; 0.97]) for respectively panel 1 and BCCT.core 1 and “excellent” interobserver agreement (ICC 0.94 [95% CI: 0.90; 0.96] and 0.85 [95% CI: 0.77; 0.91]) respectively for panel and BCCT.core. Association between panel and BCCT.core varied, ICC 0.59-0.69. Only the PROM BREAST-Q showed a significant association with both panel evaluations and BCCT.core observers (panel 1 and BCCT.core 1; $R^2$ of .157 [P = .002] and .178 [P = .001]). Both panel and BCCT.core showed comparable "excellent" intraobserver and interobserver agreement. For future trials evaluating cosmetic outcome following BCT, one of those can be chosen. Solely, the PROM BREAST-Q showed a significant association with panel and/or BCCT.core evaluation. To enable standardized cosmetic outcome evaluation and corresponding patient satisfaction in future trials, at least the BREAST-Q should be combined with a panel or BCCT.core evaluation.

KEYWORDS
breast cancer, breast-conserving surgery, cosmetic outcome, patient-reported outcome measures
Cosmetic outcome following breast-conserving therapy (BCT) is an important quality of life-related end point. Around 60 percent of early-stage breast cancer patients are treated by BCT (ie, breast-conserving surgery with additional whole-breast irradiation) in the Netherlands.²³ Cosmetic outcome is reported as unfavorable in up to 30% of patients treated through BCT.²⁻⁴ Suboptimal cosmetic outcome following BCT decreases health-related quality of life.⁶⁻⁸

To date, no consensus has been reached on the method of cosmetic evaluation for trials in breast cancer surgery.⁹¹⁰ Frequently used parameters are symmetry, scar visibility, and aspect of the nipple-areolar complex.⁹ Within trials, a panel evaluation of photographs of the breast(s) is commonly used to evaluate cosmetic outcome. Panel evaluation is a time-consuming and nonstandardized measurement technique. In recent years, digitalized assessment techniques were introduced within trials.¹¹⁻¹³ An example of such a technique is the BCCT.core software.¹² The agreement for cosmetic outcome results by the digitalized assessment techniques and panel evaluations differ between studies.¹⁴⁻¹⁶

Patient-reported outcome measures (PROMs) are expected to be of great value in the evaluation of individual patient care. The value of PROMs to evaluate cosmetic outcome in trials is not clear yet. A standard set of PROMs is however recently proposed to evaluate breast cancer patients.¹⁷ This set contains generic (EQ-SD-5L and EORTC-QLQ-C30) and disease-specific (EORTC-QLQ-BR23 and BREAST-Q breast-conserving module) questionnaires. This set of PROMs will enable a comparison between patients and studies with, possibly, a simultaneous evaluating of cosmetic outcome.

This study aims to advice on the most appropriate method to evaluate cosmetic outcome following BCT in trials. The goal of this study was to evaluate reproducibility and repeatability of both panel and BCCT.core evaluation. Additionally, it was aimed to evaluate the relationship between panel or BCCT.core evaluation and PROMs.

### 2.1 Patients

The study population consists of patients treated through BCT between 2007 and 2012 at a tertiary referral center. Patients were initially included, following a preoperative breast MRI, to evaluate the influence of tumor and breast volume on the cosmetic outcome following breast conserving therapy (BCT).³ Following approval from the local research ethic committee, an additional informed consent was obtained for all participants to obtain PROMs.

### 2.2 Cosmetic result

Postoperatively, a professional medical photographer took 4 standardized photographs. The photographs were taken (1) face-view with arms down, (2) face-view with arms in flanks (ie, with tension on m. pectoralis), (3) lateral view at 45°, and (4) lateral view at 90° on the affected side. Two separate panel evaluations were performed to conduct the cosmetic outcome based on panel evaluation. A panel consisted of a plastic surgeon, a general surgeon, a radiotherapist, a family doctor, a layperson, and a woman with a history of breast cancer as advised by Cardoso and coauthors.⁹ Both panels consisted of male and female panel members. To assess the repeatability (ie, intraobserver agreement), all members of the first panel repeated the evaluation in 20 patients after a time period of at least 12 months. Cosmetic outcome was assessed through the self-developed "Erasmus MC Panel questionnaire" based on the advised criteria by Cardoso et al.⁶⁻⁹ (Figure S1). The questionnaire consisted of 11 questions: (1) overall cosmetic result, (2) appearance of scar, (3) size and (4) shape of the breast, (5) position, (6) size, and (7) color of the nipple-areola complex, (8) position of infra-mammary fold, (9) color of the skin, (10) telangiectasia, and (11) symmetry. All questions were scored on a four-point scale (0 = "Excellent," 1 = "Good," 2 = "Moderate," and 3 = "Bad"). For each panel member, the mean score over the 11 questions was obtained. For the overall panel evaluation, the mean score of all 6-panel members combined was calculated.

Two observers used the BCCT.core software. As specified by the manufacturer's guideline, specific landmarks are placed on the face-view photograph off all patients. The BCCT.core software than automatically evaluates cosmetic outcome.¹⁵ The software thereby measures among others the symmetry, skin color, and scar visibility. Cosmetic outcome is scored based on a four-point scale (ie, 1 = "Excellent," 2 = "Good," 3 = "Fair," and 4 = "Poor"). To evaluate the intraobserver agreement, the first observer repeated the evaluation in 20 patients at 3 months after initial evaluation.

### 2.3 Patient reported outcome measures

The PROMs used were as follows: EORTC-QLQ-C30 (version 3.0), EORTC-QLQ-BR23 (version 1.0), EQ-SD-5L, and BREAST-Q, "post-operative breast-conserving module."¹⁸ To evaluate the relationship between the cosmetic outcome (panel or BCCT.core) and PROMs, one score from each PROM was chosen. This resulted in 4 PROM scores, namely the "Global Health Status," the "body image," the "health score," and the "satisfaction with breast" of respectively the EORTC-QLQ-C30, EORTC-QLQ-BR23, EQ-SD-5L, and BREAST-Q. PROM scores were calculated according to the questionnaires' scoring protocol. Modules were judged as incomplete according to the questionnaires' protocol; if >50% of the questions are unanswered, the module was classified as incomplete and excluded from analysis.¹⁹

### 2.4 Statistical analysis

Baseline characteristics are presented using the median and interquartile range (IQR) for the age of participants, the time from operation to cosmetic evaluation and the time from operation to completion of PROMs. For the cosmetic outcome by panel and BCCT.core software, the mean and standard deviation (±2.5SD)
were used to present the cosmetic outcome and PROM scores. In case of proportions, the number and percentages for the total group were presented. Interobserver and intraobserver agreements for both panel and BCCT.core were evaluated by calculating the relatedness based on the interclass correlation coefficients (ICC) with 95% confidence interval (CI). An ICC of < 0.40 was considered as an "poor" association, 0.40-0.59 as "fair," 0.60-0.74 as "good," and 0.75-1.00 as "excellent." The relationship between panel or BCCT.core and PROMs was evaluated using a scatter plot with a linear regression line. The goodness of fit was expressed using the $R^2$ and corresponding $P$-value. All analyses were performed using IBM SPSS Statistics 21 (IBM Corp., Armonk, NY).

3 | RESULTS

A total of 68 patients was included with a median age at time of BCS of 54.7 (IQR 50-62) years (Table 1). Five patients (7.4%) had undergone a relumpectomy due to an irradical primary excision. All patients underwent whole-breast irradiation following BCS. Sixty-four (94.1%) received a boost at the tumor-bed. The median total dose of radiotherapy administered was 65 Gy (IQR 65-67.76). Median time postoperatively to medical photographs and PROM assessment was respectively 33 (IQR 18-48), 70 (IQR 63-82), and 41 (IQR 31-59) months.

3.1 | Panel and BCCT.core

Interobserver agreement, ICC, between both panel evaluations and BCCT.core observers was respectively 0.94 (95% CI: 0.90; 0.96) and 0.85 (95% CI: 0.77; 0.91) ICC (Table 2). Intraobserver agreement for panel and BCCT.core observer was respectively 0.93 (95% CI: 0.83; 0.97) and 0.93 (95% CI: 0.84; 0.97) ICC. Relatedness between panel and BCCT.core was ICC 0.69 (95% CI 0.54; 0.79), 0.61 (95% CI 0.43; 0.73), 0.66 (95% CI 0.50; 0.77), and 0.59 (95% CI 0.42; 0.72) respectively for panel 1 versus BCCT.core observer 1, panel 1 versus observer 2, panel 2 versus BCCT.core observer 1, and panel 2 versus observer 2 (Table 2).

3.2 | Panel or BCCT.core and PROMs

In total, 64 (94.1%) patients completed the EORTC-QLQ-C30/BR23, and 58 (85.3%), the EQ-5D-5L and the BREAST-Q. The BREAST-Q showed a significant association with both panel evaluations and BCCT.core observers. For panel 1 and BCCT.core observer 1, the goodness of fit with BREAST-Q was respectively $R^2$ of .157 ($P = .002$) and 0.178 ($P = .001$; Figures 1 and 2). For panel 2 and BCCT.core observer 2, the goodness of fit for the BREAST-Q was $R^2 = .150$ ($P = .003$) and $R^2 = .224$ ($P < .001$) respectively (Figures S2 and S3). For the EORTC-QLQ-C30, EORTC-QLQ-BR23, and EQ-5D-5L no significant association was found for both panel evaluations or BCCT.core observers (see Figure 1 for panel 1 and BCCT.core observer 1, Figure S2 for panel 2 and BCCT.core observer 2).

4 | DISCUSSION

Up to 30% of patients has poor cosmetic outcome following BCT. This emphasizes the strong need for trials aiming to improve...
### TABLE 2 Inter-, intraobserver, and overall agreement for cosmetic outcome—interclass correlation coefficient (95% CI)

|               | Panel 1          | Panel 2          | BCCT.core 1      | BCCT.core 2      |
|---------------|------------------|------------------|------------------|------------------|
| Panel 1       | 0.99 (0.98-0.99) | 0.94 (0.90-0.96) | 0.69 (0.54-0.79) | 0.61 (0.43-0.73) |
| Panel 2       |                  | 0.66 (0.5-0.77)  | 0.59 (0.42-0.72) |                  |
| BCCT.core 1   | 0.93 (0.84-0.97) |                  |                  |                  |
| BCCT.core 2   |                  |                  |                  |                  |

Interpretation ICC: 
- $<0.40$ = "poor," 
- $0.40-0.59$ = "fair,“ 
- $0.60-0.75$ = "good," 
- $0.75-1.00$ = "excellent."

**FIGURE 1** Linear regression for panel 1 and PROMs. Higher panel scores represent unfavorable cosmetic outcome. (A) "Global health status" EORTC-QLQ-C30, higher scores represent higher quality of life. (B) "Body Image" EORCT-QLQ-BR23, higher scores represent a better quality of life. (C) "Health index" EQ-5D-5L, higher scores represent a higher quality of life. (D) "Satisfaction with breast" BREAST-Q, breast-conserving surgery module. Higher scores represent a higher satisfaction.
cosmetic outcome. To date, there is no gold standard technique for cosmetic outcome measurement following BCT within trials. In the current study, intraobserver and interobserver agreements were “excellent” for both the cosmetic outcome obtained through panel or BCCT.core evaluation. Relatedness between panel and BCCT.core was “good” ranging between 0.59 and 0.69 ICC. The association between PROMs and cosmetic outcome evaluation was limited.

So far, the PROM BREAST-Q showed a significant association with both panel and BCCT.core evaluation.

4.1 | Panel and BCCT.core

In contrast to the literature, our results showed an almost perfect agreement between two separate panel evaluations. The number of
panel members and their expertise possibly explains differences found in literature.\(^9\)\(^{10}\) Haloua et al evaluated agreement for different 3-member panel compositions showing a weighted kappa of 0.66.\(^{14}\) In contrast, Racz et al showed high concordance with a weighted kappa of 0.80-0.87 for a 3-member panel with multiple evaluations using an identical questionnaire.\(^{21}\) The identical panel questionnaire and identical composition explain the “excellent” interobserver agreement and forms strength of the current study. The two BCCT.core observers showed comparable interobserver and intraobserver agreements as compared to the panel evaluations.

The association between panel and BCCT.core within the current cohort confirmed results of previous studies showing a weighted kappa ranging 0.64-0.69.\(^{14}\)\(^{15}\) Comparing panel to the BCCT.core, results by Heil et al showed lower agreement, weighted kappa 0.24-0.45.\(^{16}\) Although panel formations differed between studies, the association of panel and BCCT.core is thus stable. Importantly, the good relatedness shows that, although both techniques showed excellent interobserver and intraobserver agreements, the results of a panel or BCCT.core are not interchangeable. The panel evaluated the cosmetic outcome consistently higher than the BCCT.core (data not shown).

The BCCT.core is less time-consuming than a panel evaluation. In line with Haloua et al, an average of 3 minutes was needed to complete the scoring using the BCCT.core (data not shown).\(^{14}\) The interobserver agreement of the BCCT.core was mainly influenced by the marker placement identifying the “upper breast region.” This is a weakness of the software program. In our opinion, a lateral view should be added to ensure more precise marker placement and therewith generate better cosmetic analysis.

### 4.2 PROMs and cosmetic outcome evaluation

A set of PROMs is selected within in the standard breast cancer set by International Consortium of Health Outcome Measurements and is expected to lead the way for current literature comparing PROMs to more traditional outcome measures.\(^{17}\) Since October 2015, the Erasmus MC has incorporated these PROMs at predetermined time points as part of the standard care of all primary breast cancer patients. This study is, to our knowledge, the first to compare the “EQ-5D-5L” and “BREAST-Q” with both panel and BCCT.core evaluation following BCT. The BREAST-Q showed a significant association with the panel and BCCT.core evaluation. Previously, the BREAST-Q has also shown to be associated with differences found in cosmetic outcome following different types of breast surgeries.\(^{7}\)\(^{19}\) In contrast to Kim and coauthors, the “EORTC-QLQ-BR23,” “Body Image” showed a nonsignificant association with both panel and BCCT.-core.\(^7\) This is possibly due to limited patient numbers, which forms a limitation of the current study. PROMs were not administered pre-treatment or at predetermined time points during follow-up which limited the comparison with other cohorts.

The limited relationship of PROMs and cosmetic outcome could be explained by the lack of baseline PROM scores which is a weakness of our study. An adjustment for preoperative PROM or PROMs obtained at predetermined time points could better reflect postoperative treatment effects. In future trials, panel or BCCT.core ideally should be combined with PROMs. In this way, the value of PROMs in research will be further elucidated. When cosmetic outcome evaluation in clinical trials is standardized by either panel or BCCT.core in combination with PROMs, the results can be compared.

### 5 CONCLUSION

This study shows excellent inter- and intraobserver agreements for both panel and BCCT.core evaluation of cosmetic result following BCT. Within trials, one of those can be chosen. The PROM “BREAST-Q” showed significant association with cosmetic outcome as evaluated by panel and BCCT.core. Future trials should at least combine PROMs with either panel or BCCT.core, to evaluate cosmetic outcome and corresponding patient satisfaction following BCT.

### ORCID

Mirelle Lagendijk [http://orcid.org/0000-0002-7738-2698](http://orcid.org/0000-0002-7738-2698)

### REFERENCES

1. NABON breast cancer audit (NBCA) poDIfcaD. Jaarrapportage 2013. [http://wwwclinicalauditnl/jaarrapportage/2013/#dica_rapportages_nbca_borstsparende_therapie_20. Accessed September 1, 2016.](http://wwwclinicalauditnl/jaarrapportage/2013/#dica_rapportages_nbca_borstsparende_therapie_20)
2. Aaronson N, Bartelink H, Vandongen JA, Vandam F. Evaluation of breast conserving therapy: clinical, methodological and psychosocial perspectives. *Eur J Surg Oncol*. 1988;14:133-140.
3. Vos EL, Koning AHJ, Obdeijn I-M, van Verschuer VMT, Verhoef C, van der Spek PJ, et al. Preoperative prediction of cosmetic results in breast conserving surgery. *J Surg Oncol*. 2015;111:178-184.
4. Wang HT, Barone CM, Steigelman MB, Kahlenberg M, Rousseau D, Berger J, et al. Aesthetic outcomes in breast conservation therapy. *Aesthet Surg J*. 2008;28:165-170.
5. Haloua MH, Krekel NM, Winters HA, Rietveld DH, Meijer S, Bloemers FW, et al. A systematic review of oncoplastic breast-conserving surgery: current weaknesses and future prospects. *Ann Surg*. 2013;257:609-620.
6. Hau E, Browne L, Capp A, Delaney GP, Fox C, Kearsley JH, et al. The impact of breast cosmetic and functional outcomes on quality of life: long-term results from the St. George and Wollongong randomized breast boost trial. *Breast Cancer Res Treat*. 2013;139:115-123.
7. Kim MK, Kim T, Moon HG, Jin US, Kim K, Kim J. Effect of cosmetic outcome on quality of life after breast cancer surgery. *Eur J Surg Oncol*. 2015;41:426-432.
8. Waljee JF, Hu ES, Ubel PA, Smith DM, Newman LA, Alderman AK. Effect of esthetic outcome after breast-conserving surgery on psychosocial functioning and quality of life. *Journal of Clinical Oncology*. 2008;26:3331-3337.
9. Cardoso MJ, Cardoso JS, Vrielting C, Macmillan D, Rainsbury D, Heil J, et al. Recommendations for the aesthetic evaluation of breast cancer conservative treatment. *Breast Cancer Res Treat*. 2012;135:629-637.
10. Vrielting C, Collette L, Fourquet A, Hoogenraad WJ, Horiot JH, Jager JJ, et al. The influence of patient, tumor and treatment factors on
the cosmetic results after breast-conserving therapy in the EORTC ‘boost vs. no boost’ trial. EORTC Radiotherapy and Breast Cancer Cooperative Groups. Radiother Oncol. 2000;55:219-232.

11. Cardoso JS, Cardoso MJ. Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. Artif Intell Med. 2007;40:115-126.

12. Cardoso MJ, Cardoso J, Amaral N, Azevedo I. Turning subjective into objective: the BCCT.core software for evaluation of cosmetic results in breast cancer conservative treatment. Breast. 2007;16:456-461.

13. Eder M, Waldenfels F, Swobodnik A, Klöppel M. Objective breast symmetry evaluation using 3-D surface imaging. Breast. 2012;21:152-158.

14. Haloua MH, Krekel NM, Jacobs GJ, Zonderhuis B, Bouman MB, Buncamper ME, Niessen FB, Winters HA, Terwee C, Meijer S, van den Tol MP. Cosmetic outcome assessment following breast-conserving therapy: a comparison between BCCT.core software and panel evaluation. Int J Breast Cancer. 2014;2014:716860.

15. Cardoso MJ, Cardoso JS, Wild T, Krois W, Fitzal F. Comparing two objective methods for the aesthetic evaluation of breast cancer conservative treatment. Breast Cancer Res Treat. 2009;116:149-152.

16. Heil J, Carolus A, Dahlkamp J, Golatta M, Domschke C, Schuetz F, et al. Objective assessment of aesthetic outcome after breast conserving therapy: subjective third party panel rating and objective BCCT.core software evaluation. Breast. 2012;21:61-65.

17. Ong WL, Schouwenburg MG, van Bommel AC, Stowell C, Allison KH, Benn KE, et al. A standard set of value-based patient-centered outcomes for breast cancer: the International Consortium for Health Outcomes Measurement (ICHOM) initiative. JAMA Oncol. 2017;3:677-685.

18. Pusic AL, Klassen AF, Scott AM, Klok JA, Cordeiro PG, Cano SJ. Development of a new patient-reported outcome measure for breast surgery: the BREAST-Q. Plast Reconstr Surg. 2009;124:345-353.

19. Atisha D, Rushing C, Samsa G, Locklear T, Cox C, Shelley Hwang E, et al. A national snapshot of satisfaction with breast cancer procedures. Ann Surg Oncol. 2015;22:361-369.

20. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994;6:284-290.

21. Racz JM, Hong NL, Latosinsky S. In search of a gold standard scoring system for the subjective evaluation of cosmetic outcomes following breast-conserving therapy. Breast J. 2015;21:345-351.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.