Patient-reported outcome measurement compared with professional judgment of cosmetic results after breast-conserving therapy

A.T.P.M. Brands-Appeldoorn MSc,* A.J.G. Maaskant-Braat MD PhD,* W.A.R. Zwaans MD,* J.P. Dieleman PhD,† K.E. Schenk MD,* C.L. Broekhuysen MD,‡ H. Weerdenburg MD,§ R. Daniels,* V.C.G. Tjan-Heijnen MD PhD,|| and R.M.H. Roumen MD PhD*||

ABSTRACT

Background In the present study, we set out to compare patient-reported outcomes with professional judgment about cosmesis after breast-conserving therapy (BCT) and to evaluate which items (position of the nipple, color, scar, size, shape, and firmness) correlate best with subjective outcome.

Methods Dutch patients treated with BCT between 2008 and 2009 were analyzed. Exclusion criteria were prior amputation or BCT of the contralateral breast, metastatic disease, local recurrence, or any prior cosmetic breast surgery. Structured questionnaires and standardized six-view photographs were obtained with a minimum of 3 years' follow-up. Cosmetic outcome was judged by the patients and, based on photographs, by 5 different medical professionals using 3 different scoring systems: the Harvard scale, the Sneeuw questionnaire, and a numeric rating scale. Agreement was scored using the intraclass correlation coefficient (ICC). The association between items of the Sneeuw questionnaire and a fair–poor Harvard score was estimated using logistic regression analysis.

Results The study included 108 female patients (age: 40–91 years). Based on the Harvard scale, agreement on cosmetic outcome between the professionals was good (ICC: 0.78). In contrast, agreement between professionals as a group compared with the patients was found to be fair to moderate (ICC range: 0.38–0.50). The items “size” and “shape” were identified as the strongest determinants of cosmetic outcome.

Conclusions Cosmetic outcome was scored differently by patients and professionals. Agreement was greater between the professionals than between the patients and the professionals as a group. In general, size and shape were the most prominent items on which cosmetic outcome was judged by patients and professionals alike.

Key Words Breast neoplasms, breast-conserving surgery, cosmetic outcomes, cosmetic evaluations, patient satisfaction

Curr Oncol. 2018 Dec;25(6):e553-e561 www.current-oncology.com

INTRODUCTION

Breast-conserving therapy (BCT) has become a widely used treatment for patients with breast cancer or ductal carcinoma in situ. The principle of BCT is wide local tumour excision and postoperative whole-breast irradiation. The goals of BCT are to achieve optimal local tumour control and a good cosmetic outcome. However, breast deformation will likely occur after this dual treatment modality, particularly during the first 3 post-treatment years.

Because of the ongoing surgical and radiotherapeutic developments in BCT and the increasing use of oncoplastic...
surgery, it is important to continue evaluating cosmetic results not only shortly after the primary treatment, but also in the years thereafter. Although the oncologic results of bct are routinely being assessed during long-term follow-up, evaluation of cosmetic results is not yet a standard part of the treatment and follow-up process.

Cosmetic results can be evaluated using various methods. The Harvard scale, which compares an overall cosmetic impression of the treated breast with the untreated breast, is the most widely used, validated, and subjective scoring system, categorizing results as excellent, good, fair, or poor⁴,⁵. Using that scale, “good” or “excellent” cosmetic outcomes have been reported in 71%–93% of patients⁶–¹⁰. Alternatively, a more detailed questionnaire developed by Sneeuw et al.¹¹ addresses the surgical result—shape, size, shape, firmness, and color; and nipple position.

The primary aim of the present study was to compare patient satisfaction about the cosmetic outcome after bct with professional judgment. The secondary objectives were to evaluate which items of the Sneeuw questionnaire correlated best with the scores resulting from use of the Harvard scale.

METHODS

In this cross-sectional study, patients and professionals used the Harvard scale, the Sneeuw questionnaire, and a numeric rating scale to rate the cosmetic results of bct.

Ethics Approval

Review by the Medical Ethical Review Board of Máxima Medical Center [MMC (10 July 2012)] confirmed that the study did not require formal ethics approval. Data were collected anonymously and were treated according to current applicable Dutch law and in accordance with the Declaration of Helsinki¹².

Patients

Patients treated for invasive breast cancer or ductal carcinoma in situ by bct at the breast clinic of MMC in 2008 and 2009 were included in the study. Patients with a prior amputation of the contralateral breast, metastatic disease at presentation or during follow-up, local recurrence, or any prior plastic surgery to the breast were excluded. Between September and October 2012, 3–4 years after their bct, patients were asked to participate in the study.

Photographs and Questionnaires

After obtaining informed consent, patients received questionnaires about the cosmetic outcome after bct, and medical photographs were taken of both breasts. Standardized 6-view photographs (frontal with arms up, frontal with arms at sides, left and right lateral, and left and right lateral oblique) were obtained, omitting the patient’s face. All photographs were saved anonymously into the MMC’s digital program Vision Review Web Client.

Five medical professionals (2 breast surgeons, 1 breast radiologist, 1 plastic surgeon, and 1 breast oncology nurse) were asked to participate and to judge the cosmetic outcomes based on the photographs. Patients did not view their photographs, but provided scores based on their own opinion. Patients and professionals were both asked to evaluate the cosmetic outcome using 3 different scoring systems:

- The Harvard scale
  Four outcomes are possible. An “excellent” cosmetic result score is assigned when the treated breast is nearly identical to the untreated breast [Figure 1(A–F)]. A “good” cosmetic score is assigned when the treated breast is slightly different from the untreated breast [Figure 2(A–F)]. A “fair” score indicates that the treated breast is clearly different from the untreated one, although not seriously distorted [Figure 3(A–F)]. A “poor” score is assigned when the treated breast is seriously distorted [Figure 4(A–F)].
- Sneeuw et al. questionnaire
  The appearances of the treated and untreated breasts are compared based on the surgical scar; breast size, shape, firmness, and color; and nipple position. Four answers are possible for each item: no difference, small difference, moderate difference, or large difference.
- A numeric rating scale scored cosmesis of the treated breast from 1 (“very poor”) to 10 (“excellent”).

Statistical Analysis

The Explora software application [NVZ Plus, Utrecht, Netherlands (https://www.nvz-ziekenhuizen.nl/trainingen/explora/)] was used to collect and process the results of the questionnaires.

Data were analyzed using the IBM SPSS Statistics software application (version 22: IBM, Armonk, NY, U.S.A.). Data are expressed as means with standard deviation and medians with range, as appropriate.

Rating agreement between the professionals and between the patients and the professionals as a group are expressed as intraclass correlation coefficients with corresponding 95% confidence intervals (cIs), using the 2-way mixed model in IBM SPSS Statistics. A logistic regression analysis was performed to identify the items in the Sneeuw questionnaire that were associated with “poor” results on the Harvard scale by assessor type (professional or patient). For that purpose, the Harvard scale was dichotomized into excellent/good (“good”) and fair/poor (“poor”). A Likert scale for the Sneeuw questionnaire items was included in the regression model as a continuous predictor. The corresponding odds ratios and 95% confidence intervals were calculated. Statistical significance was accepted at a 2-sided p value less than 0.05.

RESULTS

General Information

Of 279 patients consecutively treated by bct during the study period, 196 were eligible for the study, and 112 (40%) agreed to participate. A reason for nonparticipation in the study was given by 60 patients (Table 1).

Ultimately, 108 patients [96% (median age: 67 years; range: 40–91 years)] responded to the questionnaires, and pictures of both breasts could be taken. Table 1 shows other patient characteristics. Cosmetic outcome according to the Sneeuw questionnaire was assessed by 108 women; 92
women had evaluable numeric ratings and scores on the Harvard scale.

**Outcome**

Using the Harvard scale, 62% of patients classified their cosmetic outcome as good or excellent. The average judgment of good or excellent by the professionals was 56%. On the Harvard scale, agreement between the professionals was substantial, with an intraclass correlation coefficient of 0.78 (95% CI: 0.72 to 0.83). Agreement between the professionals as a group and the patients was fair to moderate (intraclass correlation coefficients: 0.38–0.50; Table iv).

Table iv illustrates the agreement between the professionals and the patients for the most relevant items of the Sneeuw questionnaire on cosmetic outcome. In particular, fair correlation between the professionals and the patients was observed for the items “size” and “shape.”

Table v shows odds ratios for the most relevant items of the Sneeuw questionnaire compared with outcomes defined on the Harvard scale as “poor.” Generally, a poor

**FIGURE 1** (A–F) Example of a patient who scored her outcome as “excellent.” The mean Harvard score as judged by professionals was also “excellent.”
rating for “size” and “shape” on the Sneeuw questionnaire was strongly associated with a poor outcome rating on the Harvard scale.

Figure 5 illustrates the correlations of the numeric ratings by individual patients with the average ratings by the 5 professionals. Among the lower outcome ratings (<7), scores given by the professionals were slightly higher than those given by the patients. Among the higher ratings (>7), the opposite was observed: scores given by the patients were slightly better than those given by the professionals.

DISCUSSION

In the present study, we compared professional judgment and patient satisfaction with the patient’s cosmetic outcome after BCT. Further, we evaluated how the items of the Sneeuw questionnaire best correlated with scores on the Harvard scaling system.

The present study shows that the judgments of cosmetic outcome after BCT by various professionals are quite comparable and accord with one another. However, agreement on outcomes between patients and the professionals as a group was only moderate. Based on a numeric rating scale, patients give higher ratings when they consider the outcome good or excellent, but lower ratings when they consider the outcome moderate or poor. Patients and professionals both regarded size and shape of the treated breast as the most important aspects when judging cosmetic outcome.

The literature shows that the cosmetic effects of BCT can be evaluated in various ways. Cardoso et al. discussed current methods of esthetic evaluation after BCT and the lack of a “gold standard.” Furthermore, their review provided a set of recommendations that could be used as guidance for the esthetic evaluation of BCT. They concluded that patient self-evaluation, observer evaluation, patient digital photographs, and timing

FIGURE 2  (A–F) Example of a patient who scored her outcome as “good.” The mean Harvard score as judged by professionals was also “good.”
are all components that should be evaluated when scoring the esthetic outcome of BCT. On the other hand, the Harvard scale has been reported to be the scale most widely used for such clinical evaluation. In the present study, use of all of the foregoing methods was analyzed.

Objective measurement of cosmetic outcome with specific software, as performed by Cardoso et al. and Yu et al., is considered to be the most accurate evaluation of asymmetry only. In the present study, we choose to use subjective tools rather than an objective tool to examine cosmetic outcome after BCT because we were interested in more aspects than asymmetry alone. Furthermore, we were interested in studying the grading of cosmetic results by the patients and the professionals, and any correlations. In our

FIGURE 3  (A–F) Example of a patient who scored her outcome as “fair.” The mean Harvard score as judged by professionals was also “fair.”
opinion, such an evaluation could only be accomplished using subjective tools.

The resulting data show that cosmetic outcome by the Harvard scale, when scored by patients, was excellent or good in 62% of evaluations, and that approximately 70% of the patients gave the outcome a score of 7 or more on a numeric scale. Those data closely accord with results in other studies, which reported excellent or good outcomes in 57%–88% of patients, at a median follow-up in the range of 2–5 years.

The original study by Sneeuw et al. showed good correlation between professional assessments of cosmetic results, in accord with our results. However, in the present study, agreement about outcome between

| Reason                              | Patients (n) |
|-------------------------------------|--------------|
| Confrontation with the effects of treatment | 31           |
| Private problems                    | 3            |
| Health problems at present          | 17           |
| Stay in nursing home                | 1            |
| Distance to the hospital            | 8            |
a patient and the professionals as a group appeared to be less and only moderate. We found only a few reports of this specific comparison in the literature. One study reported results of assessments by patients and by oncologists, showing low agreement between them, as observed in the present series.

Our study has some limitations. First is the possible selection bias, given that 60 patients refused participation for specific reasons, and some hundred or more did not participate for unknown reasons. Next, the cross-sectional nature of the design might result in a heterogeneous study population because of variation in the time elapsed since surgery. One of the recommendations in the review by Cardoso et al. involved the importance of standardizing the timing of image capture. In our study, cosmetic outcomes were evaluated after at least 3 years of follow-up, because progression of fibrosis is known to play a part in cosmesis. Images should be acquired before any treatment and at 1 year after radiotherapy. Ideally, images should be repeated at 5 and 10 years’ follow-up. Obtaining long-term follow-up pictures as often as possible is important, because the esthetic result continues to change over time. Another possible flaw in the study is the low number of patients included as a reference group for the items in the logistic regression analysis. Consequently, the corresponding 95% CIs are relatively broad, and thus the level of the association for items with a “poor” result on the Harvard scale might have low reliability.

**CONCLUSIONS**

Scores for cosmetic outcome after bct differ between patients and professionals. Agreement between the professionals is greater than agreement between the patients and the professionals as a group. Size and shape are the most prominent items on which cosmetic outcome is judged. Further research is desirable to develop a good, simple tool to score the cosmetic outcome of bct more extensively during the entire follow-up period.

Notably, a discussion of the patient’s expectations of the cosmetic effects of bct, before treatment, is important. As a result of the present research, we have built the Harvard scale and the numeric rating scale into our electronic patient file to evaluate cosmetic outcome scores by patients and professionals during the follow-up process. By taking that step, discussion about the cosmetic effects of bct will most probably become easier.

**CONFLICT OF INTEREST DISCLOSURES**

We have read and understood Current Oncology’s policy on disclosing conflicts of interest, and we declare that we have none.

**AUTHOR AFFILIATIONS**

*Department of Surgery, †Department of MMC Academy, ‡Department of Plastic Surgery, and §Department of Radiology, Máxima Medical Center, Veldhoven, and ||Division of Medical Oncology, LUMC–School for Oncology and Developmental Biology, Maastricht, Netherlands.

**REFERENCES**

1. Nationaal Borstkanker Overleg Nederland [NABON (Dutch National Breast Cancer Organization)]. Breast Cancer: Dutch Guideline, Version 2.0. Leiden, Netherlands: NABON; 2012. [Available online at: https://oncoline.nl/uploaded/docs/mammacarcinoom/Dutch%20Breast%20Cancer%20Guideline%202012.pdf; cited 3 November 2012]

2. Beadle G, Silver B, Botnick L, Hellman S, Harris J. Cosmetic results following primary radiation therapy for early breast cancer. Cancer 1984;54:2911–18.
### TABLE III

Agreement, expressed as intraclass correlation coefficients (ICCs), \(^a\) of Harvard scale ratings between 108 patients and 5 professionals.

|                | Surgeon 1 |            | Surgeon 2 |            | Plastic surgeon |            | Radiologist |            | Nurse |            | Patients |            |
|----------------|-----------|------------|-----------|------------|----------------|------------|-------------|------------|-------|------------|----------|------------|
|                | ICC 95% CI| ICC 95% CI | ICC 95% CI| ICC 95% CI | ICC 95% CI     | ICC 95% CI | ICC 95% CI  | ICC 95% CI | ICC 95% CI| ICC 95% CI | ICC 95% CI| ICC 95% CI |
| Surgeon 1      |           | 0.76 0.67 to 0.83 | 0.75 0.65 to 0.82 | 0.80 0.71 to 0.86 | 0.67 0.55 to 0.76 | 0.38 0.19 to 0.54 |
| Surgeon 2      |           | 0.79 0.70 to 0.85 | 0.81 0.74 to 0.87 | 0.77 0.68 to 0.84 | 0.46 0.29 to 0.61 |
| Plastic surgeon|           |             |           | 0.85 0.78 to 0.89 | 0.77 0.68 to 0.84 | 0.47 0.30 to 0.62 |
| Radiologist    |           |             |           |             | 0.82 0.75 to 0.87 | 0.48 0.31 to 0.63 |
| Nurse          |           |             |           |             |               |            |             |            |       |            |          |

\(^a\) Two-way mixed-effects model (ICC < 0 = poor agreement; 0–0.20 = slight agreement; 0.21–0.4 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.8 = substantial agreement; ≥0.81 = excellent agreement). CI = confidence interval.

### TABLE IV

Agreement, expressed as intraclass correlation coefficients (ICCs), \(^b\) between 108 patients and 5 professionals on six questionnaire items.

|                | Scar |            | Size |            | Shape |            | Firmness |            | Color |            | Nipple |            |
|----------------|------|------------|------|------------|-------|------------|----------|------------|-------|------------|--------|------------|
|                | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI | ICC 95% CI |
| Surgeon 1      | 0.29 0.10 to 0.46 | 0.52 0.36 to 0.65 | 0.32 0.13 to 0.49 | 0.20 0.00 to 0.38 | 0.26 0.06 to 0.44 | 0.38 0.19 to 0.54 |
| Surgeon 2      | 0.21 0.02 to 0.39 | 0.48 0.32 to 0.62 | 0.42 0.24 to 0.57 | 0.19 0.00 to 0.38 | 0.28 0.09 to 0.45 | 0.38 0.19 to 0.54 |
| Plastic surgeon| 0.22 0.03 to 0.40 | 0.44 0.27 to 0.59 | 0.53 0.36 to 0.66 | 0.26 0.06 to 0.43 | 0.30 0.11 to 0.47 | 0.40 0.21 to 0.36 |
| Radiologist    | 0.24 0.04 to 0.42 | 0.40 0.22 to 0.55 | 0.43 0.25 to 0.58 | 0.24 0.05 to 0.42 | –0.02 –0.22 to 0.18 | 0.38 0.19 to 0.54 |
| Nurse          | 0.22 0.02 to 0.40 | 0.51 0.35 to 0.65 | 0.42 0.24 to 0.57 | 0.11 0.09 to 0.30 | 0.16 0.04 to 0.35 | 0.37 0.18 to 0.54 |
| Professionals mutually | 0.60 0.51 to 0.68 | 0.69 0.62 to 0.76 | 0.70 0.63 to 0.77 | 0.59 0.50 to 0.67 | 0.38 0.29 to 0.48 | 0.67 0.59 to 0.74 |

\(^b\) <0 = poor agreement; 0–0.20 = slight agreement; 0.21–0.4 = fair agreement; 0.41–0.60 = moderate agreement; 0.61–0.8 = substantial agreement; ≥0.81 = excellent agreement.

CI = confidence interval.

### TABLE V

Association between six questionnaire items \(^a\) and poor cosmetic outcome expressed as Harvard scale ratings assessed by 92 patients and 5 professionals.

|                | Scar |            | Size |            | Shape |            | Firmness |            | Color |            | Nipple |            |
|----------------|------|------------|------|------------|-------|------------|----------|------------|-------|------------|--------|------------|
|                | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI | OR 95% CI |
| Surgeon 1      | 3.8 2.2 to 6.6 | 13.6 5.3 to 35.3 | 21.9 5.3 to 90.0 | 7.2 3.3 to 15.6 | 3.3 1.0 to 11.6 | 6.5 3.2 to 13.5 |
| Surgeon 2      | 8.3 3.7 to 18.6 | 14.0 5.3 to 37.1 | –b | –b | 4.7 2.1 to 10.5 | 26.3 7.4 to 92.9 |
| Plastic surgeon| 5.0 2.7 to 9.1 | 81.9 16.5 to 405.4 | 54.8 14.2 to 211.4 | 49.9 13.5 to 184.5 | 5.5 2.7 to 11.1 | 10.0 4.5 to 22.5 |
| Radiologist    | 11.2 4.7 to 26.8 | 62.5 8.1 to 484.5 | 101.1 17.9 to 571.7 | –b | 10.2 4.1 to 24.9 | 31.5 7.4 to 133.6 |
| Nurse          | 5.2 2.8 to 9.6 | 13.4 5.4 to 33.1 | 26.6 6.6 to 107.4 | 11.5 4.2 to 31.4 | 5.9 0.7 to 49.4 | 8.9 3.8 to 20.8 |
| Patients       | 3.4 1.7 to 6.6 | 19.7 6.0 to 65.1 | 37.2 9.1 to 151.3 | 4.0 2.2 to 7.3 | 1.8 0.9 to 3.5 | 4.2 2.2 to 8.2 |

\(^a\) Sneeuw et al., 1992.

\(^b\) Unable to calculate because of low numbers.

\(^c\) Calculated to estimate the relative risk of a poor Harvard score (that is, the fair and poor ratings combined) with a higher rating on the 6-item questionnaire.

OR = odds ratio; CI = confidence interval.

3. Hennings A, Hartmann B, Rauch G, et al. Long-term objective esthetic outcome after breast-conserving therapy. *Breast Cancer Res Treat* 2015;153:345–51.

4. Harris J, Levene M, Svensson G, Hellman S. Analysis of cosmetic results following primary radiation therapy for stages I and II carcinoma of the breast. *Int J Radiat Oncol Biol Phys* 1979;5:257–61.

5. Cardoso M, Cardoso J, Vrieling C, et al. Recommendations for the aesthetic evaluation of breast cancer conservative treatment. *Breast Cancer Res Treat* 2012;135:629–37.

6. Arenas M, Sabater S, Hernández V, et al. Cosmetic outcome of breast conservative treatment for early stage breast cancer. *Clin Transl Oncol* 2006;8:334–8.

7. Wang H, Barone C, Steigelman M, et al. Aesthetic outcomes in breast conservation therapy. *Aesthet Surg J* 2008;28:165–70.

8. Bajaj A, Kon P, Oberk K, Miles D. Aesthetic outcomes in patients undergoing breast conservation therapy for the treatment of localized breast cancer. *Plast Reconstr Surg* 2004;114:1442–9.

9. Hill-Kayser CE, Vachani C, Hampshire MK, Di Lullo GA, Metz JM. Cosmetic outcomes and complications reported by
patients having undergone breast-conserving treatment. *Int J Radiat Oncol Biol Phys* 2012;83:839–44.

10. Ciamella P, Podgornii A, Galeandro M, et al. Toxicity and cosmetic outcome of hypofractionated whole-breast radiotherapy: predictive clinical and dosimetric factors. *Radiat Oncol* 2014;9:97.

11. Sneeuw K, Aaronson N, Yarnold J, et al. Cosmetic and functional outcomes of breast conserving treatment for early stage breast cancer. 1. Comparison of patients’ ratings, observers’ ratings and objective assessments. *Radiother Oncol* 1992;25:153–9.

12. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA* 2013;310:2191–4.

13. Cardoso MJ, Cardoso JS, Oliveira HP, Gouveia P. The breast cancer conservative treatment. Cosmetic results – BCCT. core – Software for objective assessment of esthetic outcome in breast cancer conservative treatment: a narrative review. *Comput Methods Programs Biomed* 2016;126:154–9.

14. Yu T, Eom KY, Young Jang N, et al. Objective measurement of cosmetic outcomes of breast conserving therapy using BCCT. core. *Cancer Res Treat* 2016;48:491–8.

15. Bartelink H, van Dam F, van Dongen J. Psychological effects of breast-conserving therapy in comparison with radical mastectomy. *Int J Radiat Oncol Biol Phys* 1985;11:381–5.

16. de la Rocheftor A, Abner AL, Silver B, Vicini F, Recht A, Harris JR. Are cosmetic results following conservative surgery and radiation therapy for early breast cancer dependent on technique? *Int J Radiat Oncol Biol Phys* 1992;23:925–31.

17. Delouche G, Bachelot F, Premont M, Kurtz JM. Conservation treatment of early breast cancer: long term results and complications. *Int J Radiat Oncol Biol Phys* 1987;13:29–34.

18. Fagundes MA, Fagundes HM, Brito CS, etal. Breast-conserving surgery and definitive radiation: a comparison between quadrantectomy and local excision with special focus on local-regional control and cosmesis. *Int J Radiat Oncol Biol Phys* 1993;27:553–60.

19. Rose MA, Olivatto I, Cady B, et al. Conservative surgery and radiation therapy for early breast cancer. Long-term cosmetic results. *Arch Surg* 1989;124:153–7.

20. Sacchini V, Luini A, Agresti R, et al. The influence of radiotherapy on cosmetic outcome after breast conserving surgery. *Int J Radiat Oncol Biol Phys* 1995;33:59–64.

21. Sarin R, Dinshaw KA, Shrivastava SK, Sharma V, Deore SM. Therapeutic factors influencing the cosmetic outcome and late complications in the conservative management of early breast cancer. *Int J Radiat Oncol Biol Phys* 1993;27:285–92.