Breast Cancer Res Treat
DOI 10.1007/s10549-008-9896-5

CLINICAL TRIAL

Is face-only photographic view enough for the aesthetic evaluation of breast cancer conservative treatment?

Maria João Cardoso · André Magalhães · Teresa Almeida · Susy Costa · Conny Vrieling · David Christie · Jørgen Johansen · Jaime S. Cardoso

Received: 3 January 2008 / Accepted: 4 January 2008

Abstract The breast cancer conservative treatment. cosmetic results (BCCT.core) is a new software tool created for the automatic and objective evaluation of the aesthetic result of BCCT. It makes use of a face-only photographic view of each patient and might thus have been considered insufficient for an accurate evaluation, as others have used multiple views of each patient. The purpose of this work is to compare the performance of the BCCT.core (using face-only views) with a subjective expert analysis using both the face-only and four-view assessment. Photographs in four-views of 150 patients, were evaluated by a panel of experts and a consensus classification was obtained. The agreement between the consensus and the BCCT.core (face-only view) was calculated using the kappa (k) and weighted kappa (wk) statistics. Face-only views, of the same 150 patients, were subsequently sorted out in a different order and sent for individual evaluation by three specialists from the previous panel of experts. The individual agreement between the face-only view and the four-view evaluation by each of the three experts and the consensus was calculated using the same methods. Obtained results were compared to the BCCT.core performance. The software obtained a moderate agreement with the consensus ($k = 0.57$; wk $= 0.68$). The highest value of agreement, from the three experts, between the four-view evaluation and the consensus was identical to the software agreement ($k = 0.55$; wk $= 0.67$). In the face-only view experiment, the highest value of agreement between the experts and the consensus was only fair ($k = 0.37$; wk $= 0.54$). Performance of the software was thus considered equal to that obtained by experts using a four-view evaluation.

Keywords Aesthetic results · Agreement · Breast cancer · Conservative treatment · Expert observers · Face-only view · Four-views · Objective evaluation · Software

M. J. Cardoso (✉) · A. Magalhães · T. Almeida · S. Costa
Department of Surgery, Faculdade de Medicina do Porto, 4200-319 Porto, Portugal
e-mail: mjcard@mail.med.up.pt

C. Vrieling
Genolier Swiss Medical Network, Genolier, Switzerland

D. Christie
East Coast Cancer Center, Tugun, Australia

J. Johansen
Odense University Hospital, Odense, Denmark

M. J. Cardoso · J. S. Cardoso
INESC Porto, Porto, Portugal

Introduction

The oncological outcome of breast cancer conservative treatment (BCCT) is equivalent to mastectomy in terms of disease free survival and overall survival. Absence of a widely accepted standardized tool for aesthetic evaluation of this kind of treatment, however, limits the applicability of any comparative analysis of cosmetic outcome [1, 2]. Methods for evaluating BCCT are traditionally considered as subjective or objective [3–6]. Subjective methods usually evaluate either a patient's appearance directly or through a photograph, by one or several observers [3]. However, results of subjective evaluation show only a modest interobserver agreement, even when performed by expert observers [7, 8]. Objective methods use measurements taken directly from the patient or from photographs, and are based essentially on asymmetries between treated

and non-treated breasts [9, 10]. These methods increase the reproducibility of assessment but it has been argued that they do not take into account the global appearance of aesthetic results, failing to include other aspects such as scar appearance and differences in colour between breasts [4, 11].

The BCCT.core software was developed to provide an objective and automatic evaluation of aesthetical results, not only based on asymmetries, but also based on other parameters extracted from patients photographs. This objective score was translated into a result according to Harris scale [12]. The aim was to develop a reproducible and widely available methodology for the evaluation of aesthetic results in BCCT, enabling effective comparison of outcome between centers [13].

Still, the current version of the BCCT.core software makes use only of face-only photographic view of the patient and no other evaluation is done on the side or oblique views, due to the obvious limitation in standardizing these additional positions [14].

Considering the fact, however, that the face-only view seems to be the one contributing in a major percentage to the final aesthetic result, we decided to measure the agreement between a subjective evaluation by experts, using both face-only or four-view photographs of the same set of patients, and the results obtained by the BCCT.core.

Material and methods

The photographs in four positions of 150 patients, were evaluated by a panel of international experts. Treatment interventions (other than hormonal treatment) had ended at least 1 year before photographs were taken. All patients signed an informed consent to participate in the study. A digital camera with a resolution of at least four megapixels was used to take photographs in four positions: face arms down; face arms up; left side arms up; right side arms up and a consensus classification was obtained. The observers were asked to classify the aesthetic results of all cases in one of four categories of the Harris scale [12]: excellent-treated breast nearly identical to untreated breast; good-treated breast slightly different from untreated; fair-treated breast clearly different from untreated but not seriously distorted; poor-treated breast seriously distorted. In order to obtain a consensus between observers, a Delphi process was used [15, 16]. The method is an attempt to obtain expert opinion in a systematic manner. Experts are recruited individually and anonymously. The survey is conducted over several rounds, and the results are analysed and then reported to the group. The process is considered complete when there is a convergence of opinion or when a point of diminishing returns is reached [17]. The evaluation of each case was considered consensual when more than 50% of observers provided the same classification on aesthetic result [18].The BCCT.core software was then used on the face-only photographic view of the same 150 patients. The agreement between the consensus classification and the BCCT.core evaluation was calculated using the kappa (k) and weighted kappa (wk) statistics, the latter allowing some deviation from perfect agreement. A k score equal to zero was considered to indicate poor agreement; 0.01–0.20 slight agreement; 0.21–0.40 fair agreement; 0.41–0.60 moderate agreement; 0.61–0.80 substantial agreement; 0.81–0.99 almost perfect; and 1.00 perfect agreement [19].

Subsequently, the 150 face-only views of the studied cases were sorted out and sent for additional evaluation by three of the best evaluators of the panel of experts. Participants were chosen from the initial expert panel, and had, amongst all the panellists, the highest number of coincident answers with the final consensus. Individual classification of these three experts in four-view evaluation was also used for comparison of agreement.

The agreement with the consensus of the face-only view and the four-view evaluation by each of the three experts was calculated using the multiple k and wk statistics and compared with the BCCT.core performance.

Results

Main results of agreement analysis are considered in Tables 1 and 2.

The classification shift between the consensus and the BCCT.core was identical to the shifts between the consensus and the evaluations by the three experts (29–33%). However, the BCCT.core generally evaluated patients better (more excellent and good evaluations) than the consensus or each of the three experts. Agreement between computer and consensus slightly passed the moderate level ($k = 0.57$, wk $= 0.67$) and was similar to values obtained by each of the three experts individually using the four-view evaluation of the patients (Table 1).

Considering the face-only view evaluation the BCCT.core, with its moderate agreement with the

Table 1 Four-view (individual), BCCT.core against consensus

Consensus	Errors			Agreement	
	Nicer N (%)	Worser N (%)	Total N (%)	k	wk
Expert 1	19 (13%)	24 (17%)	43 (30%)	0.56	0.69
Expert 2	6 (4%)	31 (22%)	37 (26%)	0.57	0.68
Expert 3	14 (10%)	33 (23%)	47 (33%)	0.51	0.63
BCCT.core	33 (23%)	8 (6%)	41 (29%)	0.57	0.67

Table 2 Face-view (individual), BCCT.core against consensus

Consensus	Errors			Agreement	
	Nicer *N* (%)	Worser *N* (%)	Total *N* (%)	*k*	wk
Expert 1	13 (9%)	54 (38%)	67 (47%)	0.32	0.52
Expert 2	8 (6%)	39 (27%)	47 (33%)	0.46	0.60
Expert 3	31 (22%)	35 (24%)	66 (46%)	0.32	0.49
BCCT.core	33 (23%)	8 (6%)	41 (29%)	0.57	0.67

consensus ($k = 0.57$, wk $= 0.67$), was far better than the agreement between the three experts and the consensus ($k = 0.32$, wk $= 0.46$) (Table 2). As in the four-view evaluation experts showed a trend to more unfavourable evaluations compared with the software.

Discussion

The limited reproducibility of subjective aesthetic evaluation in BCCT has been well documented [8, 18] and is one of the main reasons behind the development of objective methods for this purpose. The latter usually produces extremely reproducible results but the majority of published papers stress out that its isolated use doesn't reflect the complete evaluation of aesthetic results [4–6, 20, 21].

The BCCT.core software arrived as a new software capable of evaluating objectively aesthetic results of BCCT by comparing symmetry, differences in colour and scar appearance in the treated breast. For the first time in an objective analysis tool other than asymmetry differences are taken into account. However, even if more parameters are evaluated, the software uses only a face-view photograph of the patient to evaluate the final aesthetic result [13].

All the previously described objective methods of evaluation made use of the face-only view to attain the final aesthetic result [5, 9–11, 22, 23]. This isolated view seemed enough to get all the asymmetry measurements needed, and all these methods were based on asymmetry only. As a novelty, Fitzal et al. [24] described the Breast Analysis Tool (BAT©) who for the first time uses the Breast Symmetry Index (BSI©) in both the front and lateral views. The Harris scale for subjective cosmetic analyses was correlated with the BSI. The reported inter-observer reproducibility was excellent (Pearson correlation $r = 0.9$; $P < 0.05$) and the BSI was able to significantly differentiate between good and bad cosmesis.

The good results obtained with the BSI when compared with the ones from the BCCT.core emphasize the initial question. Is face-only photographic view enough for getting an accurate evaluation of the final aesthetic result or the inclusion of side views will make the difference between an expert or a panel and an automatic analysis, as it is suggested by Fitzal et al. [24]. In our study the four-view evaluation by each of the expert observers didn't obtain a higher agreement when compared with the software performance and this inclusion of side views will be obviously difficult to standardize, as we previously stated, particularly if we include other characteristics as color differences and scar visibility. Could it be concluded that since our software evaluation is based on more than just asymmetry, it combines enough information to accurately evaluate the cosmetic outcome based on the face-only photographic view.

To test this hypothesis, and as stressed out recently by Zgajnar [25], it is time to test out this new tool in larger trials to accentuate its weak and strongest points so it could be improved to attain the level demanded by patients and doctors.

Acknowledgements This work was partially funded by Fundação para a Ciência e a Tecnologia (FCT) – Portugal through project PTDC/EIA/64914/2006.

Conflicts of interest The first two authors are the main researchers involved in the development of the BCCT.core software. No commercial conflicts of interest exist.

References

1. Fisher B, Anderson S, Bryant J, Margolese RG, Deutsch M, Fisher ER, Jeong JH, Wolmark N (2002) Twenty-year follow-up of a randomized trial comparing total mastectomy, lumpectomy, and lumpectomy plus irradiation for the treatment of invasive breast cancer. N Engl J Med 347:1233–1241
2. Veronesi U, Cascinelli N, Mariani L, Greco M, Saccozzi R, Luini A, Aguilar M, Marubini E (2002) Twenty-year follow-up of a randomized study comparing breast-conserving surgery with radical mastectomy for early breast cancer. N Engl J Med 347:1227–1232
3. Al-Ghazal SK, Blamey RW (1999) Cosmetic assessment of breast-conserving surgery for primary breast cancer. Breast 8:162–168
4. Christie D (1996) A comparison of methods of cosmetic assessment in breast conservation treatment. Breast 5:358–367
5. Sacchini V, Luini A, Tana S, Lozza L, Galimberti V, Merson M, Agresti R, Veronesi P, Greco M (1991) Quantitative and qualitative cosmetic evaluation after conservative treatment for breast cancer. Eur J Cancer 27:1395–1400
6. Vrieling C, Collette L, Bartelink E, Borger JH, Brenninkmeyer SJ, Horiot JC, Pierart M, Poortmans PM, Struikmans H, Van der Schueren E, Van Dongen JA, Van Limbergen E, Bartelink H (1999) Validation of the methods of cosmetic assessment after breast-conserving therapy in the eortc "Boost versus no boost" trial. Eortc radiotherapy and breast cancer cooperative groups. European organization for research and treatment of cancer. Int J Radiat Oncol Biol Phys 45:667–676
7. Cardoso MJ, Santos AC, Cardoso J, Barros H, Cardoso De Oliveira M (2005) Choosing observers for evaluation of aesthetic results in breast cancer conservative treatment. Int J Radiat Oncol Biol Phys 61:879–881

8. Pezner RD, Lipsett JA, Vora NL, Desai KR (1985) Limited usefulness of observer-based cosmesis scales employed to evaluate patients treated conservatively for breast cancer. Int J Radiat Oncol Biol Phys 11:1117–1119
9. Pezner RD, Patterson MP, Hill LR, Vora N, Desai KR, Archambeau JO, Lipsett JA (1985) Breast retraction assessment: an objective evaluation of cosmetic results of patients treated conservatively for breast cancer. Int J Radiat Oncol Biol Phys 11:575–578
10. Van Limbergen E, van der Schueren E, Van Tongelen K (1989) Cosmetic evaluation of breast conserving treatment for mammary cancer. 1. Proposal of a quantitative scoring system. Radiother Oncol 16:159–167
11. Krishnan L, Stanton AL, Collins CA, Liston VE, Jewell WR (2001) Form or function? Part 2. Objective cosmetic and functional correlates of quality of life in women treated with breast-conserving surgical procedures and radiotherapy. Cancer 91:2282–2287
12. Harris JR, Levene MB, Svensson G, Hellman S (1979) Analysis of cosmetic results following primary radiation therapy for stages i and ii carcinoma of the breast. Int J Radiat Oncol Biol Phys 5:257–261
13. Cardoso JS, Cardoso MJ (2007) Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. Artif Intell Med 40:115–126
14. Cardoso MJ, Cardoso J, Amaral N, Azevedo I, Barreau L, Bernardo M, Christie D, Costa S, Fitzal F, Fougo JL, Johansen J, Macmillan D, Mano MP, Regolo L, Rosa J, Teixeira L, Vrieling C (2007) Turning subjective into objective: the BCCT.Core software for evaluation of cosmetic results in breast cancer conservative treatment. Breast 16:456–461
15. Hasson F, Keeney S, McKenna H (2000) Research guidelines for the delphi survey technique. J Adv Nurs 32:1008–1015
16. Jones J, Hunter D (1995) Consensus methods for medical and health services research. BMJ 311:376–380
17. Fink A, Kosecoff J, Chassin M, Brook RH (1984) Consensus methods: Characteristics and guidelines for use. Am J Public Health 74:979–983
18. Cardoso MJ, Cardoso J, Santos AC, Barros H, Oliveira MC (2005) Interobserver agreement, consensus over the esthetic evaluation of conservative treatment for breast cancer. Breast 15:52–57
19. Seigel DG, Podgor MJ, Remaley NA (1992) Acceptable values of kappa for comparison of two groups. Am J Epidemiol 135:571–578
20. Al-Ghazal SK, Blamey RW, Stewart J, Morgan AA (1999) The cosmetic outcome in early breast cancer treated with breast conservation. Eur J Surg Oncol 25:566–570
21. Sneeuw KC, Aaronson NK, Yarnold JR, Broderick M, Regan J, Ross G, Goddard A (1992) 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 25:153–159
22. Noguchi M, Miyazaki I (1994) Breast conserving surgery and radiation in the treatment of operable breast cancer. Int Surg 79:142–147
23. Tsouskas LI, Fentiman IS (1990) Breast compliance: a new method for evaluation of cosmetic outcome after conservative treatment of early breast cancer. Breast Cancer Res Treat 15:185–190
24. Fitzal F, Krois W, Trischler H, Wutzel L, Riedl O, Kuhbelbock U, Wintersteiner B, Cardoso MJ, Dubsky P, Gnant M, Jakesz R, Wild T (2007) The use of a breast symmetry index for objective evaluation of breast cosmesis. Breast 16:429–435
25. Zgajnar J (2007) Digital evaluation of breast cosmesis after breast-conserving treatment: end of the beginning. Breast 16:441–442