CLINICAL TRIAL

Comparing two objective methods for the aesthetic evaluation of breast cancer conservative treatment

Maria João Cardoso · Jaime S. Cardoso · Thomas Wild · Wilfried Krois · Florian Fitzal

Abstract Two programs were recently developed for the aesthetic evaluation of results in breast cancer conservative treatment: the *B*reast *C*ancer *C*onservative *T*reatment *co*smetic *re*sults (BCCT.core) and the *B*reast *A*nalyzing *T*ool (BAT). Both make use of a face-only photographic view of the patient and were developed to overcome the lack of reproducibility observed with subjective visual evaluation. The BCCT.core analyses several parameters related to asymmetry, color differences and scar appearance, while the BAT considers only asymmetry measurements. The purpose of this study was to compare the performance of these two methods. *Material and methods* Digital pictures of 59 patients from Porto and 60 from Vienna were evaluated subjectively by two panels using the four-class Harris scale. The Porto photographs had a similar backlight and better quality, and were evaluated by an international panel of 23 experts. The Vienna photographs had different backlight and lower quality, and were evaluated by four students and two breast cancer specialists. All 119 cases were submitted to analysis using the BCCT.core and BAT. Agreement between the software programs and the subjective evaluation was calculated using kappa (k), weighted kappa statistics (wk) and error rate (er). *Results* In overall analysis, BCCT.core program obtained a better agreement with the subjective evaluation ($k = 0.56$; $wk = 0,64$; $er = 0.20$) than the BAT software ($k = 0.39$; $wk = 0.46$; $er = 0.42$) ($P < 0.0007$). Results were again better for the BCCT.core program, when analysing the photographs obtained in Porto ($k = 0.71$; $wk = 0.78$; $er = 0.14$) than for the BAT ($k = 0.35$; $wk = 0.41$; $er = 0.51$) ($P < 0.0003$) while no significant differences in agreement were obtained regarding the Vienna images ($P > 0.1$). *Conclusions* The results suggest that the inclusion of multiple parameters in image analyses of aesthetic results has the potential to improve results. However, picture quality is probably important for analysis of other features besides asymmetry.

Keywords Breast cancer conservative treatment · Aesthetic evaluation · Objective methods · Software · BCCT.core · BAT

M. J. Cardoso (✉)
Department of Surgery, Faculdade de Medicina do Porto, Hospital S. João, Alameda do Prof. Hernâni Monteiro, 4200-319 Porto, Portugal
e-mail: mjcard@mail.med.up.pt

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

J. S. Cardoso
University of Porto – Faculty of Engineering, Porto, Portugal

T. Wild · W. Krois · F. Fitzal
Department of Surgery, Medical University, Vienna, Austria

Introduction

Oncological outcome of breast cancer conservative treatment (BCCT) is considered to be equivalent to mastectomy in terms of disease free survival and overall survival [1, 2]. The absence of standardized tools for the aesthetic evaluation of this kind of treatment, however, limits the applicability of any comparative analysis of cosmetic outcome, one of its main objectives. Methods for evaluating aesthetic results of BCCT are traditionally considered as subjective or objective [3–6]. Subjective methods have been largely used, either through direct patient evaluation or through a photograph, performed by one or several observers [3]. However, results of subjective evaluation

show only a modest interobserver agreement, even when performed by experts [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]. Until now they have not been widely used in clinical practice, probably due to the absence of photographic documentation in most treated patients [11]. They are associated with increased reproducibility of assessment but it has been argued that they do not take into account the global appearance of aesthetic results [4–6, 12].

Two objective methods were recently developed for the objective evaluation of aesthetic results in BCCT, the breast cancer conservative treatment cosmetic result (BCCT.core) [13] and the breast analyzing tool (BAT) [14]. Both methods make use of a face-only photographic view of the patient. The BCCT.core analyses several parameters related to asymmetry, color differences and scar appearance while the BAT takes in consideration only measurements of asymmetry. In order to obtain an evaluation of software performance, we analyzed photographs from patients submitted to BCCT with both tools and compared the results with those obtained in a subjective evaluation performed by experts.

Material and methods

A total of 120 patients submitted to BCCT were photographed and evaluated. Of these 60 were photographed in Porto with identical backlight and background. Another 60 patients were photographed in Vienna using heterogeneous backlight and background.

Subjective evaluation

The 60 cases photographed in Porto were evaluated by a panel of 23 international experts, who were asked to classify the aesthetic result of each case in one of four Harris scale categories: 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. A Delphi procedure was used to obtain a consensus between observers, which was possible in 59 of the 60 cases [7].

The 60 cases photographed in Vienna were evaluated by two experts (breast cancer surgeons from a cancer centre with more than 300 cases per year) and four non-experts (medical students) using a three-point scale (good, fair or poor). The mean values of evaluations were used as the final subjective score [14].

To obtain a single scale for comparison with the software programs, excellent and good results from the Harris scale in the Porto series were merged together to form a three-point scale.

Objective evaluation

The 119 cases where a final subjective classification was obtained were evaluated by the BAT software. This software uses well-defined landmarks (jugulomamillary distance and distances from the nipples to the edge of the breast) and calculates the difference between left and right breasts. This difference in length is multiplied by the surface area difference and is noted as percent difference and as difference factor. The values obtained can be converted to a simplified 3-point Harris scale (good, fair, poor) [14].

The BCCT.core software was also applied to evaluate the 119 digital photographs. This program automatically evaluates several indices used for the aesthetic evaluation of BCCT (asymmetry, skin colour change and scar visibility). The BCCT.core then uses artificial intelligence techniques to translate these measures into an overall objective classification of aesthetical results reported to the user as excellent, good, fair or poor [15].

To obtain a homogeneous final classification, the BCCT.core evaluations were translated into a 3-point scale by merging good and excellent cases together.

Statistical analysis

Agreement between the subjective evaluation and the BCCT.core and the BAT classifications was calculated using the kappa (k) and weighted kappa (wk) statistics, the latter allowing for some deviation from perfect agreement. A kappa score equal to 0 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. The error rate of each method was also calculated, defined as the percentage of cases in which the software evaluation did not agree with the consensus. *P* values of 0.05 or less were considered statistically significant.

Results

The results of agreement between software and subjective evaluation are displayed in Table 1.

For the Porto photographs, agreement was substantially higher for the BCCT.core program. In the Vienna photographs, however the agreement with the subjective classification was only moderate, for both methods. Overall

Table 1 Agreement and error rate between objective evaluation using BCCT.core and BAT software and subjective classifications over the Porto, Vienna and all patients

	Porto patients (59)		Vienna patients (60)		All patients (119)	
	Consensus of experts		Mean of experts and students			
	BAT	BCCT.core	BAT	BCCT.core	BAT	BCCT.core
Kappa	0.35 ($P = 1.61E-5$)	0.71 ($P = 1.14E-12$)	0.41 ($P = 3.68E-6$)	0.43 ($P = 8.52E-6$)	0.39 ($P = 2.35E-10$)	0.56 ($P = 0$)
Weighted kappa	0.41 ($P = 3.85E-6$)	0.78 ($P = 3.17E-13$)	0.50 ($P = 2.67E-8$)	0.51 ($P = 2.21E-7$)	0.46 ($P = 8.15E-13$)	0.64 ($P = 0$)
Error rate	0.51	0.14	0.35	0.26	0.42	0.20
P-value	0.0003		0.16		0.0007	

results revealed a significantly higher agreement with the BCCT.core program.

Discussion

With the warranty of oncological results identical to mastectomy [1, 2] cosmetic results are now also a major endpoint of BCCT. Their evaluation is fraught with difficulty, subjectivity and limited reproducibility. The BCCT.core and BAT softwares, were developed in different institutions but share the goals of providing a simple tool for reproducible evaluation of aesthetic results in BCCT [14, 15]

Different methodologies were used to arrive at a consensus analysis in the two datasets of photographs [7, 14]. The difficult and time consuming process of obtaining a consensus analysis in a large number of cases compelled us to use previous results obtained with different observers and scales.

This subset analysis quite unexpectedly leads to the most curious results obtained in this study. The BCCT.core program agreed more with the consensus in the Porto photographs and overall, but performed similarly to the BAT software in the Vienna patients.

Given the higher definition of the Porto photographs it can be hypothesised that the BAT software using only asymmetry measurements, maintains its performance independently of picture quality, while the BCCT.core that uses other parameters such as colour differences and scar appearance, performs less well in lower quality pictures [14, 15] .

A good method of evaluation should be consistent and simple to use with all kinds of photographs but, on the other hand, some standards for picture quality (definition, backlight, background) are needed, if one is to expect discriminative power in evaluation of aesthetic results.

References

1. Fisher B, Anderson S, Bryant J et al (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(16):1233–1241. doi:10.1056/NEJMoa022152
2. Veronesi U, Cascinelli N, Mariani L et al (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(16):1227–1232. doi:10.1056/NEJMoa020989
3. Al-Ghazal SK, Blamey RW (1999) Cosmetic assessment of breast-conserving surgery for primary breast cancer. Breast 8(4):162–168. doi:10.1054/brst.1999.0017
4. Christie D, O'Brien M, Christie J et al (1996) A comparison of methods of cosmetic assessment in breast conservation treatment. Breast 5:358–367. doi:10.1016/S0960-9776(96)90004-1
5. Sacchini V, Luini A, Tana S et al (1991) Quantitative and qualitative cosmetic evaluation after conservative treatment for breast cancer. Eur J Cancer 27(11):1395–1400
6. Vrieling C, Collette L, Bartelink E et al (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(3):667–676. doi:10.1016/S0360-3016(99)00215-1
7. Cardoso MJ, Cardoso J, Santos AC, Barros H, Oliveira MC (2005) Interobserver agreement and consensus over the esthetic evaluation of conservative treatment for breast cancer. Breast 15:52–57
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(6):1117–1119
9. Pezner RD, Patterson MP, Hill LR et al (1985) Breast retraction assessment: an objective evaluation of cosmetic results of patients treated conservatively for breast cancer. Int J Radiat Oncol Biol Phys 11(3):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(3):159–167. doi:10.1016/0167-8140(89)90016-9
11. Van Limbergen E, Rijnders A, van der Schueren E, Lerut T, Christiaens R (1989) Cosmetic evaluation of breast conserving

treatment for mammary cancer. 2. A quantitative analysis of the influence of radiation dose, fractionation schedules and surgical treatment techniques on cosmetic results. Radiother Oncol 16(4):253–267. doi:10.1016/0167-8140(89)90037-6

12. 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(12):2282–2287. doi:10.1002/1097-0142(20010615)91:12<2282::AID-CNCR1259>3.0.CO;2-0

13. Cardoso MJ, Cardoso J, Amaral N et al (2007) Turning subjective into objective: the BCCT.core software for evaluation of cosmetic results in breast cancer conservative treatment. Breast 16(5):456–461. doi:10.1016/j.breast.2007.05.002

14. Fitzal F, Krois W, Trischler H et al (2007) The use of a breast symmetry index for objective evaluation of breast cosmesis. Breast 16(4):429–435. doi:10.1016/j.breast.2007.01.013

15. Cardoso JS, Cardoso MJ (2007) Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. Artif Intell Med 40(2):115–126. doi:10.1016/j.artmed.2007.02.007