

[Article type]

Original article

[Title]

Cosmetic Evaluation Methods Adapted to Asian Patients after Breast Conserving Surgery and Examination of the Necessarily Elements for Cosmetic Evaluation

[Authors information]

Yuki Nohara^{1,2*}, Tomoko Ogawa¹, Noriko Hanamura¹, Hisamitsu Zaha², Hiroko Kimura¹, Yumi Kashikura¹, Takashi Nakamura¹, Aya Noro¹, Nao Imai¹, Mai Shibusawa¹

¹Department of Breast Surgery, Mie University Hospital, Mie, Japan

²Department of surgery, Nakagami Hospital, Okinawa, Japan

[Corresponding author]

Yuki Nohara

Mailing address:

Department of Breast Surgery, Mie University Hospital

2-174 Edobashi Tsu, Mie, Japan, 514-8507

Telephone:

+81-59-231-5584

Fax number:

+81-59-231-5584

E-mail address:

y-nohara@clin.medic.mie-u.ac.jp

[Running title]

Cosmetic Evaluation Method for Asian Patients with Breast Conserving Surgery

[Abstract]

Purpose

Although various strategies have been reported, there are no defined criteria for cosmetic evaluation methods after breast conserving surgery (BCS). Since Asians tend to have smaller breasts, indistinct inframammary folds, and conspicuous scars, differences in the cosmetic results are expected. So we examined 2 subjective methods and one objective method to determine the differences, and elements necessary for a cosmetic evaluation after BCS.

Methods

Frontal photographs of 190 Japanese were evaluated using the Harris scale (Harris) and the evaluation method proposed by the Japanese Breast Cancer Society Sawai group (Sawai group) as the subjective methods, and the Breast Cancer Conservation Treatment cosmetic results (BCCT.core) as the objective method, respectively. In order to examine the necessary elements for developing a new ideal method, 100 out of 190 were selected and assessed separately by 6 raters using both the Harris and modified Sawai group methods in the observer assessment. The correlation between the 2 methods was examined using the Spearman's rank-correlation coefficient.

Results

The results of the BCCT.core and the other 2 methods were clearly different. In the observer assessment, the consensuses of the 6 raters were evaluated as follows; 27, 27, 26, and 20 cases were evaluated as Excellent, Good, Fair, and Poor, respectively. For the Spearman's rank-correlation coefficient, values higher than 0.7 indicated a strong correlation, as seen by the values of 0.909 for the breast shape and 0.345 for the scar. The breast shape accounted for the most significant part of the evaluation, and the scar had very little correlation.

Conclusion

In this study, we recognized a clear difference between the subjective and objective evaluation methods, and identified the necessary elements for cosmetic evaluation. We would like to continue developing an ideal cosmetic evaluation that is similar to subjective one and is independent from raters.

[Key words]

Aesthetics, Asian Continental Ancestry Group, Mastectomy, segmental, Outcome assessment

Introduction

Breast cancer is the most common cancer, and is the leading cause of cancer death among women worldwide [1]. Despite the recent increasing incidence of breast cancer, the mortality due to the disease has been decreasing because of advanced treatment methods and early detection. Consequently, even patients who developed breast cancer can survive for longer periods and the requirements for breast cancer surgery have changed. In breast cancer conserving surgery (BCS), in particular, favorable cosmetic outcome has been defined as an important endpoint.

Although various evaluation methods have been reported, there are no defined criteria for a “favorable cosmetic outcome”. Additionally, there are 2 methods used in cosmetic evaluation: the subjective method and objective method, respectively. Harris et al. initially proposed a 4-staged subjective evaluation method scored as “excellent,” “good,” “fair,” “poor” in 1979 (Harris) [2-5]. In 1988, Aaronson et al. reported a scoring method where a comprehensive evaluation was performed by scoring the subjective assessment [6]. In Japan, the evaluation of postoperative cosmetic outcome by the Japanese Breast Cancer Society Sawai group (Sawai group) was proposed in 2004. In this scoring method, the evaluation was conducted using 8 items related to the breast forms (breast size, breast shape, scar, breast firmness, nipple and areola size/shape, nipple and areola color tone, nipple position, and position of the maximum descent point of the breast) [7].

A desirable cosmetic evaluation method is an objective method that is independent of raters. As an objective method, Pezner et al. proposed the Breast Retraction Assessment (BRA) in 1985 [4, 5, 8]. In 2007, cosmetic evaluation methods utilizing photographic analysis software were proposed by 2 groups; the Breast Analyzing Tool (BAT), which was developed by Fitzal et al. [4, 9-11], and the Breast Cancer Conservative Treatment cosmetic results (BCCT.core), which was developed by Cardoso et al. [3-5, 10, 12-16]. With high-quality images, the comparison between BAT and BCCT.core showed that BCCT.core, in which color assessment and scar assessment would also be carried out, showed a higher agreement rate to consensus [5, 10]. In Europe, BCCT.core was used in a number of papers, and it was considered to be comparable to the several raters [3, 5, 17, 18]. Nevertheless, the κ (weighted kappa ; $w\kappa$) of BCCT.core compared to the consensus was 0.34 (0.53) and the agreement rate with the consensus was not necessarily high [10]. In addition, Asians including Japanese have smaller breasts than Caucasians, and therefore, inframammary fold is not as apparent in some cases. Since scars in Asians are more noticeable than in Caucasians, the BCCT.core, which had been

developed based on the data in Europeans, was expected to affect the evaluation results. In this study, we compared and evaluated Harris (as a subjective evaluation), the Sawai group (as a subjective evaluation) and the BCCT.core (as an objective evaluation) to discuss the differences between the subjective and objective methods as well as address the issues that occurred when those methods were adapted to Asians including Japanese. Then we conducted observer assessment and examined the necessary elements for developing an ideal cosmetic evaluation method that is similar to subjective methods and is not influenced by raters.

Methods

This study consisted of 2 parts: the primary evaluation and an observer assessment.

In the primary evaluation, the head author himself compared and examined the Harris, the Sawai group and the BCCT.core (developed and owned by Cardoso J, Cardoso M and INESC Porto Breast Research Group) using the frontal photographs of breasts from 190 post-BCS cases. In the observer assessment, 100 out of 190 cases were selected so that the 4-staged scales (Excellent/Good/Fair/Poor) were distributed as equally as possible. Six raters compared the evaluations obtained by Harris and the modified Sawai group and considered the relationship between the elements, which constituted the cosmetic and 4-staged evaluations. Data were collected in accordance with guidelines for human subjects research, as approved by the Institutional Review Board of Mie University Hospital (2817) and Nakagami Hospital (2014025-2).

Photographs

Consent for photographs was obtained from the 190 patients. Photographs were taken with at least 7 mega pixels, in an examination room of approximately 10 square meters size, under ordinary fluorescent lights and a white background wall. The patient faced front with her chest thrust forward so that the outline of outer side of the breasts became apparent. When taking the photographs, we informed the patients that the photographs taken were to be used to compare the pre- and post-operational conditions of breasts and for research on cosmetic evaluation, and obtained their consents.

Primary Evaluation

Methods of cosmetic evaluation

The Harris [2, 4] method evaluated the overall impression using a 4-staged scale. An “excellent” rating means that the treated breast was nearly identical to the untreated breast. A “good” rating means that the treated breast was slightly different from the untreated breast. A “fair” rating means that the treated breast was not seriously distorted but clearly different from the untreated breast, while a “poor” rating means that treated breast was seriously distorted.

The Sawai group was a method that was initially proposed in 2004 and supported by the Japanese Breast Cancer Society [7]. It scores the following 8 items and conducts evaluations with the highest total score of 12 points; breast size (0-2 points), breast shape (0-2 points), scar (0-2 points), breast firmness (0-2 points), nipple-areola complex (NAC) size/shape (0-1 point), NAC color tone (0-1 point), nipple position (0-1 point), and position of the maximum descent point of the breast (0-1 point). In this method, total scores of 11 to 12, 8 to 10, 5 to 7, and 0 to 4 points were defined as “excellent,” “good,” “fair” and “poor,” respectively, indicating that the subjective evaluation results might be obtained pursuant to Harris. This is a subjective scoring method and identifies the elements that influence the cosmetic outcome. However, as it contains many evaluation items, is complicated, and is not easy to use.

In the BCCT.core [3-5, 10, 12-17] several objective evaluations were conducted for a comprehensive evaluation based on the measurement values of the computer software using frontal photographs of breasts. The software conducted cosmetic analysis by putting marks on the jugular notch and both nipples, and outlining the breasts with lines. The endpoints included the BRA, lower breast contour (LBC) [19,20], upward nipple retraction (UNR) [19,20], breast compliance evaluation (BCE) [21], breast contour length difference, breast area difference and breast overlap difference. In addition, the breast image was divided into 12 fractions by 30 degrees each. Color and scar assessments were conducted simultaneously for comparison between the left and right breasts. All the items were automatically conducted by the software. Eventually, 4-staged evaluation results may be obtained with results similar to that of Harris. This assessment is based on frontal breast photographs, is easy to use, and is independent from raters. However, results are influenced by the conditions during the photo session (i.e., poses, lights). In addition, since it is a planar analysis of the frontal photograph of the breasts, 3-dimensional forms may not be measured precisely.

Observer Assessment

One hundred photographs out of 190 cases (approximately 25 cases each from 4 evaluation categories) were selected from the primary evaluation to ensure the equal distribution of cases in the 4-staged scale of evaluations scale. In each of the 25 cases, 15 cases had evaluation

results by BCCT.core that were different from the 2 other methods, while the remaining 10 cases had evaluation results which were the same in all 3 methods. In cases with “poor” ratings however, at least 1 of the all 3 methods were mainly selected because of the small number of cases. The evaluations were carried out by a total of 6 physicians: 5 expert breast surgeons who conduct BCS routinely and 1 plastic surgeon with experience in breast reconstruction. The 6 raters initially assessed the 100 cases according to the Harris method. Then, they evaluated the 100 cases using the modified Sawai group in order to determine the types of elements that are required for a cosmetic evaluation. Modified parts included the following 7 items: breast size, breast shape, scar, NAC size/shape, NAC color tone, nipple position, and the position of the maximum descent point). Each item was scored between 0 to 5 points. In order to create a consensus for the raters with Harris as a whole, the first results was checked with a Delphi method, feedback was given and investigated. The relationship between the total consensus of the raters and the total scores of the 7 items from the modified Sawai group was investigated using of a correlation coefficient.

Statistical Analysis

During the primary evaluation, the degree of coincidence of the 3 evaluations methods were compared using the κ and w_k statistics in IBM® SPSS® Statistics version22 (IBM® Corporation, New York, United States). The degree of coincidence of κ and w_k is considered as follows: 0 as “poor”, 0.01 to 0.20 as “slight,” 0.21 to 0.40 as “fair,” 0.41 to 0.60 as “moderate,” 0.61 to 0.80 as “substantial,” 0.81 to 0.99 as “almost perfect,” and 1.00 as “perfect” [22]. In the observer assessment, the correlation was evaluated using Spearman’s rank-correlation coefficient. The relationships of values are shown as follows; <0.2 as “none,” 0.2 to 0.4 as “weak”, 0.4 to 0.7 as “moderate,” and >0.7 as “strong.”

Results

Primary Evaluation

Using the Harris method, 44 cases were evaluated as “excellent,” 105 cases as “good,” 24 cases as “fair” and 17 cases as “poor,” respectively. In the Sawai group, 49 cases were evaluated as “excellent,” 95 cases as “good,” 29 cases as “fair” and 17 cases as “poor”. In the BCCT.core, 31 cases were evaluated as “excellent,” 102 cases as “good,” 50 cases as “fair” and 7 cases as “poor”. Of these, 72 cases (37.9%) obtained the same results in the 3 evaluation methods while another 3 cases (1.6%) obtained 3 different results in all methods. For the level of coincidence, $\kappa(w_k)$ was 0.096(0.025)(slight) in the BCCT.core vs. Harris, 0.128(0.013)(slight) in the

BCCT.core vs. Sawai group, and 0.802(0.796)(substantial) in the Harris vs. Sawai group (Table 1) , respectively. There is a significant difference between the BCCT.core as an objective evaluation and the Harris and Sawai group as subjective evaluations.

The photographs used in the primary evaluations were taken using heterogeneous backlight and background. Cardoso et al. defined $k(wk)$ in BCCT.core under such conditions as 0.43 (0.51) [5, 10]. Based on this result, the degree of agreement rate of BCCT.core was inferior to the Harris and Sawai group.

For the details of each case, 13 cases were “excellent” or “good” in BCCT.core and “fair” or “poor” in other 2 methods. Cases shown in (Figure1) were “excellent” in BCCT.core and “fair” in other 2 methods. As such, the different evaluation results were considered to be caused by the lack of recognition of retraction due to scars in the 12 of 13 cases. On the other hand, 25 cases were “fair” or “poor” in BCCT.core and “excellent” or “good” in other 2 methods. Although the forms are equal in the right and left breasts (Figure2), measured values in BCCT.core were different due to the lack of volume, leading to “fair” or “poor” evaluation results in 15 cases. Some cases may be evaluated as “fair” depending on the degree of difference between the right and left breasts as shown in (Figure 3). However, even for the same “fair” cases, the definition of “fair” becomes different depending on the reason for the evaluation; due to retraction of the breasts, or lack of volume. There were also 10 cases for which evaluation results in BCCT.core were different at least 2 stages from that of the other 2 methods.

Observer Assessment

In the consensus of whole raters obtained by the Delphi method, 27 cases were evaluated as “excellent,” 27 cases as “good,” 26 cases as “fair,” and 20 cases as “poor”. In the BCCT.core, 24 cases were evaluated as “excellent,” 30 cases as “good,” 39 cases as “fair,” and 7 cases as “poor”. The degree of agreement between consensus and the BCCT.core was $k=0.134$ (slight), showing low degree of agreement between the subjective methods and the BCCT.core. In particular, cases evaluated as “fair” and “poor” showed diverse evaluation results. Among cases evaluated as “fair” in the whole consensus, 2 cases were evaluated as “excellent” and 5 cases as “good” in the BCCT.core. Among cases evaluated as “poor” in the whole consensus, only 1 case was evaluated as “excellent” and 4 cases as “good” in BCCT.core.

Scores rated by the 6 raters were calculated (30 scores in total) by items of the modified Sawai group, and the relationship between these scores and evaluation results obtained from the whole

consensus (excellent/good/fair/poor) was shown in graphs (Figure 4A-G). In addition, the correlation coefficient is shown as Spearman's rank-correlation coefficient (Table 2). The shape of breasts showed the strongest correlation coefficient at 0.909 (Figure4B), as indicated by the upward-sloping straight line in the figure. This was followed by the breast size 0.791 (Figure4A) and the position of the maximum descent points 0.758 (Figure4G). These items also showed upward-sloping lines in graphs of whole consensus (from "poor" to "excellent"), showing positive correlation. On the other hand, the scar showed the weakest correlation coefficient at 0.345 (Figure4C), which means that it had almost no influence on the whole consensus. The line in the graph was also flat and showed that the scar had minimal influence on the evaluation results. Moderate correlation coefficient was observed in this order: nipple position, 0.690 (Figure4F); nipple and areola size /shape, 0.647 (Figure4D); and nipple and areola color tone, 0.542 (Figure4E). The cosmetic evaluation is composed in order of breast shape>breast size>position of the maximum descent>nipple position>nipple and areola size/shape>nipple and areola color tone>scar.

Discussion

Cosmetic outcome is an important element as well as a permanent cure for BCS. Although there are no defined criteria for cosmetic evaluation, the Harris evaluation method is one of the widely prevalent methods used for cosmetic evaluation of BCS in actual medical practice. A major concern for this method remains the differences caused by various raters. However, this can be resolved by using several raters for the same case. The BCCT.core uses software and is an objective method that not influenced by raters. In addition to integrated existing objective evaluation methods such as BRA, BCE, LBC and UNR, employing color assessment in order to reflect the result of scar assessment and pigmentation secondary to radiation therapy is noteworthy. However, because of influences of photography conditions and limitations of color assessment, insufficient scar assessment is a concern.

As shown in the result of the primary evaluation, insufficient volume was noted in 15 out of 25 patients who were "fair" or "poor" in the BCCT.core but "excellent" or "good" in the other 2 methods. The reasons were as follows: 1) simultaneous breast cancer surgery for the affected breast and reduction mammoplasty for the non-affected breast is not covered by national health insurance and is uncommon in Japan, and 2) the Japanese often resist surgery for the non-affected breast (difference in the cultural backgrounds of Japan and western countries).

When assessed by BCCT.core, differences between the right and left breasts become significant because the assessment is conducted depending on the measurement values, leading to an assessment of “fair” or lower. In addition, the impact of the deterioration of cosmetic outcome due to the retraction of scars is different from that due to insufficient volume, but it is impossible to distinguish these assessment results. In the observer assessment also revealed that scar was not considered to be an important element on cosmetic evaluation. Breast conserving therapy is a combination of BCS and irradiation of remaining breast. Since irradiation can make scar less noticeable to some extent, scars might be unregarded.

An ideal cosmetic evaluation method should not differ from a subjective one and should not be influenced by the raters. Close examination of the subjective assessment method and the development of a computer-based software to prevent differences among raters is essential. The agreement rates of BCCT.core and BAT to consensus were κ ($\omega\kappa$) = 0.56 (0.64) and 0.39 (0.46), respectively, showing an unnecessarily high agreement rate to subjective evaluation [10]. This was because specific software conducted both the evaluations based on measured values in a large part. They did not integrate and reflect the forms that were based on subjective impressions or the balance of the relationship between breasts and other body parts.

The recent price reduction of 3D scanners may trigger their use in the development of a cosmetic evaluation [4, 5, 23]. Considering their convenience in a routine medical practice, however, 2D assessment with frontal photographs of breasts will continue.

In adapting the BCCT.core, an objective evaluation method, to Asian breasts, a clear difference between the objective and subjective methods was revealed. An ideal cosmetic evaluation method for BCS should be an objective method that is not far from the subjective method and is independent from raters. An examination of the elements that comprise a cosmetic evaluation, showed that those that related to breast configuration composed the most part of the evaluations. In a future study, we would like to continue the development of an ideal cosmetic evaluation method that is simple, similar to subjective evaluations and not influenced by raters.

[Reference]

1. Althuis MD, Dozier JM, Anderson WF, Devesa SS, Brinton LA. Global trends in breast cancer incidence and mortality 1973-1997. *International journal of Epidemiology* 2005;34:405-12
2. Harris JR, Levene MB, Svensson G, Hellman S. Analysis of cosmetic results following primary radiation therapy for stages I and II carcinoma of the breast. *International journal of radiation oncology, biology, physics* 1979;5:257-61.
3. Preuss J, Lester L, Saunders C. BCCT.core - can a computer program be used for the assessment of aesthetic outcome after breast reconstructive surgery? *Breast* 2012;21:597-600.
4. Oliveira HIP, Cardoso JS, Magalhães A, Cardoso MJ. Methods for the Aesthetic Evaluation of Breast Cancer Conservation Treatment- A Technological Review. *Current Medical Imaging Reviews* 2013;9:32-46.
5. Cardoso MJ, Cardoso JS, Vrieling C, Macmillan D, Rainsbury D, Heil J, et al. Recommendations for the aesthetic evaluation of breast cancer conservative treatment. *Breast cancer research and treatment* 2012;135:629-37.
6. Aaronson NK, Bartelink H, van Dongen JA, van Dam FS: Evaluation of breast conserving therapy: clinical, methodological and psychosocial perspectives. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology* 1988;14:133-40.
7. Sawai K, Nakajima H, Ichihara S, Yano K, Watanabe O, Kitamura K, et al. Research of cosmetic evaluation and extent of resection for breast conserving surgery. 12th Annual Meeting of the Japanese Breast Cancer Society 2004;12. Abstract #107-8.
8. Pezner RD, Patterson MP, Hill LR, Vora N, Desai KR, Archambeau JO, et al. Breast retraction assessment: an objective evaluation of cosmetic results of patients treated conservatively for breast cancer. *International Journal of Radiation Oncology Biology Physics* 1985;11:575-8.
9. Fitzal F, Krois W, Trischler H, Wutzel L, Riedl O, Kuhbelbock U, et al. The use of a breast symmetry index for objective evaluation of breast cosmesis. *Breast* 2007;16:429-35.
10. 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 research and treatment* 2009;116:149-52.

11. Fitzal F. Analysing breast cosmesis : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology 2009; 35:222.
12. Heil J, Dahlkamp J, Golatta M, Rom J, Domschke C, Rauch G, et al. Aesthetics in breast conserving therapy: do objectively measured results match patients' evaluations? Annals of surgical oncology 2011;18:134-8.
13. Oliveira HP, Magalhaes A, Cardoso MJ, Cardoso JS. An accurate and interpretable model for BCCT.core. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference 2010;6158-61.
14. Cardoso MJ, Magalhaes A, Almeida T, Costa S, Vrieling C, Christie D, et al. Is face-only photographic view enough for the aesthetic evaluation of breast cancer conservative treatment? Breast cancer research and treatment 2008;112:565-8.
15. Cardoso MJ, Cardoso J, Amaral N, Azevedo I, Barreau L, Bernardo M, et al. Turning subjective into objective: the BCCT.core software for evaluation of cosmetic results in breast cancer conservative treatment. Breast 2007;16:456-61.
16. Cardoso JS, Cardoso MJ. Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. Artificial intelligence in medicine 2007;40:115-26.
17. 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;21:61-5.
18. Hau E, Browne LH, Khanna S, Cail S, Cert G, Chin Y, et al. Radiotherapy breast boost with reduced whole-breast dose is associated with improved cosmesis: the results of a comprehensive assessment from the St. George and Wollongong randomized breast boost trial. International journal of radiation oncology, biology, physics 2012;82:682-9.
19. Van Limbergen E, Rijnders A, van der Schueren E, Lerut T, Christiaens R. 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. Radiotherapy and oncology. journal of the European Society for Therapeutic Radiology and Oncology 1989;16:253-67.
20. Van Limbergen E, van der Schueren E, Van Tongelen K. Cosmetic evaluation of breast conserving treatment for mammary cancer. 1. Proposal of a quantitative scoring system. Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology

- and Oncology 1989;16:159-67.
21. Tsouskas LI, Fentiman IS. Breast compliance: a new method for evaluation of cosmetic outcome after conservative treatment of early breast cancer. *Breast cancer research and treatment* 1990;15:185-90.
 22. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. *Family medicine* 2005;37:360-3.
 23. Eder M, Waldenfels FV, Swobodnik A, Kloppel M, Pape AK, Schuster T, et al.. Objective breast symmetry evaluation using 3-D surface imaging. *Breast* 2012;21:152-8.

[Figure legends]

【Figure1】 BCCT.core evaluated better than the others

This case was evaluated as Excellent in BCCT.core and Fair or Poor in 2 other methods.

【Figure2】 BCCT.core evaluated worse than the others

This case was evaluated as Fair in BCCT.core and Excellent or Good in 2 other methods.

【Figure3】 A no retraction case evaluated as “fair”

This case was evaluated as Fair in BCCT.core and Fair or Poor in 2 other methods.

【Figure4】

Relationship between scores rated by items of modified Sawai group and evaluation results obtained from the consensus

【Figure4A】

Relationship between breast size and consensus

【Figure4B】

Relationship between breast shape and consensus

【Figure4C】

Relationship between scar and consensus

【Figure4D】

Relationship between nipple and areola size/shape and consensus

【Figure4E】

Relationship between nipple and areola color tone and consensus

【Figure4F】

Relationship between nipple position and consensus

【Figure4G】

Relationship between position of the maximum descent point and consensus

[Table footnotes]

【Table1】 Result of Primary Evaluation

*BCCT.core = Breast Cancer Conservation Treatment cosmetic results

【Table2】 Spearman's rank correlation coefficient of Consensus vs. 7 Sawai group items
(n=100)

Table1. Result of Primary Evaluation

Comparisons of 2 methods	Kappa	Weighted kappa	Error rate
BCCT.core* vs Harris	0.096	0.025	0.059
BCCT.core* vs Sawai group	0.128	0.013	0.051
Harris vs Sawai group	0.802	0.796	0.038

*BCCT.core = Breast Cancer Conservation Treatment cosmetic results

Table2. Spearman's rank correlation coefficient of Consensus vs. 7 Sawai group items (n=100)

7 Sawai group items	Correlation coefficient	<i>p</i> -value
Breast size	0.791	<0.001
Breast shape	0.909	<0.001
Scar	0.345	<0.001
Nipple and areola size/shape	0.647	<0.001
Nipple and areola color tone	0.542	<0.001
Nipple position	0.690	<0.001
Position of the maximum descent	0.758	<0.001



[Figure1]

This case was evaluated as Excellent in BCCT.core and Fair or Poor in 2 other methods.



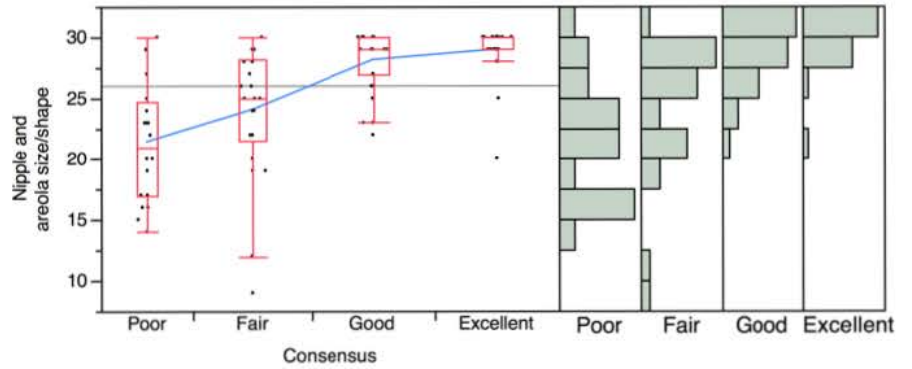
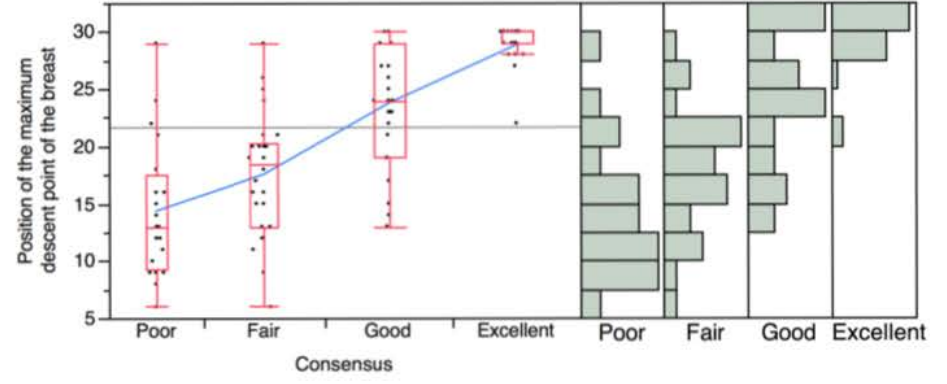
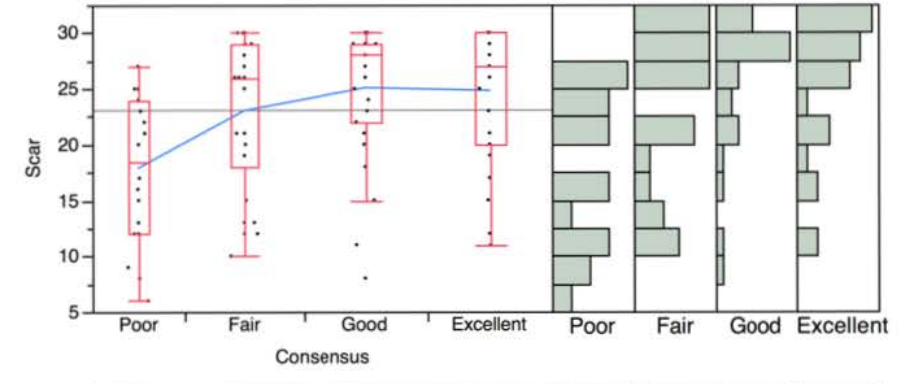
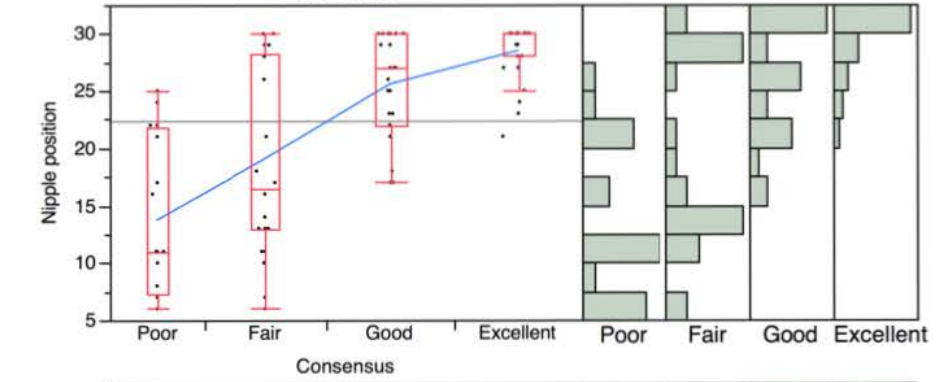
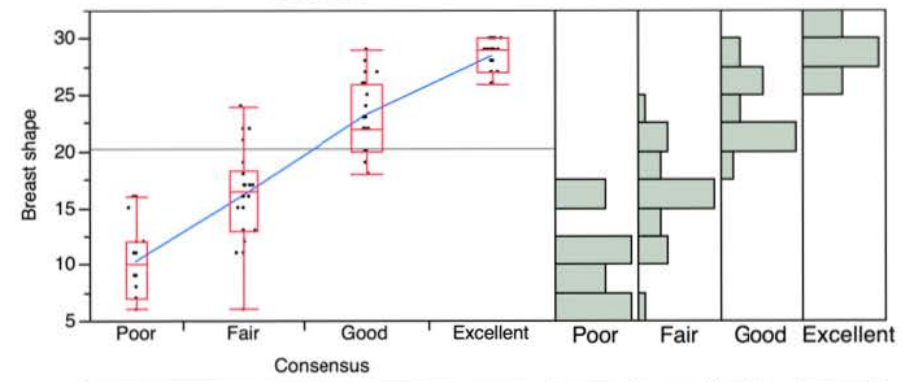
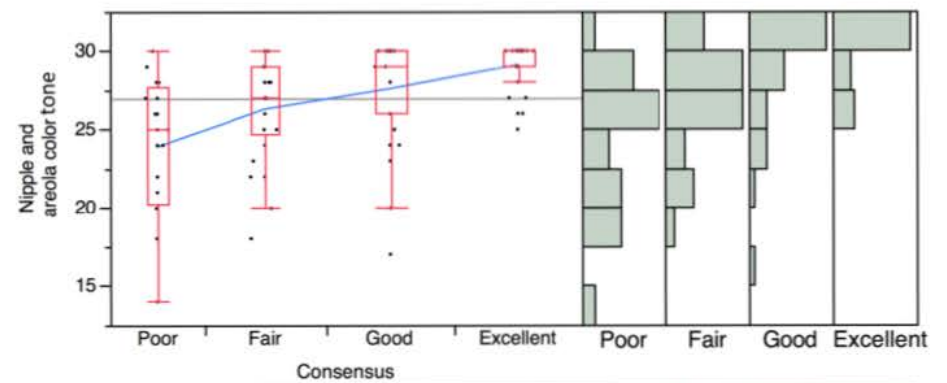
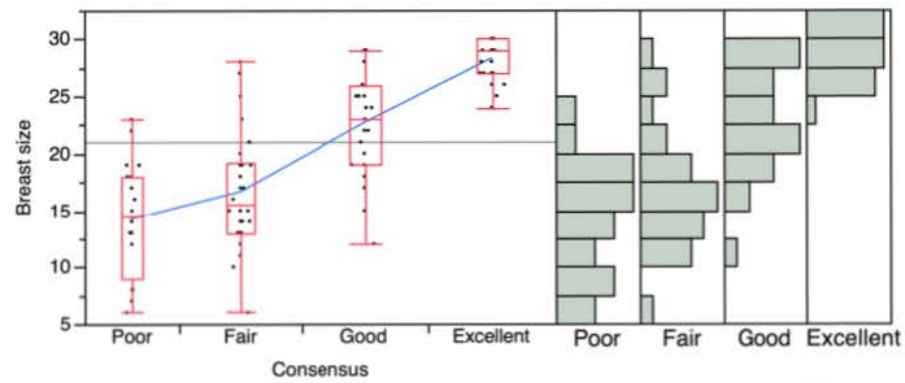
[Figure2]

This case evaluated as Fair in BCCT.core and EG in 2 other methods.



[Figure3]

This case evaluated as Fair in BCCT.core and FP in 2 other methods.



A	E
B	F
C	G
D	

[Figure4]
 Relationship between scores rated by items of modified Sawai group and evaluation results obtained from the consensus