Published in final edited form as:

J Eval Clin Pract. 2009 February ; 15(1): 20–31. doi:10.1111/j.1365-2753.2008.00945.x.

Toward Quantifying the Aesthetic Outcomes of Breast Cancer Treatment: Comparison of Clinical Photography and Colorimetry

Min Soon Kim¹, William N. Rodney¹, Tara Cooper¹, Chris Kite², Gregory P. Reece³, and Mia K. Markey¹

¹ The University of Texas Department of Biomedical Engineering, Austin

² Department of Computer Science, The University of Texas at Austin

³ Department of Plastic Surgery, The University of Texas M. D. Anderson Cancer Center

Abstract

Rationale, aims and objectives—Scarring is a significant cause of dissatisfaction for women who undergo breast surgery. Scar tissue may be clinically distinguished from normal skin by aberrant color, rough surface texture, increased thickness (hypertrophy), and firmness. Colorimeters or spectrophotometers can be used to quantitatively assess scar color, but they require direct patient interaction and can cost thousands of dollars. By comparison, digital photography is already in widespread use to document clinical outcomes and requires less patient interaction. Thus, assessment of scar coloration by digital photography is an attractive alternative. The goal of this study was to compare color measurements obtained by digital photography and colorimetry.

Method—Agreement between photographic and colorimetric measurements of color were evaluated. Experimental conditions were controlled by performing measurements on artificial scars created by a makeup artist. The colorimetric measurements of the artificial scars were compared to those reported in the literature for real scars in order to confirm the validity of this approach. We assessed the agreement between the colorimetric and photographic measurements of color using a hypothesis test for equivalence, the intra-class correlation coefficient (ICC), and the Bland-Altman method.

Results—Overall, good agreement was obtained for three parameters ($L^*a^*b^*$) measured by colorimetry and photography from the results of three statistical analyses.

Conclusion—Color measurements obtained by digital photography were equivalent to those obtained using colorimetry. Thus, digital photography is a reliable, cost-effective measurement method of skin color and should be further investigated for quantitative analysis of surgical outcomes.

Keywords

Aesthetics; Breast Neoplasm; Clinical Photography; Reconstructive Surgical Procedures; Surgical Scars; Treatment Outcome

INTRODUCTION

Breast cancer is the most common type of cancer among American women, with one in eight women diagnosed with breast cancer during her lifetime. It is predicted that in 2007, there will be 176,295 new cases diagnosed (1). Improved screening and treatment methods have increased the breast cancer survival rate. Thus, there is an increased emphasis on the quality of life that women experience after breast cancer treatment.

Mastectomy followed by breast reconstruction is a common form of cancer treatment for the patient with early staged breast cancer or who has contraindications to radiotherapy (2,3). Reconstruction can help breast cancer survivors regain their quality of life by restoring physical balance and helping them carry out their daily lives without being constantly reminded that they are cancer survivors. The goal of reconstructive plastic surgery of the breast is to recreate a natural appearance that is satisfying to the patient. The American Society of Plastic Surgeons (ASPS) estimates that approximately 56,000 breast reconstructive surgeries were performed in the US in 2006 (4).

However, breast surgery, including breast reconstruction, leaves scars that can be disfiguring, aesthetically unpleasant, and can negatively affect a patient's quality of life (5–7). Scars can cause itching, tenderness, pain, sleep disturbances, anxiety, depression, and disruption of daily activities. Other psychosocial complications include development of post-traumatic stress reactions, loss of self-esteem, and stigmatization (5,6).

Assessment of aesthetic outcomes, including scarring, is important to protect patients' health and well-being (8), but there is no accepted method for assessing the appearance of surgical scars. Methods have been proposed to measure the mechanical aspects of scars (9,10). However, while mechanical properties of scars influence the comfort of a patient, they do not impact the visibility, and thus the aesthetics, of scars. The visibility of a scar is significantly influenced by surface area, vascularity, and pigmentation (11).

Since the size and color of scars have the largest influences on the overall visibility, quantitative ways of obtaining these data have been of great interest (9,11–13). Vascularity and pigmentation can be separately evaluated by observing the color of the scar before and after it is blanched using a tool such as a piece of clear plastic (9,11). However, for purposes of quantifying the visibility of scars, it is not necessary to distinguish between coloration due to vascularity vs. pigmentation (14). A cost-effective, practical means of quantitatively assessing scar color is needed. Digital photography has high potential as a tool to achieve this aim.

Color assessment

To standardize the quantification of color, CIE (International Commission on Illumination) recommended tri-stimulus color values XYZ, which are read using spectrometric reflectance data from an object and three primary stimuli strictly defined by the commission. CIE $L^*a^*b^*$ (CIELAB) is the most complete color model used conventionally to describe all the colors visible to the human eye. The three parameters in the model represent the lightness of the color (L^* , $L^* = 0$ indicates black and $L^* = 100$ indicates white), its position between magenta and green (a^* , $a^* = -60$ indicates green and $a^* = 60$ indicates magenta), and its position between yellow and blue (b^* , $b^* = -60$ indicates blue and $b^* = 60$ indicates yellow) (15).

While for purposes of assessing scar visibility it is not necessary to separate the color contributions from the underlying physiological factors of vascularity and pigmentation, for completeness we briefly review how the color parameters are known to relate to the skin

physiology. Studies using the simulated skin model have shown that L^* and melanin index correlate almost linearly if the Hb (oxyhemoglobin) amount is constant. (12,16). These findings suggest that L^* , b^* , or combinations of them could be used with less accuracy for evaluation of pigmentation since the cutaneous blood volume and flow is not similar at all sites. The a^* index is a good parameter for evaluating the degree of erythema or cutaneous blood volume. The difference values between the test and control sites are the best parameters for evaluating erythema and pigmentation (16). These findings suggest that a combination of L^* and b^* can be used for measuring pigmentation and a^* can be used for evaluating the redness of surgical scars. Although all of the $L^*a^*b^*$ parameters are important indicators of vascular changes or pigmentation, the a^* parameter is the most relevant and sensitive one for erythemous or vasodilatory change (15).

The drawback to the use of colorimeters is that they require direct patient interaction and can cost thousands of dollars. By comparison, digital photography is already in widespread use to document clinical outcomes. Moreover, digital photographs can be evaluated after the fact and so required less patient interaction. Thus, assessment of scar coloration by digital photography is an attractive alternative.

Towards this aim, it is encouraging that digital photography has been demonstrated to be a viable alternative to color measurements by a colorimeter for objective evaluation of UV erythema. Coelho *et al.* used the change in the a^* parameter to quantify erythema, and L was used to evaluate the level of pigmentation that develops during tanning for three subjects in each of three ethnic/racial groups: Asian, Black or African American, and White (17). They reported their computer assisted digital image evaluation (CADIE) system, which uses a commercial digital SLR camera, a Nikon D1 (Nikon, Tokyo, Japan), and image processing software, Adobe Photoshop 6.0 (Adobe Systems, San Jose, CA), produced results comparable to those obtained with a spectroscopy system, the Minolta CM 2002 (Minolta, Ramsey, New Jersey). They found good agreement between the a^* (Pearson $r = 0.77-0.97$) and L^* (Pearson $r = 0.66-0.96$) generated by the two methods.

Careful experimental control and standardization is needed in order to collect precise color information. Past studies found that an artificial skin model could be successfully created to control the experimental environment and to demonstrate the diversity of color of skin regions (18,19). In particular, a multilayered skin model was developed to measure the amount of melanin and hemoglobin with a microscope interfaced with a computer and the values were compared with those of real skin regions. Similarly, in this study, artificial scars created by a makeup artist were employed in order to achieve maximal control of the experimental set up, i.e., there was no hemoglobin in the makeup used to create the artificial scar.

Overview of paper

In this study, we compared color measurements of the artificial scars obtained by clinical digital photography to those from a standard colorimeter. Experimental conditions were controlled by performing measurements on artificial scars created by a makeup artist (TC). The colorimetric measurements of the artificial scars were compared to those reported in the literature for real scars in order to confirm the validity of this approach. Agreement in photographic and colorimetric measurements was statistically analyzed using the hypothesis test for equivalence (20), the ICC coefficient (21), and the Bland-Altman method (22,23).

MATERIALS AND METHODS

Artificial Scar Creation on Models

Two female models in their early twenties representing two major race groups (Model A: White and not Latino or Hispanic, Model B: Black or African American and not Latino or Hispanic) were hired for the study. Their demographic and clinical background information is summarized in Table 1. Neither of the two models has a history of pregnancy or lactation, but Model B has a history of a breast biopsy on her left side approximately 7 years ago.

A makeup artist (TC) created 6 types of scars while taking into consideration the reality of the scar appearance from a distance, excessive shine, and the utilization of latex or collodion makeup. An experienced plastic surgeon (GPR) guided the creation of artificial scars that emulate the appearance of scars that would result after breast reconstruction with a latissimus dorsi (LD) flap or a transverse rectus abdominis myocutaneous (TRAM) flap. In LD reconstruction, the flap is composed of an elliptical patch of skin that lies over the latissimus dorsi muscle. Typically, the flap is used to replace the breast skin removed during the mastectomy and the breast volume is replaced with a breast implant (24). The TRAM is the most popular form of tissue-based breast reconstruction, wherein a flap from the abdomen, either with blood vessels intact or completely severed from the body, is transferred to the chest either via a subcutaneous tunnel or using microsurgical technique to replace the breast skin and volume (25–29). Note that artificial donor site scars were not created in this study.

The scars created for Model A resemble normal scars (Figure 1, Figure 2,) and hypertrophic scars (Figure 3), raised scars that remain within the boundaries of the original lesion, as typically seen 3 to 6 months after surgery. Scar A1 was created on left breast around the nipple-areola complex (NAC) with the incision running into the low superior lateral area. Scar A2 was started from Scar A1 and the scar was expanded around the NAC into the medial region of the left breast. Liquid latex was added on Scar A2 to make the nipple more asymmetric in photographs of Scar A3. The scars created for Model B resemble normal scars (Figure 4), hypertrophic scars (Figure 5), and a keloid scar (Figure 6), a scar that has spread beyond the margins of the original wound and invades the surrounding normal skin (8). The scars on Model B resemble those typically seen 6 months to 1 year after surgery. Scar B1 was created on left breast around the nipple-areola complex (NAC) with incision running into the low superior lateral area. Scar B2 was started from Scar B1 and crescent shaped region to lateral side of areola was added. Scar B3 was built off from Scar B2 and added rigid collodion to create a keloid appearance.

The basic scar shape was initially created using liquid latex, rigid collodion, or a combination of these products on top of alcohol cleansed skin in order to get the desired effect. The simulated scars were then covered with a natural skin-toned base foundation to help blend the edges of the created scars with the bare skin. Top detailing color makeup in various tones was next meticulously placed over the simulated scars to give them a more organic appearance. Neutral set powder was pressed over the simulated scars to finalize the makeup and to reduce shine. For Model A, the following top detailing colors were used to create all three artificial scars (Mehron Inc., Chestnut Ridge, New York): Mehron ProColorRing Bruise “Bloody Rose”, Mehron ProColorRing Bruise “Midnight Sky”, Mehron ProColorRing Bruise “Burnt Maroon”. For Model B, Mehron ProColorRing Bruise “Midnight Sky” was used to create scar B1 (Figure 4) and scar B2 (Figure 5) whereas Mehron ProColorRing Bruise “Midnight Sky” and Mehron ProColorRing Bruise “Burnt Maroon” was used to create scar B3 (Figure 6).

Color Measurements using a Colorimeter

One of the widely used opto-electric tri-stimulus instruments commercially available for measuring skin colors in dermatology area is the Chroma-Meter (Konica Minolta, Osaka, Japan) (30–32). Previous studies of the Chroma-Meter CR300 on normal skin have reported excellent intra-observer agreement ($ICC \geq 0.98$) and good to excellent inter-observer agreement ($ICC \geq 0.85$) for all of the three parameters ($L^*a^*b^*$) (31).

A Chroma-Meter CR300 was used to record color information of specific regions of interest defined by a plastic surgeon (GPR). Before measurements were taken, the instrument was calibrated to a standard white plate provided by the manufacturer. One measurement consisted of three consecutive flashes of illumination to obtain a mean value. Before any scar was created, color was measured between the breasts, on each breast mound, on the tan line (if applicable), in each nipple region, and of any natural scars for retrospective analyses. Only the skin color measured between the breasts were grouped and the mean value was taken as the control value for normal skin to which the color values of artificial scars and natural scars were compared.

For each artificial scar, color was measured on the scar region and multiple measurements were performed if the scar varied substantially in width. For accuracy and consistency of measurement, calibration was performed before a scar was created and each time a new scar was created. A total of seven calibrations were used for the measurements. Past studies have shown that there are significant differences between the color parameters of hypertrophic scar regions (aged 6–12 months) due to burns, scalds, trauma and surgical incisions and normal skin regions, including upper arm, forearm, hand, chest, abdomen, back, and thigh (33). The L^* and b^* values of normal skin collected using a spectro-colorimeter, Lab Scan XE (Hunter Associates Laboratory, Inc. Reston, VA), were significantly higher than those of hypertrophic scars, while measurements of a^* of hypertrophic scars were significantly higher than those of normal skin. The authors did not report data on variability across racial groups in these trends of normal vs. scar measurements. To verify that our artificial scars also exhibit these properties, we analyzed the differences between normal skin regions and hypertrophic scars in CIE $L^*a^*b^*$.

Color Measurements using a Camera

30 photographs, including before and after the artificial scars were applied, in sets of five standard views (AP, left oblique, left lateral, right oblique, Right lateral), were taken. A Nikon 8400 was employed for all scar photographs because it is the model most commonly used for record keeping at The University of Texas M. D. Anderson Cancer Center. The main light source was overhead fluorescent lighting, and the camera was set to white balance, fluorescent lighting with ISO of 400. The aperture priority mode was set to the smallest value with spot metering for exposure level selection. Image quality was set to highest level possible to catch the details of the scars. The model stood in front of a dark sky blue background wall. Distance from subject to lens barrel was standardized to 150 cm.

For this study, scar regions of interest (ROIs) were delineated using the polygon selection tool of ImageJ (Rasband, W.S., ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, <http://rsb.info.nih.gov/ij/>, 1997–2007) by looking at the images at 400% zoom ratio on the computer screen with consistent resolution and size (1490×900). The RGB values were collected from each image for both the scar, and normal skin regions (between breasts, breast mounds). The reference values of “C” white used for the computations were provided by the Chroma-Meter CR300 instruction manual. The CIE $L^*a^*b^*$ data of scar regions from photography were compared with those of colorimeter.

Statistical Analysis

The agreement between color measurements obtained by colorimetry and digital photography was analyzed separately for each model. Three different statistical methods were used because each method has its technical limitations (34). The methods used were: a hypothesis test for equivalence (20), the ICC (21), and the Bland-Altman methods (22,23). The Bland-Altman method provides a qualitative assessment while the hypothesis test for equivalence and ICC methods provide quantitative measures of agreement between measurements.

In testing for equivalence between two measurements, the null hypothesis is that the measurements by photography are not equivalent to those by colorimeter and the alternative hypothesis is that they are equivalent. This hypothesis is specifically intended for assessing equivalence and is distinct from that of the familiar paired t -test wherein a null hypothesis of equal value and an alternative hypothesis of unequal value are compared. The test statistic for assessing equivalence is

$$t = \frac{\sqrt{n}(\bar{x} \pm \delta)}{s}$$

where $\bar{x}$ and s denote the mean and standard deviation, respectively, of the differences between the photographic and colorimetric measurements. The value of δ is computed as a factor multiplied by the mean of the measurements by colorimeter. In this study, the factor was 0.20, 0.25, or 0.30. The variable δ accounts for the expected variability in the measurements made by two methods. A smaller value of δ implies stricter criteria for demonstrating that the measurements of the two methods are equivalent, in other words, photography can be used in place of colorimetry.

To report the agreement between two measurements, a single-score two way random model was used since the set of images is a random subset of images from the class of representative cases of post-operative images and the camera and colorimeter are also randomly selected from the population of cameras and colorimeters.

$$ICC = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where n denotes the number of scar measurements by photography and colorimeter and, k denotes the number of measurement methods, MS_R is the mean square error, MS_E is the residual mean square error, and MS_C is the mean square error between photographic and colorimeter measurements, respectively. The guideline for the interpretation of ICC used is that an ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility (35).

Bland-Altman method is used for qualitatively measuring the degree of agreement, where the differences between photographic and colorimeter measurements are plotted against the average value of these measurements. If 95% of the differences are within ± 1.96 standard deviations of the mean of the differences, then the two methods are considered to be in good agreement.

RESULTS

Colorimetric measurements (CIE $L^*a^*b^*$) of the artificial scars and normal skin were evaluated for characteristic differences (Table 2, Table 3) anticipated based on prior knowledge. Recall that the $L^*a^*b^*$ parameters indicate lightness, redness, and yellowness, respectively. The normal skin regions of model B had significantly higher a^* and b^* values than the normal skin regions of model A, while model A had a significantly higher L^* value. This result is consistent with expectations since model B has a darker skin tone than Model A (36).

A similar pattern was seen when color measurements of the artificial scar regions were compared between the two models. Recall that the scars created for Model A resemble normal scars (Figure 1, Figure 2) and hypertrophic (Figure 3) scars typically seen 3 to 6 months after surgery. The scars created for Model B resemble normal scars (Figure 4), hypertrophic scars (Figure 5), and keloid scar (Figure 6), typically seen 6 months to 1 year after surgery. The artificial scars of model B had substantially higher a^* and b^* values than the artificial scars of model A, while model A had a substantially higher L^* value. This result is consistent with our expectations from clinical experience that scar colors vary across racial groups, but we are unaware of any systematic studies that have evaluated this trend. There were no notable differences between the L^* and a^* values of Model B's natural scar and those of the artificial scars created upon her skin. However, the artificial scars had lower b^* values (less yellow) than her natural scar. We are unaware of any prior studies reporting how colorimeter measurements of scars change as scars mature. Thus, it is difficult to compare the measurements of the artificial scars, which were designed to emulate scars 6–12 months following surgery, to the subject's real scar, which was several years old.

Table 2 shows the colorimetric measurements of artificial scars created on Model A, and their comparison with those of her normal skin. Likewise, Table 3 shows the colorimetric measurements of artificial scars and normal skin of Model B. Artificial scars on model A had very slightly lower L^* values and substantially higher a^* values than her normal skin. Similarly, artificial scars on model B also had slightly lower L^* values and slightly higher a^* values than her normal skin. In other words, as expected, the artificial scar regions look darker and redder than normal skin regions, as is the case for real scars. A previous study reported similar trends in the L^* and a^* measurements of real hypertrophic scars (6–12 months after skin injury) as compared with normal skin regions (33). In contrast, they reported that scars are less yellow (lower b^* value) than normal skin (33). However, there was no notable difference in the b^* values of Model A's artificial scars and her normal skin or between the artificial scars and normal skin of Model B. Note that the authors (33) did not report data on variability across racial groups in these trends of normal vs. scar measurements

The $L^*a^*b^*$ values of the artificial keloid scar (scar B3) were significantly lower than those of the artificial hypertrophic scars (scars B1 and B2). In other words, the artificial keloid scar was darker, less red, and less yellow than the artificial hypertrophic scars. This result is consistent with our expectations since hypertrophic scars should be lighter due to less melanin, redder due to the increase in vascularity, and less yellow. However, we are unaware of any studies in which colorimeter measurements of keloid and hypertrophic scars were compared.

The agreement between photographic and colorimetric measurements was analyzed using a hypothesis test for equivalence, the Bland-Altman method, and ICC. Agreement was assessed separately for each model and was based on measurements of three artificial scars per model.

Using a hypothesis test for equivalence, agreement between the photographic and colorimetric measurements was demonstrated for both models at the 20% level for the a^* ($p = 0.03$, $p = 0.03$) and b^* ($p = 0.01$, $p = 0.01$) parameters (Table 4). Equivalence was also achieved at the 20% level for L^* in Model B ($p = 0.04$), but was not achieved even at the 30% level for the L^* value for Model A ($p = 0.08$).

The degree of agreement between the measurements was also analyzed using the Bland-Altman method. Good agreement between the photographic and colorimetric measurements was obtained for all three parameters for both Models A and B as indicated by the fact that 95% of the differences were within the limit of agreement defined by Bland-Altman (Figure 7).

Similarly, the degree of agreement between the measurements was assessed using the ICC (Table 5). ICC results were consistent with the hypothesis test for equivalence and Bland-Altman in that there was “excellent” agreement between photographic and colorimetric measurements for the L^* value of Model A (ICC = 0.99) and “fair to good agreement” for the L^* value (ICC = 0.56) for Model B. In contrast to the other statistical assessments, ICC indicated unacceptable agreement in the a^* and b^* values (0.05–0.34).

DISCUSSION

While colorimeters have been widely accepted as reliable tools for the assessment of skin color in Dermatology, they have not been routinely utilized in the Plastic and Reconstructive Surgery. However, digital photography is commonly used to document surgical outcomes. Thus, it would be valuable if digital photography could also be used to quantitatively evaluate skin color. In order to assess the validity of color measurements obtained by clinical photography, we compared the measurements of a digital camera and a colorimeter. In order to conduct controlled experiments, artificial scars created by a makeup artist (TC) on professional models under the guidance of an experienced plastic surgeon (GPR) were used in this study. Comparisons between normal skin regions and artificial scars in CIE $L^*a^*b^*$ were consistent with expectations based on prior experience and previous studies (33). Using multiple statistical methods, we demonstrated that color measurements obtained by digital photography were equivalent to those obtained using a colorimeter. Differences across the statistical methods may be due to the limited sample size ($N=3$ for each scar groups) we employed in this study. This study suggests that clinical photography can be employed for reliable, cost-effective measurement methods of skin color for the analysis of surgical outcomes, but future, larger analyses employing real scars are needed.

While our results support the use of clinical photography for quantitative measurements of skin properties such as scarring, it is important to note that standardization of photographic conditions is critical. Substantial variability may be observed if the photographic conditions are not controlled. Important parameters to standardize include: lighting intensity, distance and direction from the light to the subject, camera lighting settings, background color, and the subject's pose relative to the camera. While studio lighting is typically recommended for standardization of photography (37), in practice, this may not be feasible since it requires special equipment and training and might not be available to all clinicians, and it also requires enough room with lights placed a standard distance from each other, the patient, and the background. Fortunately, reliable images can be collected without studio conditions provided consistent settings are used. In this study, variability was controlled by avoiding the use of any built-in or external flashes and by simply setting the camera to white balance under fluorescent lamp, an option available on any commercial digital camera. This setting was also found to be advantageous in that it prevented a “washed-out” appearance of scars and surrounding skin colors.

Consistent with our previous study that employed manual identification of fiducial points on clinical photographs (38), we found that non-clinical observers could delineate regions of interest surrounding scars if a minimal amount of basic training was provided. However, a more automated method would be better for large outcome studies or for developing intra-operative tools since it took about 2 minutes to identify and delineate a region of interests.

In conclusion, we found that photography can be used in place of a colorimeter for measuring color properties of skin in a controlled setting. We also demonstrated that artificial scars created by a makeup artist under the guidance of a surgeon exhibit color properties consistent with real scars.

Acknowledgments

This work was supported in part by grant number 1 R21 CA109040-01A1 from the National Cancer Institute.

References

1. American Cancer Society. Cancer Facts and Figures 2007. 2007.
2. Hunt, KK.; Robb, GL.; Strom, EA. Breast Cancer. New York: Springer; 2001.
3. Bold RJ. Surgical management of breast cancer: today and tomorrow. Cancer Biotherapy & Radiopharmaceuticals. 2002; 17(1):1–9. [PubMed: 11915165]
4. American Society of Plastic Surgeons. 2000/2005/2006 National Plastic Surgery Statistics. 2007.
5. Bayat A, McGrouther DA, Ferguson MWJ. Skin scarring. BMJ. 2003; 326(7380):88–92. [PubMed: 12521975]
6. Atiyeh BS, Ioannovich J, Al-Amm CA, El-Musa KA, Dham R. Improving scar quality: a prospective clinical study. Aesthetic Plastic Surgery. 2002; 26(6):470–6. [PubMed: 12621572]
7. Godwin Y. A comparison of the patient and surgeon opinion on the long-term aesthetic outcome of reduction mammoplasty. British Journal of Plastic Surgery. 1998; 51:444–9. [PubMed: 9849364]
8. Kim, MS.; Sbalchiero, JC.; Reece, GP.; Miller, MJ.; Beahm, EK.; Markey, MK. Assessment of Breast Aesthetics. Plastic & Reconstructive Surgery; 2007. in press
9. Baryza MJ, Baryza GA. The Vancouver Scar Scale: an administration tool and its interrater reliability. Journal of Burn Care & Rehabilitation. 1995; 16(5):535–8. [PubMed: 8537427]
10. Truong PT, Abnoui F, Yong CM, Hayashi A, Runkel JA, Phillips T, Olivotto IA. Standardized assessment of breast cancer surgical scars integrating the Vancouver Scar Scale, Short-Form McGill Pain Questionnaire, and patients' perspectives. Plastic & Reconstructive Surgery. 2005; 116(5):1291–9. [PubMed: 16217470]
11. van de Kar AL, Corion LU, JSM, Draaijers LJ, van der Horst CM, van Zuijlen PP. Reliable and feasible evaluation of linear scars by the Patient and Observer Scar Assessment Scale. Plastic & Reconstructive Surgery. 2005; 116(2):514–22. [PubMed: 16079683]
12. Draaijers LJ, Tempelman FRH, Botman YAM, Kreis RW, Middelkoop E, van Zuijlen PPM. Colour evaluation in scars: tristimulus colorimeter, narrow-band simple reflectance meter or subjective evaluation? Burns: Journal Of The International Society For Burn Injuries. 2004; 30(2): 103–7. [PubMed: 15019115]
13. Draaijers LJ, Tempelman FRH, Botman YAM, Tuinebreijer WE, Middelkoop E, Kreis RW, van Zuijlen PPM. The Patient and Observer Scar Assessment Scale: A Reliable and Feasible Tool for Scar Evaluation. Plastic & Reconstructive Surgery. 2004; 113(7):1960–5. [PubMed: 15253184]
14. Beausang E, Floyd H, Dunn KW, Orton CI, Ferguson MWJ. A New Quantitative Scale for Clinical Scar Assessment. Plastic & Reconstructive Surgery. 1998; 102(6):1954–61. [PubMed: 9810991]
15. Andreassi L, Flori L. Practical applications of cutaneous colorimetry. Clinics In Dermatology. 1995; 13(4):369–73. [PubMed: 8665445]
16. Takiwaki H. Measurement of skin color: practical application and theoretical considerations. The Journal Of Medical Investigation. 1998; 44(3–4):121–6. [PubMed: 9597799]

17. Coelho SG, Miller SA, Zmudzka BZ, Beer JZ. Quantification of UV-induced erythema and pigmentation using computer-assisted digital image evaluation. *Photochemistry and Photobiology*. 2006; 82(3):651–5. [PubMed: 16522135]
18. Takiwaki H, Kanno Y, Miyaoka Y, Arase S. Computer simulation of skin color based on a multilayered skin model. *skin research And Technology*. 1997; 3:36–41.
19. Wan S, Anderson RR, Parrish JA. Analytical modeling for the optical properties of the skin with in vitro and in vivo applications. *Photochemistry and Photobiology*. 1981; 34(4):493–9. [PubMed: 7312955]
20. Wellek, S. Testing statistical hypothesis of equivalence. Boca Raton, FL: CRC Press LLC; 2003.
21. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996; 1(1):30–46.
22. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986; 1(8476):307–10. [PubMed: 2868172]
23. Bland JM, Altman DG. Applying the right statistics: analyses of measurement studies. *Ultrasound in obstetrics & gynecology*. 2003; 22(1):85–93. [PubMed: 12858311]
24. Kroll, SS. The well-informed patient's guide to breast reconstruction. Houston: The University of Texas M. D. Anderson Cancer Center; 2002.
25. Moran SL, Serletti JM. Outcome comparison between free and pedicled TRAM flap breast reconstruction in the obese patient. *Plastic & Reconstructive Surgery*. 2001; 108(7):1954–60. [PubMed: 11743383]
26. Serletti JM, Moran SL. Free versus the pedicled TRAM flap: a cost comparison and outcome analysis. *Plastic & Reconstructive Surgery*. 1997; 100(6):1425–7.
27. Shestak KC. Breast reconstruction with a pedicled TRAM flap. *Clinics In Plastic Surgery*. 1998; 25(2):167–82. [PubMed: 9627774]
28. Veiga DF, SabinoNeto M, Ferreira LM, Garcia EB, VeigaFilho J, Novo NF, Rocha JL. Quality of life outcomes after pedicled TRAM flap delayed breast reconstruction. *British Journal of Plastic Surgery*. 2004; 57(3):252–7. [PubMed: 15006527]
29. Walgenbach KJ, Shestak KC. Pedicled TRAM Breast Reconstruction. *Breast Disease*. 2002; 16:73–7. [PubMed: 15687659]
30. Serup J, Agner T. Colorimetric quantification of erythema--a comparison of two colorimeters (Lange Micro Color and Minolta Chroma Meter CR-200) with a clinical scoring scheme and laser-Doppler flowmetry. *Clinical And Experimental Dermatology*. 1990; 15(4):267–72. [PubMed: 2208776]
31. Van den Kerckhove E, Staes F, Flour M, Stappaerts K, Boeckx W. Reproducibility of repeated measurements on healthy skin with Minolta Chromameter CR-300. *Skin Research and Technology*. 2001; 7(1):56–9. [PubMed: 11301642]
32. Takiwaki H, Overgaard L, Serup J. Comparison of narrow-band reflectance spectrophotometric and tristimulus colorimetric measurements of skin color Twenty-three anatomical sites evaluated by the Deraspectrometer and the Chroma Meter CR-200. *Skin Pharmacology*. 1994; 7(4):217–25. [PubMed: 8024803]
33. Li-Tsang CW, Lau JC, Liu SK. Validation of an objective scar pigmentation measurement by using a spectrophotometer. *Burns*. 2003; 29(8):779–84. [PubMed: 14636751]
34. Luiz RR, Szklo M. More than one statistical strategy to assess agreement of quantitative measurements may usefully be reported. *Journal of Clinical Epidemiology*. 2005; 58(3):215–6. [PubMed: 15718108]
35. Rosner, B. Fundamentals of Biostatistics. 6. Belmont, CA: Duxbury Press; 2005.
36. Shriver MD, Parra EJ. Comparison of narrow-band reflectance spectroscopy and tristimulus colorimetry for measurements of skin and hair color in persons of different biological ancestry. *Am J Phys Anthropol*. 2000; 112(1):17–27. [PubMed: 10766940]
37. Yavuzer R, Smirnes S, Jackson IT. Guidelines for standard photography in plastic surgery. *Plastic & Reconstructive Surgery*. 2001; 46(3):293–300.
38. Kim MS, Reece GP, Miller MJ, Beahm E, Atkinson EN, Markey MK. Objective assessment of aesthetic outcomes of breast cancer treatment: measuring ptosis from clinical photographs. *Computers in Biology and Medicine*. 2007; 37:49–59. [PubMed: 16438948]

39. Fitzpatrick TB. Soleil et peau. *Journal of Medical Esthetics*. 1975; 2:33–4.
40. Bostwick, J, III. *Plastic and reconstructive breast surgery*. 2. St. Louis, Missouri: Quality Medical Publishing, Inc; 1999.
41. Regnault P. Breast ptosis: definition and treatment. *Clinics In Plastic Surgery*. 1976; 3(2):193–203. [PubMed: 1261176]

Figure 1.

2D image of Scar A1 on Model A. The scars created for Model A resemble normal scars that are typically seen 3 to 6 months after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scar A1 was created on left breast around the nipple-areola complex (NAC) with incision running into the low superior lateral area. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these.

Figure 2.

2D image of scar A2 on Model A. The scars created for Model A resemble normal scars that are typically seen 3 to 6 months after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. The scar was expanded around the nipple area into the medial region of the breast. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these.

Figure 3.

2D image of scar A3 on Model A. The scars created for Model A resemble hypertrophic scars that are typically seen 3 to 6 months after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these. Liquid latex was added to make nipple more asymmetric on Scar A3.

Figure 4. 2D image of scar B1 on Model B. The scars created for Model B resemble normal scars that are typically seen 6 months to 1 year after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scar B1 was created on left breast around the nipple-areola complex (NAC) with incision running into the low superior lateral area. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these.

Figure 5.

2D image of scar B2 on Model B. The scars created for Model B resemble hypertrophic scars that are typically seen 6 months to 1 year after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these. Scar B2 was started from Scar B1 and crescent shaped region to lateral side of areola was added.

Figure 6.

2D image of scar B3 on Model B. The scars created for Model B resemble keloid scars that are typically seen 6 months to 1 year after breast reconstruction surgery with either LD or TRAM flap. No donor site scar was created in this study. Scars were created by adding base color and top color makeup on the base of either latex or rigid collodion or combination of these. Scar B3 was built off from Scar B2 and added rigid collodion to create a keloid appearance

Figure 7.

Bland-Altman analysis for the agreement of scar groups A (a, b, c) and B (d, e, f) between the measurements of three artificial scars ($N = 3$) for each of the three parameters that were measured by colorimeter and digital camera. The parameters measured were: (a) L value, (b) a* value, and (c) b* value. The results show that good agreement was obtained for all three parameters measured as indicated by the fact that 95% of the differences were within the limit of agreement defined by Bland-Altman.

Table 1

Clinical and demographic background information of Model A and Model B. None of two models had history of pregnancy or lactation, but Model B has a history of breast surgery on left side about 7 years ago, which might affect the aesthetic assessment of the breast.

	Model A	Model B
Age (in years)	21	22
Weight (in kg)	57	64
Height (in cm)	170	173
Bra size	36	36
Cup size (A, B, C, D)	B	C
Ethnicity	Not Hispanic or Latino	Not Hispanic or Latino
Race	White	Black or African American
Tanning booth exposure	Have had blister burns. Not more than 4 hours of sun per week	Never tan in a booth Sometimes sunbath outside no more than an hour, with sun block though. SPF 50.
Sunscreen use	No routine use of sunscreen	Always. And waterproof.
Pregnancy and lactation history	None	None
Estrogen use	Oral contraceptive	Oral contraceptive
Previous breast surgery	None	Had a benign tumor in left breast removed at 15
Fitzpatrick Skin Type (FST) (39)	II. Burn and minimal tan, Pale white	VI, Tan, no burn, Dark brown
Notes on FST	Mainly burns, but little tan	Have never had a sun burn
Ptois grade (40,41)	0: Nipple and most of gland are above IMF	2: Nipple is below IMF, but above lower contour of the breast
Pseudoptosis	No	No

Table 2

Comparison between normal skin region and scar region of Model A in L*a*b* color space. L*a*b* parameters indicate lightness, redness, and yellowness, respectively. Group A scars showed slightly lower L* value and significantly higher a* value than normal skin. This result shows that artificial scar regions look darker and redder than normal skin regions, while there was no significant difference in b* value, which indicates no prominent yellowness between artificial scars and normal skin.

	Scar A1 (normal)	Scar A2 (normal)	Scar A3 (hypertrophic)	Art. Scar Mean	Normal skin Mean
L*	74.75	53.72	70.33	66.27	67.00
a*	5.59	9.70	8.92	8.07	3.28
b*	6.00	17.78	13.72	12.50	11.12

Comparison between normal skin region and scar region of Model A in L*a*b* color space. L*a*b* parameters indicate lightness, redness, and yellowness, respectively. There were no notable difference observed in L* and a* values between group B artificial scars and natural scar, except significantly and consistently lower b* in group B. There were no notable differences in all three L*a*b* values between natural scar and normal skin, because natural scar was 7 years aged according to Model B's history, and it was stabilized enough and didn't stand out against surrounding normal skin.

Table 3

	Scar B1 (Normal)	Scar B2 (Hypertrophic)	Scar B3 (Keloid)	Art. Scar Mean	Natural scar Mean	Normal skin Mean
L*	41.02	64.07	34.43	46.51	46.74	49.28
a*	10.02	14.36	6.37	10.25	11.85	9.74
b*	10.55	10.84	6.63	9.34	17.89	19.68

Results of the hypothesis test for equivalence of scar group B for three CIE $L^*a^*b^*$ parameter values of six artificial scars ($N = 3$) as measured by a camera and a colorimeter. Equivalence was achieved at 20% variability in all three color parameters except in group A, where equivalence was not achieved even at 30% in L^* value. The null hypothesis was that the photographic and colorimeter color measurements are not equivalent. $\delta = \text{factor} \times \text{mean}$ of colorimeter measurement. A p -value less than 0.05 indicates that the color measurements obtained by photography and colorimeter are equivalent.

Table 4

Factor	L*		a*		b*	
	Group A	Group B	Group A	Group B	Group A	Group B
0.30	$p = 0.08$	$p = 0.02$	$p = 0.03$	$p = 0.03$	$p = 0.01$	$p = 0.01$
0.25	$p = 0.21$	$p = 0.02$	$p = 0.03$	$p = 0.03$	$p = 0.01$	$p = 0.01$
0.20	$p = 0.22$	$p = 0.04$	$p = 0.03$	$p = 0.03$	$p = 0.01$	$p = 0.01$

Table 5

Intra-class correlation (ICC) coefficient of group A and B for three CIE L*a*b* parameter values of six artificial scars as measured by a colorimeter and a camera. Mixed agreement was shown that group A (N = 3) showed “excellent” agreement for the L* value (ICC = 0.99), and “fair to good agreement” in group B for the L* value (ICC = 0.56), while Both group A and B showed unacceptable agreement in a* and b* values (0.05–0.34). The guideline for the interpretation of ICC used is that an ICC value of less than 0.40 indicates poor reproducibility, ICC values in the range 0.40 to 0.75 indicate fair to good reproducibility, and an ICC value of greater than 0.75 shows excellent reproducibility.

	L* value	a* value	b* value
Group A (N = 3)	ICC = 0.56	ICC = 0.34	ICC = 0.16
Group B (N = 3)	ICC = 0.99	ICC = 0.29	ICC = 0.05