



ARTICLE



Evaluation of an assessment scale for aesthetic outcome in breast reconstructions based on digital photos in both 2D and 3D format

Linda Tallroth^{a,b} , Nathalie Mobargha^c, Patrik Velander^{a,b}, Stina Klasson^{a,b} and Magnus Becker^{a,b}

^aDepartment of Plastic and Reconstructive Surgery, Skåne University Hospital, Malmö, Sweden; ^bDepartment of Clinical Sciences in Malmö, Lund University, Sweden; ^cDepartment of Reconstructive Plastic Surgery, Karolinska University Hospital, Stockholm, Sweden

ABSTRACT

The aesthetic outcome is crucial in a breast reconstruction. Our aim was to evaluate the intra- and interrater reliability of an aesthetic outcome assessment scale with digital photos of breast reconstructions in two-dimensional (2D) and three-dimensional (3D) format. Thirty-three women with delayed breast reconstructions, consecutively participating in a five-year follow-up between November 2019 and June 2021, were included in the study. Of these, 14 were reconstructed with an expander prosthesis (EP) and 19 with a deep inferior epigastric perforator (DIEP) flap. Photos of the breasts were assessed in 2D and 3D format by expert, layman and patient panels. Data were analysed with the weighted kappa (wk) statistics. The intrarater agreements were moderate to substantial, with wk between 0.66 and 0.73 for the panels. Within the panels, the interrater agreements were 0.46–0.62. Moderate agreements were found between the matched 2D and 3D format photos (wk 0.62–0.66). The patient panel graded scar appearance worse in 3D compared with 2D format. In all panels, there was a tendency towards DIEP flap reconstructions receiving higher aesthetic outcome grades compared with EP. Thus, the aesthetic outcome assessment scale demonstrated acceptable agreements between the individual panellists and within the panels. Scars captured in 3D format may provide a greater resemblance to the reality compared with 2D. Implications for clinics remain to be further studied.

ARTICLE HISTORY

Received 6 April 2022
Revised 5 September 2022
Accepted 24 November 2022

KEYWORDS

Aesthetics; expander prosthesis; DIEP flap

Background

In Sweden in 2020, 2405 women with breast cancer and no remote metastases underwent mastectomy [1]. Mastectomy has a negative impact on women's body image and quality of life and thus, a breast reconstruction is offered to mitigate these effects [2]. A breast reconstruction may be implant-based, created from autologous tissue, or potentially a combination of the two. Individual patient characteristics and patient preferences will guide the choice of breast reconstruction method and influence the result. A satisfactory aesthetic result together with a good functional outcome are essential in a breast reconstruction. Yet there is no agreement on how best to evaluate the aesthetic outcome.

The aesthetic outcome after a breast reconstruction is often evaluated with photos using an assessment scale. A variety of assessment scales have been reported [3–9]. The most common assessment scale used for professional assessment has been a four-point scale [10]. In previous reports, the number and size of panels recruited to evaluate the aesthetic outcome have differed and the agreements have in many cases been poor or have not been addressed [5–7,11]. In addition, measurement of aesthetic outcome with panel assessment has received criticism as it is time-consuming [12]. However, subjectivity is crucial for evaluating outcomes in plastic surgery as it may provide information that is not explored by objective measures. To this day, there is no gold standard assessment scale for evaluation of breast reconstructions. In a review article from 2015, the scale reported by

Visser et al. was considered the most preferable as it demonstrated high validity [4,10]. It was, however, limited by a wide range of intra- and interrater agreements [4]. An assessment scale that is reliable between assessors with similar experiences and can identify differences over time is desirable. Therefore, the main aim of this study was to evaluate the reliability of an assessment scale for aesthetic outcome in breast reconstructions. A secondary aim was to compare the aesthetic outcome following expander prosthesis (EP) breast reconstructions with deep inferior epigastric perforator (DIEP) flap breast reconstructions.

Material and methods

Patients

Thirty-four consecutive patients who had undergone unilateral delayed breast reconstruction between October 2012 and November 2016 were selected for participation in this study. The patients had been randomised to breast reconstruction with either an EP or a DIEP flap, and participated in a prospective five-year follow-up [13]. The study was approved by the Ethical Review Board in Sweden Dnr 2012/187 and Dnr 2021-00555.

Photo session

Photography of the patients was performed by a professional medical photographer or by the first author, in a hospital photo studio with standardised lightning. A two-dimensional (2D)

CONTACT Linda Tallroth  linda.tallroth@med.lu.se  Jan Waldenströms gata 18, 214 28 Malmö, Sweden

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

camera and a three-dimensional (3D) camera were used for documentation of the breasts. The photos in 2D format were taken with a single-lens reflex digital camera (Nikon Corporation, Tokyo, Japan). The lens used during photography was a NIKKOR lens with constant f/2.8 aperture and focal length of 24–70 mm (Nikon Corporation, Tokyo, Japan). The 3D camera system, 3dMD trio system (3dMD LLC, Atlanta, GA), had 12 fixed cameras. Of these 12 cameras, four were mounted frontally and four on both sides. Prior to each 3D photo session, the 3dMD trio system was calibrated. Subsequently, photos were taken from three angles, resulting in a photo possible to be viewed as a 3D photo in the 3dMD Vultus (version 2.2.024) program. The program enables the viewer to rotate the 3D photo, zoom in on details and conduct measurements, but was not used by the panellists in this study.

Aesthetic outcome assessment scale

The assessment scale used in this study was a modification of the scale reported by Visser et al. [4]. The five items—breast size, shape, symmetry, scar appearance and nipple areolar complex (NAC)—were graded using a five-point Likert scale. The five-point Likert scale ranged from very bad (1) to very good (5). The item size was graded from much smaller [1], equal [3] to much larger [5], compared with the non-reconstructed breast, different from the scale by Visser et al. The option “Cannot be evaluated” was added as a modification in the absence of a NAC. The overall aesthetic outcome was assessed using a ten-point Likert scale, very bad (1) to very good (10).

Panels

Three types of panellists were recruited for this study. Plastic surgeons and breast surgeons participated as experts. Only consultant and senior consultant physicians were invited. Laymen with varying degrees of medical knowledge were invited to join the layman panels. Twelve patients were invited to join a patient panel. Their participation included assessment of their own breast reconstruction.

Data collection

The study data were managed and collected using Research Electronic Data capture (REDCap) tools hosted at Lund University [14,15]. REDCap is a web-based platform which we used to facilitate the photo assessments. The study was performed in two phases. In the first phase, 48 sets of photos accompanied by the assessment scale were included. The same breast reconstruction appeared on two sets of photos. There were four photos per set in 2D format and five per set in 3D format (Figure 1(A–D)). The sets were arranged in a randomised order. Laterality was noted but not reconstruction type. To facilitate a high response rate, the assessments could be completed at any time. The panellists were not informed in advance that the same reconstruction appeared twice, nor that two different camera modalities were used. All panellists were asked to perform the assessment twice, a minimum of three weeks apart. They were also asked to record the time it took to perform the assessment. An expert and a layman panel assessed the photos in the study's first phase and a reliability analysis was conducted. In the second phase of the study, all breast reconstructions apart from two were replaced by new breast reconstructions. Twelve breast reconstructions were included, and in total there were 24 sets of photos. An expert, a layman and a patient panel assessed the breast reconstructions in

the second phase. The assessment was performed twice by the patients and once by the other panels.

Statistical analysis

Statistical Package for Social Sciences version 27 (IBM Corp. Armonk, NY: IBM Corp. Released 2020) was used for statistical analysis. Intrarater and interrater agreements were calculated with the weighted kappa (w_k) and were presented as median, minimum and maximum values. The w_k was used for a reliability analysis of the assessment of digital photos in 2D format with the corresponding assessments in 3D format. Interrater reliability was presented as the median of the individually pairwise calculated kappa values. Level of agreement was interpreted as poor below 0.00, slight 0.00–0.20, fair 0.21–0.40, moderate 0.41–0.60, substantial 0.61–0.80, and almost perfect 0.81–1.00 [16]. A p -value below 0.05 was considered statistically significant.

Results

Patient characteristics

Thirty-four patients completed the photography as a part of the prospective follow-up. One patient was excluded as she had undergone a contralateral breast reconstruction due to breast cancer. Thus, 33 patients were included. The photo sessions were performed between November 2019 and June 2021 at a mean of 66 (standard deviation, SD 11) months after breast reconstruction. The median age at breast reconstruction was 55 (SD 10) years. Of the included patients, 14 breasts were reconstructed with an EP and 19 with a DIEP flap.

Panel characteristics

Eleven plastic surgeons and two breast surgeons participated in the expert panel in the first phase of the study. Of these, eight were men and five were women. The age within the panel ranged from 38 to 68 years. Eleven of them performed the assessment twice. In the second phase, none of the four expert panellists had been involved in the care of the patients. The layman panel comprised of nine panellists of which four were men and five were women. Their ages ranged from 20 to 58 years. One member of the panel was a senior consultant physician working within a non-surgical specialty, two were intern physicians, and one was a medical student. The other laymen did not have any previous medical knowledge.

The median time for performing the assessment in the first phase was 60 (40–120) min for the experts and 50 (35–120) min for the laymen.

Reliability analysis

Distribution on the Likert scale

The distributions of the panels' gradings per item are shown in Tables 1–2. The expert and layman assessments are from the first phase of the study and the patient panel assessment from the second phase. Grades 4 and 5 were the most frequent for symmetry; however, grade 2 was the most common grade in the layman panel assessment of photos in 3D format. Regarding scar appearance, grade 4 was the most frequent grade in the expert and layman panels. However, photos in 2D format were most frequently assessed as grade 5 and photos in 3D format as grade 2 by the patient panel.

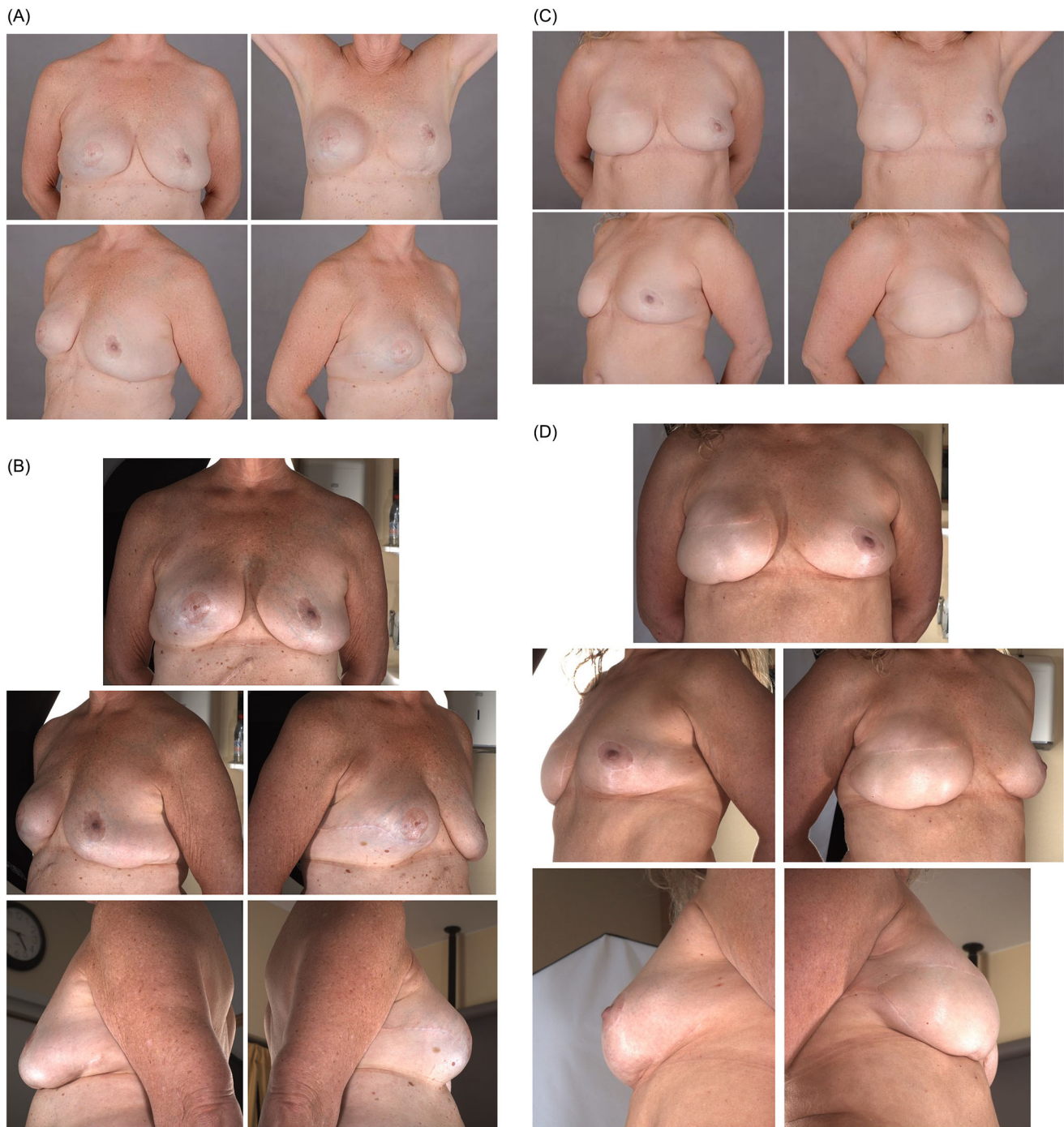


Figure 1. Examples of postoperative photographs evaluated in the study. (A) An expander prosthesis (EP) breast reconstruction in two-dimensional (2D) format and in (B) three-dimensional (3D) format. (C) A deep inferior epigastric perforator (DIEP) flap breast reconstruction in 2D format and in (D) 3D format.

Reliability for repeated assessments

In the expert panel, the intrarater agreements were moderate to substantial with a median κ of 0.70 (0.62–0.75) for photos in 2D format and 0.67 (0.54–0.80) for photos in 3D format. In the layman panel, the agreements were moderate to substantial. The median κ was 0.70 (0.58–0.74) and 0.66 (0.60–0.69) for the respective photo format. The patient panel had a median κ of 0.73 (0.50–0.89) and 0.72 (0.29–0.83) respectively, assessed in the second phase of the study. The intrarater agreements are summarised in [Table 3](#).

Reliability for assessment within panels

The interrater agreements were moderate in the expert and the layman panels. In the expert panel ($n=13$), the median κ was 0.60 (0.36–0.74) for photos in 2D format and 0.55 (0.35–0.77) for photos in 3D format. In the layman panel ($n=9$), the assessments resulted in a median κ of 0.62 (0.44–0.73) and 0.57 (0.38–0.67) for photos in 2D and 3D format respectively. The agreements in the patient panel ($n=12$) were somewhat lower with a median κ of 0.46 (0.19–0.73) for photos in 2D format and 0.48 (0.04–0.73) for photos in 3D format. The interrater agreements are presented in [Table 4](#).

Table 1. Frequency distribution of panels' grading from the panels' first assessments separated by photo format for size, shape, symmetry, scar appearance and nipple areolar complex.

Item	n	Likert scale grades					Total n (%)
		n (%)					
		1	2	3	4	5	
Size							
<i>2 D format</i>							
Expert	13	10 (4.3)	62 (27.0)	103 (44.8)	53 (23.0)	1 (0.4)	229 (99.6)
Layman	9	19 (9.2)	49 (23.7)	105 (50.7)	33 (15.9)	1 (0.5)	207 (100)
Patient	12	12 (8.3)	42 (29.2)	69 (47.9)	15 (10.4)	4 (2.8)	142 (98.6)
<i>3 D format</i>							
Expert	13	12 (5.2)	66 (28.7)	96 (41.7)	55 (23.9)	1 (0.4)	230 (100)
Layman	9	28 (13.5)	51 (24.6)	92 (44.4)	35 (16.9)	206 (99.5)	207 (100)
Patient	12	14 (9.7)	46 (31.9)	73 (50.7)	9 (6.3)	2 (1.4)	144 (100)
Shape							
<i>2 D format</i>							
Expert	13	8 (3.5)	44 (19.1)	31 (13.5)	101 (43.9)	46 (20.0)	230 (100)
Layman	9	5 (2.4)	42 (20.3)	26 (12.6)	53 (25.6)	81 (39.1)	207 (100)
Patient	12	9 (6.3)	19 (13.2)	16 (11.1)	57 (39.6)	42 (29.2)	143 (99.3)
<i>3 D format</i>							
Expert	13	15 (6.5)	50 (21.7)	42 (18.3)	92 (40.0)	31 (13.5)	230 (100)
Layman	9	18 (8.7)	45 (21.7)	25 (12.1)	56 (27.1)	63 (30.4)	207 (100)
Patient	12	12 (8.3)	26 (18.1)	16 (11.1)	45 (31.3)	43 (29.9)	142 (98.6)
Symmetry							
<i>2 D format</i>							
Expert	13	10 (4.3)	53 (23.0)	32 (13.9)	91 (39.6)	43 (18.7)	229 (99.6)
Layman	9	7 (3.4)	48 (23.2)	26 (12.6)	49 (23.7)	76 (36.7)	206 (99.5)
Patient	12	9 (6.3)	14 (9.7)	16 (11.1)	51 (35.4)	54 (37.5)	144 (100)
<i>3 D format</i>							
Expert	13	16 (7.0)	59 (25.7)	34 (14.8)	89 (38.7)	30 (13.0)	228 (99.1)
Layman	9	20 (9.7)	56 (27.1)	30 (14.5)	50 (24.2)	51 (24.6)	207 (100)
Patient	12	7 (4.9)	34 (23.6)	13 (9.0)	41 (28.5)	49 (34.0)	144 (100)
Scar appearance							
<i>2 D format</i>							
Expert	13	0	22 (9.6)	45 (19.6)	130 (56.5)	31 (13.5)	228 (99.1)
Layman	9	1 (0.5)	27 (13.0)	46 (22.2)	77 (37.2)	54 (26.1)	205 (99.0)
Patient	12	6 (4.2)	19 (13.2)	16 (11.1)	42 (29.2)	60 (41.7)	143 (99.3)
<i>3 D format</i>							
Expert	13	7 (3.0)	49 (21.3)	60 (26.1)	93 (40.4)	17 (7.4)	226 (98.3)
Layman	9	9 (4.3)	56 (27.1)	53 (25.6)	65 (31.4)	22 (10.6)	205 (99.0)
Patient	12	14 (9.7)	38 (26.4)	20 (13.9)	34 (23.6)	35 (24.3)	141 (97.9)
NAC							
<i>2 D format</i>							
Expert	13	9 (3.9)	40 (17.4)	58 (25.2)	72 (31.3)	20 (8.7)	199 (86.5)
Layman	9	11 (5.3)	38 (18.4)	35 (16.9)	56 (27.1)	35 (16.9)	175 (84.5)
Patient	12	7 (4.9)	24 (16.7)	18 (12.5)	42 (29.2)	35 (24.3)	126 (87.5)
<i>3 D format</i>							
Expert	13	7 (3.0)	51 (22.2)	39 (17.0)	82 (35.7)	19 (8.3)	198 (86.1)
Layman	9	10 (4.8)	33 (15.9)	38 (18.4)	62 (30.0)	26 (12.6)	169 (81.6)
Patient	12	8 (5.6)	25 (17.4)	12 (8.3)	43 (29.9)	41 (28.5)	129 (89.6)

The expert and layman assessments are from the first phase of the study and the patient panel assessment from the second phase. The Likert scale grade with highest frequency per panel and item is in bold.

2 D: Two dimensional; 3 D: Three dimensional; NAC: Nipple areolar complex.

Table 2. Frequency distribution of panels' grading from the panels' first assessments separated by photo format for the overall aesthetic outcome.

Item	n	Likert scale grades										Total n (%)
		n (%)										
		1	2	3	4	5	6	7	8	9	10	
Overall aesthetic outcome												
<i>2 D format</i>												
Expert	13	0	6 (2.6)	16 (7.0)	13 (5.7)	18 (7.8)	23 (10.0)	41 (18.7)	57 (24.8)	41 (17.8)	15 (6.5)	230 (100)
Layman	9	0	4 (1.9)	14 (6.8)	11 (5.3)	18 (8.7)	25 (12.1)	36 (17.4)	42 (20.3)	39 (18.8)	18 (8.7)	207 (100)
Patient	12	5 (3.5)	5 (3.5)	9 (6.3)	7 (4.9)	9 (6.3)	13 (9.0)	18 (12.5)	23 (16.0)	30 (20.8)	25 (17.4)	144 (100)
<i>3 D format</i>												
Expert	13	6 (2.6)	5 (2.2)	27 (11.7)	18 (7.8)	21 (9.1)	32 (13.9)	37 (16.1)	52 (22.6)	23 (10.0)	9 (3.9)	230 (100)
Layman	9	2 (1.0)	10 (4.8)	25 (21.1)	21 (10.1)	17 (8.2)	26 (12.6)	27 (13.0)	36 (17.4)	36 (17.4)	7 (3.4)	207 (100)
Patient	12	6 (4.2)	7 (4.9)	11 (7.6)	8 (5.6)	20 (13.9)	11 (7.6)	9 (6.3)	28 (19.4)	21 (14.6)	23 (16.0)	144 (100)

The expert and layman assessments are from the first phase of the study and the patient panel assessment from the second phase. The Likert scale grade with highest frequency per panel is in bold.

2 D: Two dimensional; 3 D: Three dimensional.

Table 3. Intrarater agreements with weighted kappa (κ_w) values.

Photo format	Expert panel <i>n</i> = 11	Layman panel <i>n</i> = 7	Patient panel <i>n</i> = 12
2 D	0.70 (0.62 – 0.75)	0.70 (0.58 – 0.74)	0.73 (0.50 – 0.89)
3 D	0.67 (0.54 – 0.80)	0.66 (0.60 – 0.69)	0.72 (0.29 – 0.83)

The patient panel's agreements were calculated based on ratings from the second phase of the study. Agreements are presented in median with minimum and maximum values in parenthesis.

Interpretations of κ_w values: 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate 0.61–0.80 substantial, >0.80 almost perfect [17].

2 D: Two dimensional; 3 D: Three dimensional.

Table 4. Interrater agreements with weighted kappa (κ_w) values.

Photo format	Expert panel <i>n</i> = 13	Layman panel <i>n</i> = 9	Patient panel <i>n</i> = 12
2 D	0.60 (0.36 – 0.74)	0.62 (0.44 – 0.73)	0.46 (0.19 – 0.73)
3 D	0.55 (0.35 – 0.77)	0.57 (0.38 – 0.67)	0.48 (0.04 – 0.73)

The patient panel agreements were calculated based on ratings from the second phase of the study. Agreements are presented in median with minimum and maximum values in parenthesis.

Interpretations of κ_w values: 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate 0.61–0.80 substantial, >0.80 almost perfect [17].

2 D: Two dimensional; 3 D: Three dimensional.

Table 5. Intrarater agreements with weighted kappa (κ_w) between assessments in 2D format with the corresponding in 3D format.

Reconstruction method	Expert panel <i>n</i> = 13	Layman panel <i>n</i> = 9	Patient panel <i>n</i> = 12
All	0.64 (0.33 – 0.78)	0.62 (0.46 – 0.73)	0.66 (0.25 – 0.77)
EP	0.55 (0.27 – 0.71)	0.56 (0.34 – 0.67)	0.57 (0.19 – 0.69)
DIEP flap	0.63 (0.32 – 0.83)	0.61 (0.43 – 0.80)	0.63 (0.13 – 0.86)

The patient panel's agreements were calculated based on ratings from the second phase of the study. Agreements are presented in median with minimum and maximum values in parenthesis.

Interpretations of κ_w values: 0.00–0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate 0.61–0.80 substantial, >0.80 almost perfect [17].

EP: Expander prosthesis; DIEP: Deep inferior epigastric perforator; 2D: Two dimensional; 3D: Three dimensional.

Digital photos in 2D and 3D format

Intrarater agreements were calculated on the matched assessments of photos in 2D and 3D format. The median κ_w was moderate in all panels. The median κ_w was 0.64 (0.33 – 0.78) in the expert panel, 0.62 (0.46 – 0.73) in the layman panel, and 0.66 (0.25 – 0.77) in the patient panel. Separated by reconstruction method, the median κ_w values were somewhat higher in assessments of DIEP flaps. The results are presented in Table 5.

Aesthetic outcome

The aesthetic outcome results are presented in Table 6. The results presented are from the second phase of the study. In all panels, there was a general tendency towards higher grades for DIEP flap breast reconstructions compared with EP. The tendency was more pronounced for the overall aesthetic outcome regarding photos in 3D format. In comparison of the panels, laymen gave the lowest median grades for overall aesthetic outcome. An in-depth review of the patients receiving lower overall outcome scores, less than or equal to 6.5, by the expert panels in phase one and two illustrated potential negative factors such as previous prosthesis exchanges (*n* = 3), increased body mass index (BMI) with more than four units (*n* = 1) and early reoperations due to complications (*n* = 2).

Table 6. Aesthetic outcome scores per item and photo format assessed by three panels.

Panel	Aesthetic outcome scores					Overall
	Size	Shape	Symmetry	Scar	NAC	
Expert						
2 D format	3 (2–5)	4 (1–5)	4 (1–5)	4 (2–5)	4 (1–5)	8 (2–10)
EP	3 (2–4)	3.5 (1–5)	3.5 (1–5)	4 (2–5)	3 (1–5)	7 (2–10)
DIEP flap	3 (2–5)	5 (4–5)	5 (4–5)	5 (2–5)	4 (2–5)	9 (7–10)
3 D format						
EP	3 (2–5)	4 (1–5)	4 (1–5)	4 (2–5)	4 (1–5)	7.5 (2–10)
DIEP flap	3 (2–4)	3.5 (1–5)	3.5 (1–5)	3 (2–5)	3 (1–5)	6 (2–10)
Layman						
2 D format	3 (1–4)	4 (1–5)	4 (1–5)	3 (1–5)	4 (1–5)	7 (1–10)
EP	2 (1–4)	4 (1–5)	4 (1–5)	3 (2–5)	3 (1–5)	6 (1–10)
DIEP flap	3 (2–4)	4.5 (3–5)	4 (3–5)	3 (1–5)	4 (3–5)	8 (6–10)
3 D format						
EP	3 (1–4)	4 (1–5)	4 (1–5)	3 (1–5)	4 (1–5)	7 (2–9)
DIEP flap	2 (1–4)	3 (1–5)	3 (1–5)	3 (1–5)	3 (1–5)	5 (2–9)
Patient						
2 D format	3 (1–5)	4 (1–5)	4 (1–5)	4 (1–5)	4 (1–5)	8 (1–10)
EP	2 (1–4)	4 (1–5)	4 (1–5)	4 (1–5)	3 (1–5)	7 (1–10)
DIEP flap	3 (2–5)	4 (2–5)	4 (2–5)	4 (1–5)	4 (1–5)	9 (2–10)
3 D format						
EP	3 (1–5)	4 (1–5)	4 (1–5)	3 (1–5)	4 (1–5)	7.5 (1–10)
DIEP flap	2 (1–4)	3 (1–5)	3 (1–5)	3 (1–5)	3 (1–5)	5 (1–10)
DIEP flap	3 (2–5)	5 (2–5)	5 (2–5)	4 (1–5)	4 (2–5)	8.5 (3–10)

On the 5-point rating scale, 1 equal very bad and 5 very good regarding items shape, symmetry, scar appearance and NAC. For the item size, 1 equal much smaller than the natural breast, 3 the same size as the natural breast and 5 much larger than the natural breast. Overall aesthetic outcome was rated on a 10-point scale on which 1 equal very bad and 10 very good.

EP: Expander prosthesis; DIEP: Deep inferior epigastric perforator; 2D: Two dimensional; 3D: Three dimensional; NAC: Nipple areolar complex.

Discussion

In this study, we report on the reliability of an aesthetic outcome assessment scale used for breast reconstructions. Median agreements were moderate to substantial for repeated assessments in expert, layman and patient panels. Between members of the same panel, somewhat lower median agreements were found, with the lowest values in the patient panel. In a comparison of matched photos in 2D and 3D format, moderate to substantial median agreements were demonstrated.

In the context of breast reconstructions, repeated evaluations are essential to identify changes postoperatively. For example, weight changes and implant disfiguration may alter the aesthetic result over time. Based on the findings from this study, the assessment scale demonstrated acceptable reproducibility. Compared with a study by Veiga et al., the agreements in our report were high. They presented intrarater agreements between 0.12 and 1 for photo evaluations of autologous breast reconstructions at three different time points [17]. The wide agreement range presented could be explained by the scale used. It may be difficult for panellists to distinguish between adjacent grades in the presence of a scale with ten grades. However, our findings concurred with the intrarater agreements reported by Godden et al. ranging from 0.4 to 0.7, using a five-point scale [18].

The use of aesthetic outcome assessment scales has been questioned, partly as a result of the high variability of interrater agreements reported in the literature. In the past, different statistical analysis methods have been used, which complicates comparisons between studies [3,5,19]. Moreover, in some studies, reliability was not analysed [6,7]. Results from this study reflect moderate agreements, similar to some previous studies [5,19]. Lindegren et al. and Gahm et al. used the κ_w and consequently, their results can be compared with ours [5,19]. Meanwhile, Visser et al. and Liu et al. used a different analysis method, the intraclass

correlation, and presented higher agreements [3,4]. We found the lowest agreements in the patient panel, indicating that there may be a heterogeneity within this panel. Plausibly, the patient's own reconstruction experience and result influences the perception of other breasts. Supposedly, a satisfied patient will assess other breast reconstructions more favourably. Although we opt for high agreements, some variability is to be expected as aesthetics are perceived differently between individuals.

The agreements between the matched assessments of photos in 2D and 3D format, together with similar intrarater agreements between the two, suggest a comparable use of the photo formats. A similar result was reported in a study evaluating cleft, lip and palate patients using 2D and 3D photos. The difference between the interrater agreements of the 2D and 3D photos was small, 0.56 and 0.62, respectively [20]. We used a 3D camera system aiming for more realistic and detailed photos compared with the standard digital photos in 2D format. Our hypothesis was that the panels would grade photos in 3D format worse due to greater enhancement of scars and skin surface irregularities. We did not find a general tendency confirming this hypothesis. Interestingly, compared with the photos in 2D format, the photos in 3D format were assessed with lower grades concerning scar appearance by the patient panel. This result may be explained by scar appearance being an important outcome for patients, and therefore assessed more critically. Also, patients may have unrealistic expectations concerning the final scar appearance. This is further supported by the findings in the study by Lindegren et al. in which patients were less satisfied with the DIEP flap donor scar compared with experts [5]. Photos in 3D format may provide a better reflection of the reality. By using photos in 3D format when informing patients preoperatively, more realistic expectations may be achieved. In the process of choosing the reconstruction method it is crucial that the patient is well-informed, with awareness of the possible aesthetic outcomes, as this may increase the postoperative satisfaction. In addition, a future perspective would be to evaluate the reliability between outpatient clinic assessments and 3D photo assessments.

Although a low number of patients were included in this study, the results tended to be in favour of the DIEP flap breast reconstructions due to the better aesthetic outcome. Superior aesthetic outcome in autologous reconstruction compared with implant-based reconstructions has been reported previously [21–23]. The difference in aesthetic outcome between the reconstruction methods may increase with time as autologous reconstructions tend to be stable over time, unlike implants. Thus, other treatments and patient characteristics may influence the aesthetic outcome. Radiation therapy had a negative effect on the overall aesthetic outcome in a study by Huis et al. [6]. In addition, higher BMI and reoperations due to complications have also been reported to negatively affect aesthetics [21]. An in-depth review of our study material supports these results as some of the patients with low overall aesthetic outcome had been through reoperations due to complications, and in one case had a large increase in BMI. However, these associations must be confirmed in a larger body of material.

In concordance with previous studies, we acknowledge that the use of an assessment scale for aesthetic evaluation of breast reconstructions is time-consuming. To facilitate a high number of participating panellists, we used REDcap, which provided a more flexible way to evaluate the photos. Although there is a considerable advantage to using an electronic platform that can be accessed easily, a drawback is the possible influence of external factors. Strengths of this study are that the patients were

randomised to breast reconstruction with an EP or a DIEP flap and that they were included consecutively. The panels were blinded to the reconstruction method and all reconstructions were assessed twice; in 2D and 3D format. The long follow-up time provided evaluation of breast reconstructions that were somewhat stable in their appearances. A weakness of the study is that the reliability analysis for the patient panel was based on different photos to those used with the expert and layman panels. The low number of patients included in this study is another weakness. Moreover, it is important to consider the drawback of not being able to rotate the 3D photos. This feature may have led to different results.

Conclusion

The result from this study suggested that continued use of the assessment scale in breast reconstructions could be recommended. A possible value of assessing scar appearance with photos in 3D format was found. A comparison between clinical assessments in the outpatient clinic and assessment of 3D photos is yet to be performed.

Acknowledgement

The authors are very grateful to the medical photographer Magnus Nilsson and to the panellists for their contributions to this study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

No financial support was received for this study.

ORCID

Linda Tallroth  <http://orcid.org/0000-0002-3469-9361>

References

- [1] Cancer NQRFB. Immediate reconstructions at mastectomy [internet]. Stockholm: National Quality Register for Breast Cancer; 2021. <https://statistik.incanet.se/brostdcancer/>
- [2] Yoon AP, Qi J, Brown DL, et al. Outcomes of immediate versus delayed breast reconstruction: results of a multicenter prospective study. *Breast*. 2018;37:72–79.
- [3] Liu T, Freijs C, Klein HJ, et al. Patients with abdominal-based free flap breast reconstruction a decade after surgery: a comprehensive long-term follow-up study. *J Plast Reconstr Aesthet Surg*. 2018;71(9):1301–1309.
- [4] Visser NJ, Damen THC, Timman R, et al. Surgical results, aesthetic outcome, and patient satisfaction after microsurgical autologous breast reconstruction following failed implant reconstruction. *Plast Reconstr Surg*. 2010;126(1):26–36.
- [5] Lindegren A, Halle M, Docherty Skogh AC, et al. Postmastectomy breast reconstruction in the irradiated breast: a comparative study of DIEP and latissimus dorsi flap outcome. *Plast Reconstr Surg*. 2012;130(1):10–18.
- [6] Huis In 't Veld EA, Long C, Sue GR, et al. Analysis of aesthetic outcomes and patient satisfaction after delayed-

- immediate autologous breast reconstruction. *Ann Plast Surg.* 2018;80(5S Suppl 5):S303–S307.
- [7] Ramon Y, Ullmann Y, Moscona R, et al. Aesthetic results and patient satisfaction with immediate breast reconstruction using tissue expansion: a follow-up study. *Plast Reconstr Surg.* 1997;99(3):686–691.
- [8] Teotia SS, Alford JA, Kadakia Y, et al. Crowdsourced assessment of aesthetic outcomes after breast reconstruction. *Plast Reconstr Surg.* 2021;147(3):570–577.
- [9] Eltahir Y, Bosma E, Teixeira N, et al. Satisfaction with cosmetic outcomes of breast reconstruction: investigations into the correlation between the patients' Breast-Q outcome and the judgment of panels. *JPRAS Open.* 2020;24:60–70.
- [10] Maass SW, Bagher S, Hofer SO, et al. Systematic review: aesthetic assessment of breast reconstruction outcomes by healthcare professionals. *Ann Surg Oncol.* 2015;22(13):4305–4316.
- [11] O'Connell RL, Di Micco R, Khabra K, et al. Comparison of immediate versus delayed DIEP flap reconstruction in women who require postmastectomy radiotherapy. *Plast Reconstr Surg.* 2018;142(3):594–605.
- [12] Dahlback C, Ringberg A, Manjer J. Aesthetic outcome following breast-conserving surgery assessed by three evaluation modalities in relation to health-related quality of life. *Br J Surg.* 2019;106(1):90–99.
- [13] Tallroth L, Velandar P, Klasson S. A short-term comparison of expander prosthesis and DIEP flap in breast reconstructions: a prospective randomized study. *J Plast Reconstr Aesthet Surg.* 2021;74(6):1193–1202.
- [14] Harris PA, Taylor R, Minor BL, REDCap Consortium, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform.* 2019;95:103208.
- [15] Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* 2009;42(2):377–381.
- [16] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–174.
- [17] Veiga DF, Neto MS, Garcia EB, et al. Evaluations of the aesthetic results and patient satisfaction with the late pedicled TRAM flap breast reconstruction. *Ann Plast Surg.* 2002;48(5):515–520.
- [18] Godden AR, Wood SH, James SE, et al. A scoring system for 3D surface images of breast reconstruction developed using the Delphi consensus process. *Eur J Surg Oncol.* 2020;46(9):1580–1587.
- [19] Gahm J, Edsander-Nord A, Jurell G, et al. No differences in aesthetic outcome or patient satisfaction between anatomically shaped and round expandable implants in bilateral breast reconstructions: a randomized study. *Plast Reconstr Surg.* 2010;126(5):1419–1427.
- [20] Mosmuller DGM, Maal TJ, Prah C, et al. Comparison of two- and three-dimensional assessment methods of nasolabial appearance in cleft lip and palate patients: do the assessment methods measure the same outcome? *J Craniomaxillofac Surg.* 2017;45(8):1220–1226.
- [21] Duraes EFR, Schwarz GS, de Sousa JB, et al. Factors influencing the aesthetic outcome and quality of life after breast reconstruction: a cross-sectional study. *Ann Plast Surg.* 2020;84(5):494–506.
- [22] Clough KB, O'Donoghue JM, Fitoussi AD, et al. Prospective evaluation of late cosmetic results following breast reconstruction: I. Implant reconstruction. *Plast Reconstr Surg.* 2001;107(7):1702–1709.
- [23] Clough KB, O'Donoghue JM, Fitoussi AD, et al. Prospective evaluation of late cosmetic results following breast reconstruction: II. Tram flap reconstruction. *Plast Reconstr Surg.* 2001;107(7):1710–1716.