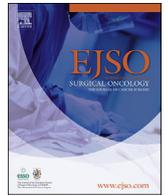




Contents lists available at ScienceDirect

European Journal of Surgical Oncology

journal homepage: www.ejso.com

A scoring system for 3D surface images of breast reconstruction developed using the Delphi consensus process

Amy R. Godden ^{a,b}, Simon H. Wood ^c, Stuart E. James ^d, Fiona A. MacNeill ^d,
Jennifer E. Rusby ^{a,b,*}

^a Royal Marsden Hospital, Downs Road, Sutton, Surrey, SM2 5PT, UK

^b Institute of Cancer Research, 15 Cotswold Road, Sutton, London, SM2 5NG, UK

^c Imperial College NHS Foundation Trust, Fulham Palace Road, W6 8RF, UK

^d Royal Marsden Hospital, Fulham Road, Chelsea, London, SW3 6JJ, UK

ARTICLE INFO

Article history:

Received 11 December 2019

Received in revised form

6 April 2020

Accepted 14 May 2020

Available online xxx

This work was presented at the Association of Breast Surgery Annual Conference, Glasgow 2019 and won the poster presentation prize.

Keywords:

Aesthetic/reconstruction/breast/cancer

ABSTRACT

Introduction: Evaluation of aesthetics after breast reconstruction is challenging. In the absence of an objective measurement, panel assessment is widely adopted. Heterogeneity of scales and poor internal consistency make comparison difficult. Development and validation of an expert panel scale using a Delphi consensus process is described. It was designed specifically for use as the gold standard for development of an objective evaluation tool using 3-Dimensional Surface Imaging (3D-SI).

Materials and methods: 20 items relating to aesthetic assessment were identified for consideration in the Delphi consensus process. Items were selected for inclusion in the definitive panel scale by iterative rounds of voting according to importance, consensus discussion, and a final vote. The Delphi-derived scale was tested on a clinical research series for intra- and inter-panellist, and intra-panel reliability, and correlation with Patient Reported Outcome Measures (PROMs).

Results: 61 surgeons participated in the Delphi process. Oncoplastic and plastic surgeons were represented. The Delphi-derived scale included symmetry, volume, shape, position of breast mound, nipple position, and a global score. Intra-panellist reliability ranged from poor to almost perfect ($\omega_k < 0.086$), inter-rater reliability was fair (ICC range 0.4–0.5) for individual items and good (ICC 0.6) for the global score, intra-panel reliability was moderate to substantial ($\omega_k 0.4–0.7$), and correlation with PROMs was moderate ($r = 0.5p < 0.01$).

Conclusions: The Delphi-derived panel evaluation is at least as good as other scales in the literature and has been developed specifically to provide expert evaluation of aesthetics after breast reconstruction. The logistical constraints of panel assessment remain, reinforcing the need to develop an objective evaluation method.

© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Aesthetic outcome from breast cancer surgery has a well-documented influence on patients' psychosocial wellbeing and quality of life [1–8]. With excellent survival expectations of 90% at

one year and 80% at ten years [9], more women experience the long term impact of treatments, reinforcing aesthetic outcome as an important survivorship focus. The guidance from the UK's National Institute for Health and Care Excellence (NICE) in 2002, that all women undergoing mastectomy in the UK should be offered reconstruction, reflects the acceptance of this in breast cancer surgery [10].

Breast reconstruction is a rapidly developing field in the UK [11], yet it is without a gold standard measure for reporting aesthetic outcome. A review by Potter et al. highlights many inconsistencies between the current methods for assessment of outcome [13].

Patient reported outcome measures (PROMs) are an important reflection of aesthetic satisfaction, however, they do not correlate

Abbreviations: PROMS, patient reported outcome measures; BCT, breast conserving treatment; 3D-SI, 3-dimensional surface imaging; ABS, association of breast surgery; MOT, Medical Outcomes Trust; ICC, intra-class correlation coefficient; ω_k , weighted kappa; NAC, nipple areola complex.

* Corresponding author. Royal Marsden Hospital, Downs Road, Sutton, Surrey, SM2 5PT, UK.

E-mail address: Jennifer.rusby@rmh.nhs.uk (J.E. Rusby).

<https://doi.org/10.1016/j.ejso.2020.05.016>

0748-7983/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

well with professional evaluation of aesthetics [8,14–16] which frequently report more favourable outcomes [17–19]. Qualitative interviews with patients who had undergone breast reconstruction shed some light on factors contributing to patients' decision making. How a patient feels about their reconstruction and how they feel they are perceived by others', the relationship with their surgeon or other practitioners, clinical outcome (i.e. complications) and how they viewed reconstruction as part of their cancer journey were described. A positive patient experience may translate into a favourable opinion of aesthetic outcome as measured by PROMs when a professional judgement may suggest the contrary, thus, PROMs may not be responsive enough to build clinical evidence and develop best practice [20]. The impact of relationships and pathway on patient perception of outcome limits the value of PROMs in evaluation of existing and new techniques in order to innovate.

An objective scoring system would provide a robust method of communication, benchmarking, and comparison of aesthetic outcome, an important asset when considering new surgical techniques and materials, especially within today's context of 'no surgical innovation without evaluation' [21–23]. Three Dimensional Surface Imaging (3D-SI) may represent the best modality for this as it provides quantitative information on projection, volume and surface symmetry in addition to linear mammometrics (distances from surface landmarks, i.e. sternal notch-to-nipple distance).

In order to develop an objective scoring system for 3D-SI, a gold standard is needed against which to perform multivariate analysis to discern how to weight each measured parameter. Panel assessment is the most widely accepted evaluation method, but has inherent bias, is costly, time-consuming, and not standardised. There are multiple panel scales reported in the literature which share common deficiencies, including lack of responsiveness, repeatability, and interpretability [12,14]. Evaluating reconstructive surgery is arguably more challenging than for breast-conserving treatment (BCT) where the overarching goal is to maintain or achieve symmetry. In reconstructive surgery there is a broader spectrum of operative goals which can be tailored to the patient desires and surgical achievability (i.e. the capacity to modify size, shape, projection, cleavage, and the position of breast mound).

In this study, a Delphi consensus process was used to develop a contemporary scale for the panel assessment for breast reconstruction aesthetics. Enabling as a step towards the development of an objective scoring system using measures derived from 3D-SI [24].

Methods

This study was part of a pilot study reviewed by the London-Surrey Research Ethics Committee (17/LO/0763) in preparation for a large multi-centre trial, available at clinicaltrials.gov (NCT03203252).

Delphi consensus process

The purpose was to identify key aesthetic criteria used by clinicians when evaluating aesthetic outcomes. These items were developed into a panel scoring system, ultimately for use as the gold standard in development of an objective score for the evaluation of breast reconstruction using 3D-SI. A systematic review of current scales within the literature by Maass et al. [12] and consultation with experts in the field shaped the online questionnaire for use in the Delphi consensus process. Items referenced in the literature which have been used in previous panel assessments for aesthetic evaluation after reconstructive breast surgery were included in the questionnaire. The questionnaire was circulated to oncoplastic and plastic surgeons via iBRA-net (a community of

research active breast/plastic surgeons in the UK) and the UK Association of Breast Surgery (ABS). Participants were required to rate each item according to their perception of its importance for inclusion in an expert panel assessment for breast reconstruction, from 1 (extremely important) to 9 (unimportant).

The Delphi process consisted of two iterative rounds of voting with predefined exclusion criteria (Fig. 2), followed by two consensus discussions and a final round using binary "in" or "out" voting. Participants must have completed prior rounds in order to participate in subsequent rounds. The rationale for hosting two consensus discussions was pragmatic, to enable maximum participation. One was held in Birmingham at the iBRA-net meeting (October 2018) and the other in London at the Royal Marsden cross-site oncoplastic research meeting (November 2018). Only votes from members who had participated in rounds one and two contributed to the final round, although others contributed to the discussion.

Additional questions posed within the questionnaire that were not part of the Delphi process pertained to the participant (profession, grade, experience [years in post], gender) and the proposed panel methodology (number of panellists, number of points on the Likert scale).

Evaluation of the aesthetic scoring system

The validity, intra- and inter-panellist and intra-panel reliability of the scoring system was tested by a multi-disciplinary panel with representation from two centres in London with high rates of autologous and implant reconstruction (Imperial College Healthcare NHS Trust and The Royal Marsden NHS Foundation Trust). Three oncoplastic breast surgeons, three clinical oncologists, and three plastic surgeons comprised the panel. The panellists were all consultants with at least 5 years' experience in post. The panellists were blinded to patient, surgeon and clinical oncologist identity and the treatment received. Lay people and patients were not included because the aim was to select items which contributed to a better technical outcome, for benchmarking and research purposes.

3D-SIs of 55 women who had undergone mastectomy (unilateral or bilateral) and immediate autologous breast reconstruction with adjuvant or neoadjuvant radiotherapy were viewed in a standardised animation (7 views in sequence). Antero-posterior (AP), oblique (left and right), lateral (left and right), cranial, and caudal views were presented (Fig. 1). 10 3D-SIs were selected at random and repeated within the panel assessment to report upon intra-panellist and intra-panel reliability. No discussion was permitted during the panel assessment. The panel was not shown example images to benchmark standards. Panellists were given written definitions for the Likert scales in advance of the panel (also available during the panel assessment for reference) in order to standardise the scoring.

Three-dimensional surface images

3D-SIs were captured using Vectra® (Canfield, USA). Women were positioned with their hands on their hips with their elbows behind the mid-axillary line to optimise visualisation of the lateral aspect of the breast. The images were taken at the end inspiratory pause during quiet breathing. Images for the adjuvant radiotherapy cohort were collected as part of a previous study [14], and the images for the neoadjuvant group were from the aesthetic subgroup of the Primary Radiotherapy And Diep flap study (PRADA) [NCT02771938].

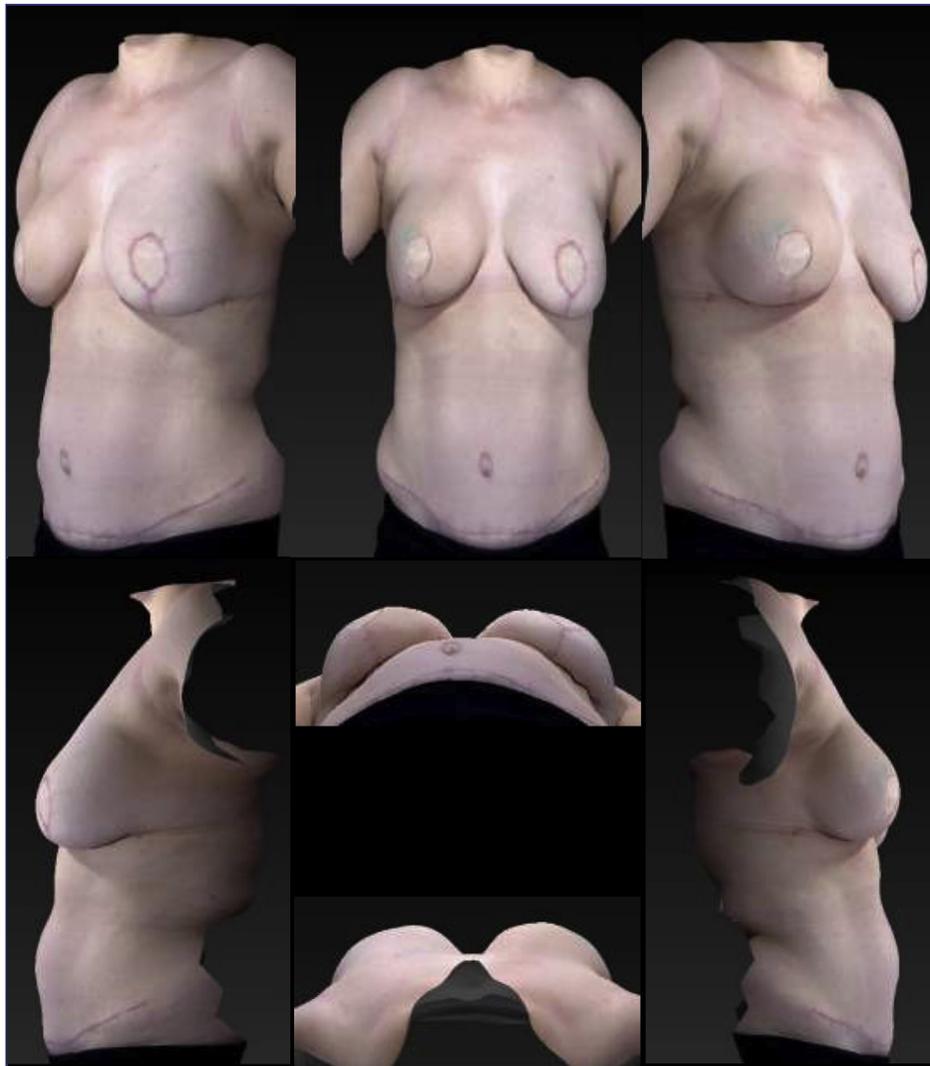


Fig. 1. Standardised views of 3D-SIs viewed by the panel including Antero-posterior (AP), Oblique (left and right), Lateral (left and right), cranial, and caudal.

Statistics

IBM SPSS 23 was used throughout. Delphi results were represented using simple descriptive statistics.

Intra-panellist and intra-panel reliability was reported using weighted kappa (w_k) for 10 repeated images. A w_k equal to 0 was considered to indicate poor agreement; 0.01–0.20 slight agreement; 0.21–0.40 fair agreement; 0.41–0.60 moderate agreement; 0.61–0.80 substantial agreement; 0.81–0.99 very good; and 1.00 perfect agreement.

Inter-panellist agreement was measured using the intra-class correlation coefficient (ICC). ICC of less than 0.40 was considered poor, between 0.40 and 0.59 fair, between 0.60 and 0.74 good and between 0.75 and 1.00 excellent [25]. Internal consistency was evaluated using Cronbach's alpha (α). An α value of ≥ 0.9 translated to excellent internal consistency, ≥ 0.8 to <0.9 good, ≥ 0.7 to <0.8 acceptable, ≥ 0.6 to <0.7 questionable, ≥ 0.5 to <0.6 poor, and <0.5 unacceptable [26].

Using Spearman's correlation coefficient (r), the correlation between each item (shape, symmetry, volume, position of breast mound, nipple position) with the global score was used to test the validity of the method. For interpretation, r of ≥ 0.9 – 1.0 (≤ -0.9 to -1.0) very high positive (negative) correlation, ≥ 0.7 to 0.9

(≤ -0.7 to -0.9) high positive (negative), ≥ 0.5 to 0.7 (≤ -0.5 to -0.7) moderate positive (negative), ≥ 0.3 to 0.5 (≤ -0.3 to -0.5) low positive (negative), 0.3 to -0.3 negligible correlation [27]. Spearman's correlation coefficient was also used to assess the correlation between the global score and PROMs (BREAST-Q satisfaction with breasts Q-score 0–100, where 100 is the best score).

Results

The Delphi consensus process

61 surgeons (88% consultants) completed round one, 49 round two, and 18 were involved in both the consensus discussion and final voting round (Fig. 2). Oncoplastic and plastic surgeons were represented (75% and 25% respectively). 46 were consultant oncoplastic surgeons (1 retired), 11 consultant plastic surgeons, and 4 were trainees. 19 (31%) had spent <3 years in their current post, 12 (20%) 3–5 years, 11 (18%) 6–10 years, 19 (31%) >10 years. 28 (46%) were male, 30 (29%) female, 1 (2%) transgender, and 2 (3%) preferred not to say.

Round one voting reduced the number of items included in the questionnaire from 20 to 17, and round two from 17 to 13, the final round established 5 items (surface symmetry, volume, shape,

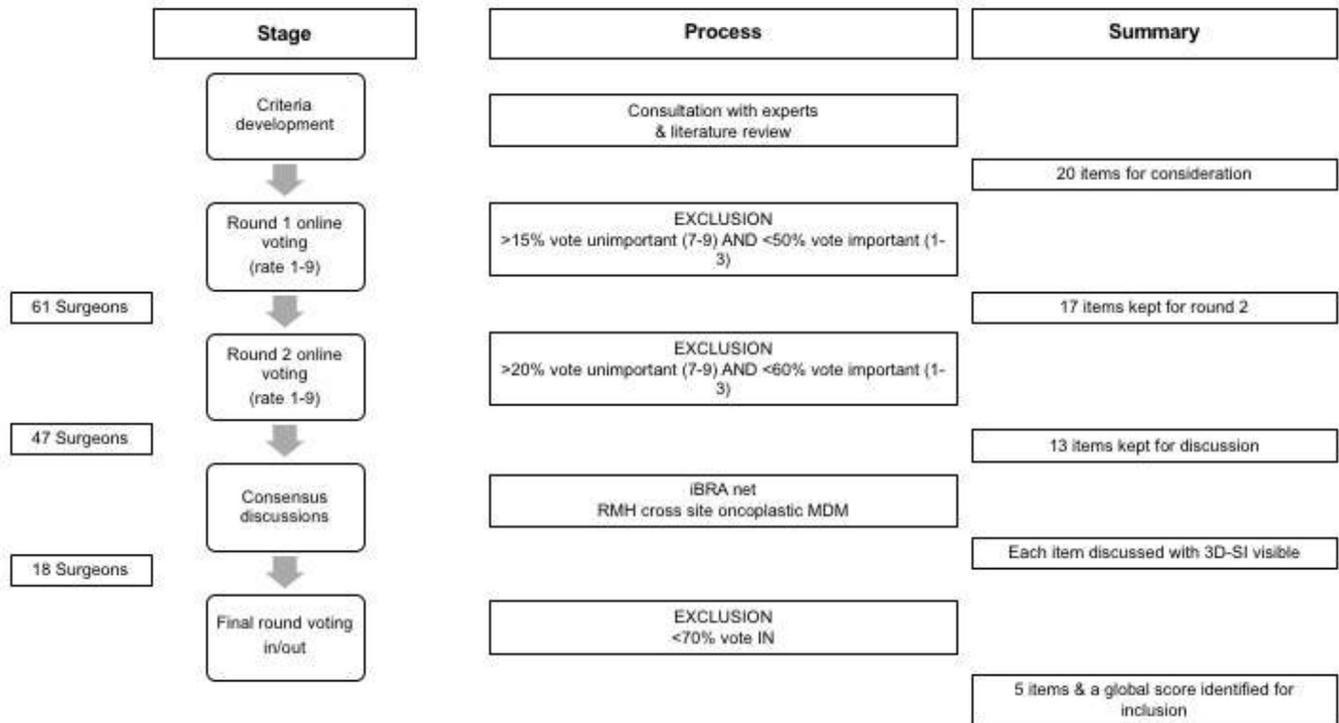


Fig. 2. The Delphi consensus process.

position of breast mound, nipple position) in addition to a global score for the Delphi-derived panel evaluation (Table 1). The majority voted for a 5-point Likert scale, and 3–5 panellists to comprise the panel.

The detailed description of scores for each item used in the panel evaluation is illustrated in Table 2. The rationale is not only to guide the panellists but to aid in the interpretation of the result making the process more clinically relevant.

Reliability of the panel methodology

Intra-panellist reliability

The intra-panellist reliability varied from poor to almost perfect agreement across the individual items ($\kappa < 0$ to 0.86), with better agreement for symmetry, volume, and the global score.

Inter-panellist reliability

The ICC could not be calculated for nipple position because there were too few cases i.e. the majority of women in the images had not had NAC conservation or reconstruction. The ICC for the other individual items was fair (range 0.4–0.5) and was good (0.6) for the global score (Table 3). Cronbach's alpha was good to excellent.

Intra-panel reliability

Intra-panel reliability for 10 repeated images showed moderate to substantial agreement between the mean panel score for the two occasions when the images were shown (range κ 0.4–0.7). Shape and symmetry demonstrating the strongest agreement (Table 4).

Validity

Each item (shape, symmetry, volume, position of breast mound, nipple position) independently showed a high or very high positive correlation with the global score (range $r = 0.88$ –0.92).

Panel evaluation and PROMs

There was a statistically significant but only moderate strength positive correlation between the mean global panel score and the 'satisfaction with breasts' domain of the BREAST-Q post-operative reconstruction module ($r = 0.5$ $p < 0.01$).

Discussion

The rationale for the development of this scale is primarily as the gold standard against which an objective measure of aesthetic outcome for breast reconstruction can be developed using measures derived from 3D-SI in a future multicentre study. 3D-SI may provide a highly reliable and robust way to communicate aesthetic outcome to enable comparison of results, benchmarking, and development of best practice guidelines, but a robust gold standard is required, against which to test these measurements.

This paper describes the successful development of a reliable, contemporary scale for evaluating breast aesthetic outcomes as captured on 3D-SI. The scale is designed specifically for breast reconstruction and was developed through a Delphi consensus process using expert oncoplastic and plastic breast surgeons. The strengths include the robust development of the scoring criteria using the Delphi process incorporating a literature review with opinion from experts representing different centres across the UK. The scale was developed specifically for breast reconstruction and has been tested by a multi-disciplinary panel. Careful consideration of the written descriptors for each individual item enabled panellists to have a clear understanding of what each score represented clinically. This is also important when interpreting scores to ensure a qualitative meaning can be gleaned from a quantitative score.

A systematic review by Maas et al. employed the modified Medical Outcomes Trust (MOT) criteria to evaluate the professional aesthetic assessment scales referenced in the literature [12]. The MOT criteria evaluate each scale based on a predefined scoring system encompassing 7 domains including development of the

Table 1
Progression of scales through the Delphi rounds with percentage voting important (1–3) and unimportant (7–9) in rounds 1 and 2 and percentage voting to keep or exclude each scale in round 3. The parameters for retaining items is explained.

Item	Round 1 (n = 61)		Progress to next round >50% vote important or < 15% vote unimportant	Round 2 (n = 47)		Progress to next round >60% vote important or < 20% vote unimportant	Round 3 (n = 18)	
	Voted Important 1–3 (%)	Voted Unimportant 7–9 (%)		Voted Important 1–3 (%)	Voted Unimportant 7–9 (%)		Percentage voting to keep the item (%)	Item used in final scale >70% voted 'in'
Symmetry	82	13	Yes	77	21	Yes	100	Yes
Global	82	11	Yes	72	21	Yes	83	Yes
Shape	80	10	Yes	66	15	Yes	89	Yes
Position of breast mound	79	15	Yes	64	26	Yes	72	Yes
Volume	74	11	Yes	68	19	Yes	89	Yes
Natural look	74	11	Yes	49	23	No	–	–
Nipple position	69	15	Yes	64	17	Yes	72	Yes
Fibrosis	67	11	Yes	51	15	Yes	22	No
Contour	64	16	Yes	55	17	Yes	55	No
Projection	59	7	Yes	40	15	Yes	61	No
Appearance of upper pole	59	15	Yes	40	23	No	–	–
Contracture	57	11	Yes	43	13	Yes	28	No
Scar	54	13	Yes	37	17	Yes	17	No
IMF height	52	10	Yes	47	17	Yes	44	No
Focal volume deficit or excess	52	13	Yes	45	19	No	–	–
Appearance of NAC	51	11	Yes	38	17	Yes	22	No
Skin paddle appearance	38	11	Yes	23	23	No	–	–
Implant edge	36	15	No	–	–	–	–	–
Skin colour match	34	23	No	–	–	–	–	–
Shape of areola	33	21	No	–	–	–	–	–

Table 2
Likert scale description used for reference during the panel assessment.

		Excellent 5	Good 4	Moderate 3	Poor 2	Very Poor 1
Shape	The overall shape of the reconstructed breast/s	Shape symmetry out of bra achieved	Shape of operated breast is pleasing but not symmetrical	Moderate difference in shape but does not detract from overall aesthetic result	Moderate focal deficits detracting from overall aesthetic result	Large focal deficits distorting contour significantly detracts from overall aesthetic result
Volume	Overall volume symmetry between breasts	Equal volume between breasts	Minor difference in volume	Moderate difference in volume but does not detract from overall aesthetic result	Volume difference impacts overall aesthetic result	Major volume mismatch significantly detracts from overall aesthetic result
Nipple Position	Nipple position in relation to the ipsilateral breast	Excellent symmetry between sides and nipple in an ideal position on reconstructed breast mound	Minor adjustments required to achieve excellence in nipple position	Noticeably suboptimal but does not influence overall aesthetic results	Nipple position slightly impacts overall aesthetic result	Nipple position significantly detracts from overall aesthetic result
Projection	Patient view of symmetry	Projection is equal	Minor differences in projection	Noticeable difference but not detracting from overall aesthetic result	Slightly impacts overall aesthetic result	Significantly detracts from overall aesthetic result
Position of Breast Mound	In relation to chest wall and other breast	Equal to the other side and in an optimal position on chest wall	Minor asymmetry of position or symmetrical but suboptimal position	Asymmetry of position or symmetrical but suboptimal position not detracting from overall aesthetic result	Slightly impacts overall aesthetic result	Significantly detracts from overall aesthetic result
Symmetry	Comparison between breasts	Out of bra symmetry achieved	Mild asymmetry	Moderate asymmetry but does not detract from overall aesthetic result	Moderate asymmetry detracting from overall aesthetic result	Significant asymmetry detracting from overall aesthetic result
Global	Taking into consideration subscale evaluation what is your overall impression of the quality of the reconstruction	Excellent	Good	Moderate	Poor	Very Poor

Table 3

Inter-panellist agreement and internal consistency of the panel methodology per scale. ICC; intra-class correlation coefficient.

Item	Cronbach's alpha	ICC (single measures)	Significance	ICC (average measures)	Significance
Position of Breast Mound	0.931	0.51	<0.01	0.903	<0.01
Nipple position	–	–	–	–	–
Symmetry	0.918	0.511	<0.01	0.904	<0.01
Volume	0.892	0.432	<0.01	0.872	<0.01
Shape	0.906	0.466	<0.01	0.887	<0.01
Global	0.938	0.564	<0.01	0.921	<0.01

Table 4

Intra-panel agreement using quadratic weighted kappa for each subscale for 10 repeated images. The mean panel score was used for evaluation.

Item	Weighted kappa
Position of breast mound	0.4
Nipple areola complex	–
Symmetry	0.7
Volume	0.4
Shape	0.7
Global	0.5

framework, reliability, validity, responsiveness, interpretability, burden (for professional and patient), and correlation to PROMs. The maximum score is 7. The aesthetic evaluation scale scoring most highly in the review by Maas (4.5 out of 7) was the ten-point Visser scale which includes 5 subscales (volume, symmetry, scar, nipple-areola complex, shape) and a global ten-point (Likert) scale. The Vrieling scale which incorporates 6 subscales (scar, size, shape, nipple position, shape of areola, skin colour) and a global four-point Likert scale from 0, excellent to 3, poor, was the most commonly referenced in the literature and scored 3 out of 7 [12]. When the MOT criteria was applied to the scale in this study, a comparable score of 5 results. Thus, until the 3D-SI objective scoring system is available, this Delphi-derived scale provides a robust alternative.

As with all panel assessments, some logistical constraints remain. The strict non-discussion policy during panel assessment was designed to mimic conditions of a 'virtual panel' i.e. independently viewing and scoring of images from a remote location, which would surmount the majority of logistical challenges associated with traditional panel assessment and potentially reduce the burden for panellists (travel, inconvenience). Discussion permits a level of benchmarking between panellists, as hearing the opinions of others may lead to re-evaluation of an individual's scoring. If this variation was significant it would be reflected in the inter-panellist variability i.e. it would be worse with no discussion.

There is a wide range of inter-panellist variability reported in the literature often using different statistical tests rendering comparison challenging. Visser et al. report an ICC ranging from poor to excellent (0.17–1) for the ten-point scale [28], and Vrieling et al. report moderate inter-rater reliability (κ 0.55) [29], and the linear numeric analogue scale (the second highest scoring scale in the Maas review) reports inter-rater agreement of 0.23–0.38. Inter-rater variability using the Visser scale for a similar cohort of patients scored in our unit was "moderate" ($w\kappa$ 0.4) for the global score and ranged from $w\kappa$ 0.2–0.36 for the subscales [14]. The inter-rater reliability for the method developed in this paper ranged from an ICC of 0.4–0.5 for the subscales and 0.6 for the global score, in line with the best scoring other scales suggesting remote panel assessment is a feasible alternative.

The number of experienced surgeons participating in the Delphi consensus was good, however, the inclusion of clinical oncologists would perhaps have broadened the experience and provided additional elements for consideration. The concept was to design a

panel to represent expert opinion on aesthetic outcome, hence the intentional exclusion of patient representation. Patient reported evaluation of aesthetics is consistently discordant with expert opinion hence the need to separate the two [28,30,31].

The Delphi-derived panel evaluation described in this paper was compared with PROMs, demonstrating a moderate correlation. Comparison between PROMs and expert panel evaluation is reported in the literature but it is difficult to draw reliable conclusions due to the different scales and methods of comparisons used. Ramon et al. report moderate correlation between the ten-point Visser scale with PROMs using spearman's correlation co-efficient ($r = 0.48$) [32] which is comparable to the results observed here ($r = 0.5$). The PROMs used in that study were in the form of a study specific questionnaire rather than the validated and widely used BREAST-Q which was used here.

Conclusions

We have successfully harnessed expert opinion to develop and test a scale for the expert evaluation of aesthetic outcome after breast reconstruction. This Delphi-derived panel evaluation is at least as good as other scales referenced in the literature and has been developed specifically for future use as a gold standard against which objective outcomes scores can be tested. The logistical flaws of panel assessment remain, giving weight to the importance of the development of an objective way to report upon aesthetic outcomes.

Compliance with ethical standards

Declarations of interest: none.

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standard. The protocol was reviewed by London-Riverside NRES committee Ref 17/LO/0763. The study is registered on a publicly accessible database, clinicaltrials.gov, NCT03203252.

Funding

Miss Amy R Godden is funded by a grant from the NIHR Royal Marsden/Institute of Cancer Research Biomedical Research Centre.

This work was also funded by a grant from the Association of Breast Surgery.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Amy R. Godden: Conceptualization, Methodology, Validation,

Formal analysis, Investigation, Writing - original draft, Funding acquisition. **Simon H. Wood:** Methodology, Writing - review & editing. **Stuart E. James:** Methodology, Writing - review & editing. **Fiona A. MacNeill:** Methodology, Writing - review & editing. **Jennifer E. Rusby:** Conceptualization, Methodology, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgements

This paper represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and the Institute of Cancer Research. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

We would like to formally acknowledge the contributions of the participants, the iBRA-net collaborative, Miss Rachel O'Connell and medical photographer Dennis Underwood to this study.

References

- [1] Al-Ghazal SK, Sully L, Fallowfield L, Blamey RW. The psychological impact of immediate rather than delayed breast reconstruction. *Eur J Surg Oncol* 2000;26(1):17–9.
- [2] Waljee JF, Hu ES, Ubel PA, Smith DM, Newman LA, Alderman AK. Effect of esthetic outcome after breast-conserving surgery on psychosocial functioning and quality of life. *J Clin Oncol* 2008;26(20):3331–7.
- [3] Evans AA, Straker VF, Rainsbury RM. Breast reconstruction at a district general hospital. *J R Soc Med* 1993;86(11):630–3.
- [4] Ringberg A, Tengrup I, Aspegren K, Palmer B. Immediate breast reconstruction after mastectomy for cancer. *Eur J Surg Oncol* 1999;25(5):470–6.
- [5] Reaby LL, Hort LK, Vandervord J. Body image, self concept, and self-esteem in women who had a mastectomy and either wore an external breast prosthesis or had breast reconstruction and women who had not experienced mastectomy. *Health Care Women Int* 1994;15:361–75.
- [6] Kim MK, Kim T, Moon HG, et al. Effect of cosmetic outcome on quality of life after breast cancer surgery. *Eur J Surg Oncol : J. Eur. Soc. Surg Oncol. Br. Assoc. Surg Oncol.* 2015;41(3):426–32.
- [7] Kim K-D, Kim Z, Kuk JC, et al. Long-term results of oncoplastic breast surgery with latissimus dorsi flap reconstruction: a pilot study of the objective cosmetic results and patient reported outcome. *Ann. Surg. Treat. Res.* 2016;90(3):117–23.
- [8] Jeevan R, Cromwell DA, Browne JP, et al. Findings of a national comparative audit of mastectomy and breast reconstruction surgery in England. *J Plast Reconstr Aesthetic Surg* 2014;67(10):1333–44.
- [9] Cancer Research Uk. Breast cancer incidence statistics. Available at: <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/incidence/uk-breast-cancer-incidence-statistics>. [Accessed October 2019].
- [10] National Institute for Clinical Excellence. Guidance on cancer services. Improving outcomes in breast cancer emanual update. 2002.
- [11] Mennie JC, Mohanna PN, O'Donoghue JM, Rainsbury R, Cromwell DA. National trends in immediate and delayed post-mastectomy reconstruction procedures in England: a seven-year population-based cohort study. *Eur J Surg Oncol* 2017;43(1):52–61.
- [12] Maass SW, Bagher S, Hofer SO, Baxter NN, Zhong T. Systematic review: aesthetic assessment of breast reconstruction outcomes by healthcare professionals. *Ann Surg Oncol* 2015;22(13):4305–16.
- [13] Potter S, Harcourt D, Cawthorn S, et al. Assessment of cosmesis after breast reconstruction surgery: a systematic review. *Ann Surg Oncol* 2011;18(3):813–23.
- [14] 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.
- [15] O'Connell RL, DiMicco R, Khabra K, et al. Initial experience of the BREAST-Q breast-conserving therapy module. *Breast Canc Res Treat* 2016;160(1):79–89.
- [16] O'Connell RL, Stevens RJ, Harris PA, Rusby JE. Review of three-dimensional (3D) surface imaging for oncoplastic, reconstructive and aesthetic breast surgery. *Breast* 2015;24(4):331–42.
- [17] Heil J, Dahlkamp J, Golatta M, et al. Aesthetics in breast conserving therapy: do objectively measured results match patients' evaluations? *Ann Surg Oncol* 2011;18(1):134–8.
- [18] Merie R, Browne L, Cardoso JS, et al. Proposal for a gold standard for cosmetic evaluation after breast conserving therapy: results from the St George and Wollongong Breast Boost trial. *Journal of medical imaging and radiation oncology* 2017;61(6):819–25.
- [19] Boyages J, Barraclough B, Middeldorp J, Gorman D, Langlands AO. Early breast cancer: cosmetic and functional results after treatment by conservative techniques. *Aust N Z J Surg* 1988;58(2):111–21.
- [20] Beesley H, Ullmer H Fau - Holcombe C, Holcombe C Fau - Salmon P, Salmon P. How patients evaluate breast reconstruction after mastectomy, and why their evaluation often differs from that of their clinicians. (1878-0539 (Electronic)).
- [21] Hirst A, Philippou Y, Blazeby J, et al. No surgical innovation without evaluation: evolution and further development of the IDEAL framework and recommendations. (1528-1140 (Electronic)).
- [22] McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet* 2009;374(9695):1105–12.
- [23] Mylvaganam S, Conroy EJ, Williamson PR, et al. Adherence to best practice consensus guidelines for implant-based breast reconstruction: results from the iBRA national practice questionnaire survey. *Eur J Surg Oncol : J. Eur. Soc. Surg Oncol. Br. Assoc. Surg Oncol.* 2018;44(5):708–16.
- [24] Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;32(4):1008–15.
- [25] Cicchetti D. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instrument in psychology. *Psychol Assess* 1994;6:284–90.
- [26] Cortina JM. What is coefficient alpha? An examination of theory and applications. *J Appl Psychol* 1993;78:98–104.
- [27] Hinkle DE, Wiersma W, Jurs SG. Applied statistics for the behavioral sciences. fifth ed. Boston: Houghton Mifflin; 2003.
- [28] Visser NJ, Damen TH, Timman R, Hofer SO, Mureau MA. Surgical results, aesthetic outcome, and patient satisfaction after microsurgical autologous breast reconstruction following failed implant reconstruction. *Plast Reconstr Surg* 2010;126(1):26–36.
- [29] Vrieling C, Collette L, Bartelink E, et al. Validation of the methods of cosmetic assessment after breast-conserving therapy in the EORTC "boost versus no boost" trial. EORTC radiotherapy and breast cancer cooperative groups. European organization for research and treatment of cancer. *Int J Radiat Oncol Biol Phys* 1999;45(3):667–76.
- [30] Potter S, Thomson HJ, Beresford M, et al. Evaluating the cosmetic outcome of breast reconstruction: a comparison of patients' and healthcare professionals' views and their associations with patient satisfaction and quality of life. *Canc Res* 2009;69(2).
- [31] Haekens CM, Enajat M Fau - Keymeulen K, Keymeulen K Fau - Van der Hulst RRWJ, Van der Hulst RR. Self-esteem and patients' satisfaction after deep inferior epigastric perforator flap breast reconstruction. (1550-1841 (Electronic)).
- [32] 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):686–91.