Original article

Validation of an oncoplastic breast simulator for assessment of technical skills in wide local excision

D. R. Leff[1,2], G. Petrou[1], S. Mavroveli[1], M. Bersihand[2], D. Cocker[1], R. Al-Mufti[2], D. J. Hadjiminas[2], A. Darzi[1] and G. B. Hanna[1]

[1]Department of BioSurgery and Surgical Technology, Imperial College London, and [2]Breast Unit, Imperial College Healthcare NHS Trust, London, UK

Correspondence to: Mr D. R. Leff, Department of BioSurgery and Surgical Technology, 10th Floor, QEQM Wing, St Mary's Hospital, Paddington, London W2 1NY, UK (e-mail: d.leff@imperial.ac.uk)

Background: Simulation enables safe practice and facilitates objective assessment of technical skills. However, simulation training in breast surgery is rare and assessment remains subjective. The primary aim was to evaluate the construct validity of technical skills assessments in wide local excision (WLE).

Methods: Surgeons of different grades performed a WLE of a 25-mm palpable tumour on an in-house synthetic breast simulator. Procedures were videotaped (blinded), reviewed retrospectively, and independently rated against a procedure-specific global rating scale by two consultant breast surgeons. Specimen radiographs were obtained and the macroscopic distance from the 'tumour' edge to the resection margin was recorded in four cardinal directions. Expert consensus was used to construct an Oncoplastic Deviation Score (ODS), assigning points for excessively wide (more than 10 mm) and, conversely, close (less than 5 mm) macroscopic margins.

Results: Thirty-four surgeons (12 consultant surgeons, 12 specialty trainees and 10 core trainees) participated in the study. Video-based rating scores varied hierarchically with operator expertise ($P < 0{\cdot}050$). Inter-rater reliability was excellent ($\alpha \geq 0{\cdot}80$, $P < 0{\cdot}050$ for all scales), and inter-rater agreement was moderate ($\kappa = 0{\cdot}132 - 0{\cdot}361$, $P < 0{\cdot}050$ for all scales). Statistically significant differences were observed on pairwise comparisons between each grade of surgeon in scores for 'exposure', 'skin flap development', 'glandular remodelling', 'skin closure' and 'final product review' ($P < 0{\cdot}050$). Consultants received significantly fewer ODS points than specialty trainees ($P = 0{\cdot}012$) and core trainees ($P = 0{\cdot}028$). Compared with experts (median 9·0 mm), wider margins were observed amongst specialty trainees (median 12·0 mm) and narrower margins amongst core trainees (median 7·1 mm) ($P = 0{\cdot}001$).

Conclusion: Video ratings of performance and a proposed ODS differentiate surgeons based on technical skills in WLE and may be useful for objective assessment of breast surgery trainees.

VIDEO AVAILABLE ONLINE

Paper accepted 15 September 2015

Published online 17 December 2015 in Wiley Online Library (www.bjs.co.uk). **DOI:** 10.1002/bjs.9970

Introduction

Simulation may facilitate the deliberate practice of index procedures in a safe environment, prepare trainees better for the real operating room experience, and help bridge the gap between skills required for core breast surgical procedures and advanced techniques in an era of oncoplastic surgery[1]. However, despite the diffusion of simulation training in certain surgical specialties such laparoscopic[2−4] and vascular[5,6] surgery, simulation training and skills assessment in breast surgery has been quiescent. Breast surgery trainees have yet to benefit from high-fidelity surgical simulations to refine and practise their skills, and assessment of technical competence amongst breast surgery trainees remains subjective. There are few publications describing simulation training in breast surgery; reports are focused on axillary sentinel node surgery as opposed to surgery on the breast gland itself, and these studies lack objective validation[7]. Moreover, following a recent needs assessment survey[8], more than half of the respondents (55 per cent) felt that deaneries could use simulation to bridge gaps in oncoplastic training.

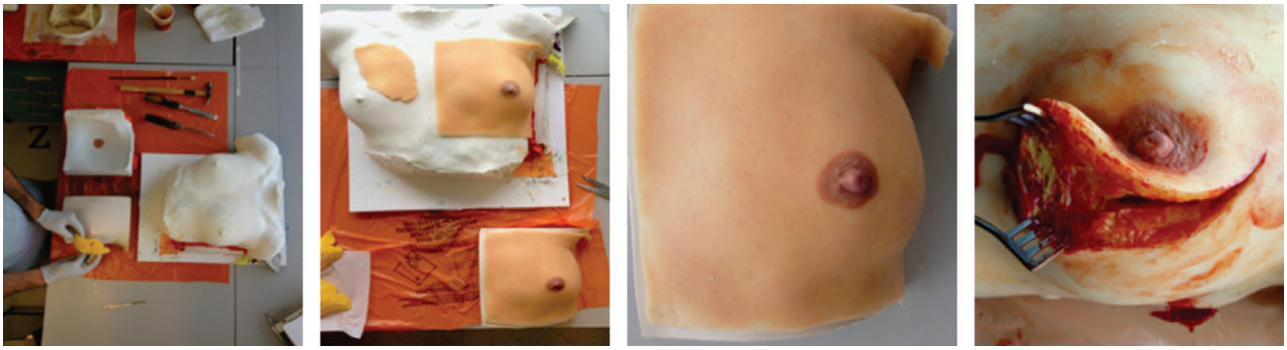

Fig. 1 Breast simulator model prototype development phases, including **a**,**b** plaster of Paris casting, **c** silicone gland construction and **d** pilot testing

Within the context of the London General Surgical Skills Programme, a simulator to improve training and skills assessment in breast surgery was developed recently (*Fig. 1*). The model facilitates the practice and assessment of technical skills in wide local excision (WLE), wire-guided WLE and mastectomy. The aim of the present study was to investigate face, content and construct validity of the simulation for assessment of technical skills in WLE. The null hypothesis was that the model insufficiently represented real WLE, and that procedure-specific global rating scales or end-product analyses would be incapable of differentiating breast surgeons with varying operative experience.

Methods

Model development and simulated wide local excision

The clinical team worked with the project designer for approximately 2 h per week for 6 months to develop the prototype simulator (*Fig. 1*). The focus was to ensure that anatomical structures were represented, the visual aesthetic and tactile features of breast pathologies were realistic, and the integral steps of WLE could be reproduced. The synthetic 'tumour' was designed using gypsum (plaster of Paris), to be readily palpable across the external surface of the simulated breast. To enhance realism, at the site of likely incisions on the breast, simulated blood vessels were introduced that appeared to bleed when transected. Silicone gels varying in colour and consistency were used to differentiate anatomical entities such as subcutaneous fat and the underlying breast gland. The latter enabled development of skin flaps over the tumour mass. During material testing, no synthetic substance sufficiently conducted electrocautery, as common materials are insulators and not conductors of energy. Therefore, sharp dissection was required. As the pectoralis fascia is the posterior limit of an oncological WLE, an anatomical layer of 'angel hair' fibre was introduced between silicone layers representing the retromammary fat and the pectoralis fascia. The entire simulated gland was mounted on a firm and gently sloping frame to represent the underlying chest wall. The base of the frame included Velcro stickers, to ensure that the simulator could be mounted securely on to the operating table.

Task paradigm, model consistency and quality assurance

Participants were asked to perform a WLE procedure from start to finish or 'skin to skin' on the oncoplastic simulator, as illustrated in *Fig. 2*. From a technical perspective, the operating surgeon was required to remove a 25-mm palpable breast 'lesion' located 30 mm from the nipple–areola complex in the 3 o'clock position, on the in-house synthetic breast simulator. Communication was standardized. All subjects received the same briefing, as follows: 'This is Mrs Betty Breast, she is a 55-year-old woman who has recently been diagnosed with an invasive cancer of the left breast. The tumour has been deemed to be palpable and does not require radiological localization. The tumour can be palpated in the 3 o'clock position of the breast. The tumour is approximately 25 mm in diameter (vertically and horizontally). Your local multidisciplinary team has approved your decision to proceed to WLE. Please perform what you would consider to be your standard WLE procedure or (for junior surgeons) that which you have observed'.

Efforts were made to control variables that might confound outcomes. To eliminate subtle differences in the ergonomics between operating on the right or left breast,

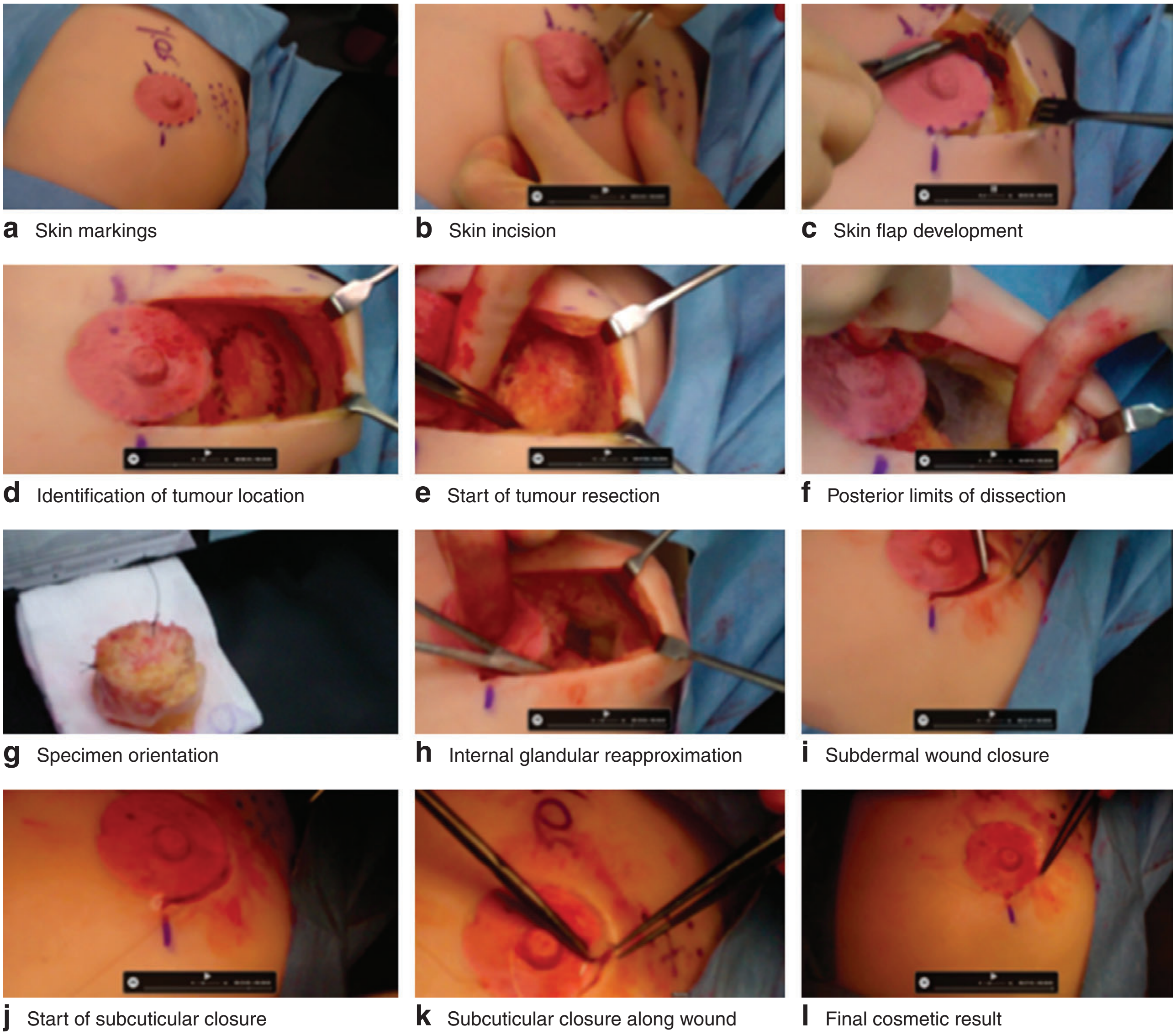

Fig. 2 Photographic images of a wide local excision procedure performed by a consultant breast surgeon. Critical steps of the procedure obtained from the video data stream and subcategorized according to: **a** external periareolar skin markings; **b** skin incision using a 10-mm scalpel blade; **c** skin flap development using MacIndoe scissors; **d** identification of the tumour location marked on the gland surface; **e** tumour resection as started at gland surface; **f** tumour resection illustrating posterior limits of dissection (angel hair fibres); **g** specimen orientation for histopathological assessment (long suture marks lateral margin, short suture marks superior margin); **h** internal glandular reapproximation; **i** subdermal wound closure using 2/0 poliglecaprone 25; **j** subcuticular wound closure commenced using 3/0 poliglecaprone 25; **k** subcuticular closure continued along the wound; and **l** final cosmetic result

only simulators cast from the left breast were used. Similarly, the WLE procedure may be influenced by glandular consistency. The breast of an elderly patient handles differently from the tough glandular breast of a younger woman. The look, consistency and handling of the silicone that simulates the gland was controlled by the designer to limit differences in glandular handling being a factor accounting for differences in technical performance. Likewise each ‘tumour’ measured 25 × 25 × 25 mm, which, although artificial, helped to ensure consistency in the target resection for each subject. Finally, the ‘tumour’ was always placed within the simulated breast at a distance of 30 mm from the nipple and at a depth of 20 mm below the simulated cutaneous surface.

Participants

Following National Regional Ethics Committee approval (CRO1587), subjects from Imperial College London-affiliated institutions were recruited. Given the novelty of the simulation and assessment methods such as the Oncoplastic Deviation Score (ODS), it was not possible to power study numbers. Subjects reflected local and regional expertise.

Data collection protocol

Subjects were invited to perform a WLE procedure on the oncoplastic simulator as described above. Procedures were recorded using a hand-held video camera (Flip Video UltraHD; Pure Digital, San Francisco, California, USA) for retrospective viewing and procedure-specific global rating analysis. To blind the assessors regarding the identity of the operating surgeon, the focus of attention was maintained on the operator's hands. Similarly, each participant was asked to wear a gown and gloves. Audio data were eliminated during performance scoring to prevent participants from being identified. Each participant was given a unique study identifier code (for example, WLE 001), and data analysis was performed using coded data. An Excel® (Microsoft, Redmond, Washington, USA) database was maintained by the investigators in case it became necessary to match codes with participants. In accordance with ethical approvals, participants were not identifiable in performance review, data analysis or report generation. Similarly, any video not used for performance scoring was destroyed ($n = 1$).

Assessment of technical skills

A fourfold approach to the assessment of participant WLE performance was conducted as follows.

Video-based ratings of performance

Two expert trainers reviewed the procedural videos and independently scored performance against a procedure-specific global rating scale. A WLE-specific global rating checklist (0–100 visual analogue scale) was devised that included categories deemed by two consultant oncoplastic breast surgeons to be integral procedural skills (for example, flap development, resection skills, glandular remodelling, end-product assessment). The rating scale (*Fig. 3*) included descriptors anchoring the extreme scores, which were designed to assist the experts when scoring technical performance (for instance, good = excellent procedural flow, incision in Kraissl's lines with optimal exposure; poor = dissection proceeding in inadequate tissue planes, poor use of instruments, hesitant performance). Inter-rater reliability and agreement of technical skills assessments were also evaluated.

Assessment of margin width and the Oncoplastic Deviation Score

The distance from the edge of the 'tumour' to the cut edge of the resected specimen was calculated in four cardinal directions from specimen radiographs, as illustrated in *Fig. 4*. A team of consultant surgeons derived the ODS, which assigns points for excessively wide (greater than 10 mm) and conversely 'close' (less than 5 mm) macroscopic margins (*Table 1*). Of an external group of 19 expert consultant breast surgeons, 16 (84 per cent) agreed that experts would be expected to receive a lower ODS than novices.

Resection weights

Specimen weights were recorded.

Tumour at the resection margin

Each WLE specimen was reviewed carefully to see whether a macroscopically involved margin could be identified (visible 'tumour', yes or no). When a grossly positive margin was identified, video data were reviewed retrospectively to determine the operator's response (for example, decision-making regarding the need for a further shave excision).

Content and face validation questionnaire

A questionnaire was developed to interrogate WLE model realism and suitability of the simulation for training and technical skills assessment (*Table S1*, supporting information). On completion of the simulated WLE, surgeons were asked to complete the questionnaire, which gauged opinions on the anatomical realism, the realism of the WLE procedure, and the value of the simulation as a training and assessment tool.

Statistical analysis

Video ratings of performance (0–100), margin width, the ODS and specimen weight provided quantitative data, analysed using SPSS® version 22 (IBM, Armonk, New York, USA). Data were observed to be non-parametric, and hence non-parametric tests of significance for continuous data were used. Regarding the global rating scores, two scores for each participant (one from each rater) were available for analysis, and participants were labelled according to their level of experience (1 for consultants, 2 for trainees,

Category	Score 0	10	20	30	40	50	60	70	80	90	100
Incision											
	Incision placed in cosmetically insensitive location/too far from lesion							Incision placed in discrete location (i.e. periareolar or inframammary fold)			
Exposure											
	Inadequate exposure (incision too small); excessive incision for lesion size (incision too large)							Incision in Langer's or Kraissl's lines with optimal exposure of lesion			
Flap development											
	Dissection proceeding in incorrect tissue planes; flap excessively thick or thin							Optimal dissection between subcutaneous tissues and breast gland; skin flap of ideal thickness to ensure margins and skin viability			
Resection skills											
	Excessive force/manipulation during dissection/clumsy/uncoordinated/ inadequate use of assistance/ skin trauma/slow or stuttering performance							Excellent procedural flow; use of instruments and assistance to achieve resection with minimal trauma to skin			
Specimen orientation/resection depth											
	Inadequate specimen orientation/ appearance suggests threatened margin or resection to inadequate depth (i.e. not down to pectoralis fascia)							Orients specimen/cylindrical circumferential resection centred on tumour (no 'tumour' visible at margin)			
Glandular remodelling											
	No attempt or inadequate attempt to refashion breast disc							Excellent reconstruction of breast pillars to minimize cosmetic insult			
Skin closure											
	Poor closure, single layer, cosmetically insensitive (e.g. interrupted); poor or uneven tissue approximation; dog-ear deformity; poor instrument manipulation							Excellent procedural flow; safe and secure two-layer closure (subdermal and subcuticular closure); no dog-ear appearance; excellent skin approximation			
End-product evaluation											
	Overall, poor incision placement, unnecessarily large or small incision, inadequate final cosmetic appearance							Quantification of distance between edge of tumour and resection of margin in four planes using portable X-ray (numerical score of combined distances)			

Fig. 3 Procedure-specific global rating scale for wide local excision, with descriptors anchoring extreme values

3 for novices). The Kruskal–Wallis test was used to compare performance scores amongst the three experience groups. For any given step of the procedure (such as 'incision placement'), the Kruskal–Wallis test was used to determine whether there was a significant difference between all three experience groups. If the test reached threshold (a difference was observed between group rating scores for that procedural step), a further statistical test was required to determine where the differences lay. *Post hoc* comparisons in performance between any two of the experience groups were done using the Mann–Whitney U test. This process of comparing rating scores between experience groups was repeated for each step of the procedure for which scores were available. Internal consistency was analysed using Crohnbach's α, and inter-rater agreement using Cohen's κ – valid methods for evaluating, respectively, the degree of internal reliability and the concordance between two independent raters for categorical items. Gross macroscopic margin positivity derived from visual inspection of resected specimens (yes or no) provided categorical data, analysed using the χ^2 test. For all statistical tests, $P < 0{\cdot}050$ was deemed to be significant. Responses to questions regarding face and content realism were semiquantitative (1–5 Likert scales), and were therefore expressed as median scores or as percentages of agreement or disagreement with specific statements.

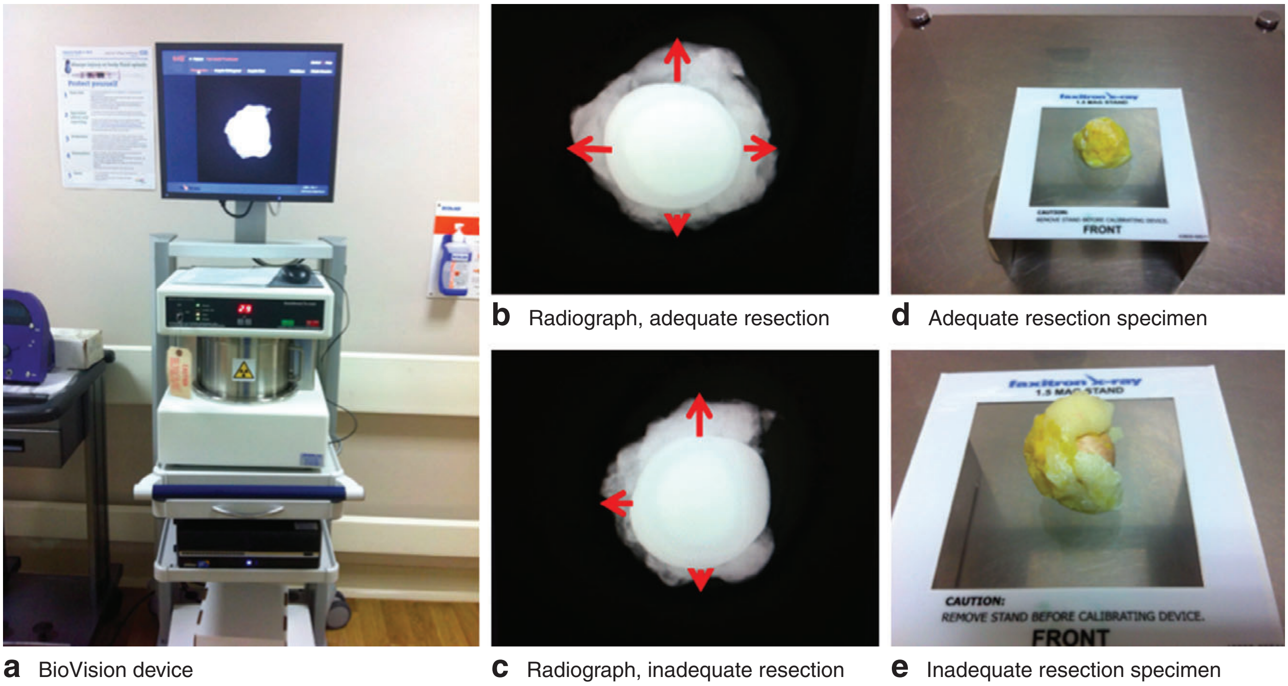

Fig. 4 **a** BioVision device (Faxitron®, Tucson, Arizona, USA) used for surrogate analysis of the oncological adequacy of resection. The distinction between **b** adequate and c inadequate resection is evident on intraoperative specimen radiographs, and from gross macroscopic appearance (**d** and **e** respectively). Arrows indicate measurements obtained in four cardinal directions

Table 1 Expert-derived Oncoplastic Deviation Score, a cumulative total of penalty points awarded for each radial margin of interest

Margin width (mm)	ODS points
0·0–1·0	300
1·1–2·0	200
2·1–5·0	20
5·1–10·0	0
10·1–15·0	10
≥15·1	100

Points are awarded for close (less than 5 mm) and wide (greater than 10 mm) margins of clearance. ODS, Oncoplastic Deviation Score.

Results

Of 34 participants, 12 were classified as experts, having performed more than 100 independent WLE procedures. Of these, nine were practising as consultant breast surgeons, one held the certificate of completion of surgical training and was waiting to take up a consultant appointment, and two were associate specialists who practised WLE independently. Data for one consultant were subsequently excluded from analysis as the participant approached the procedure as for a benign lesion. Twelve participants were breast specialty trainees who had completed sufficient training to be eligible to be invited to join the 'breast specialty skills' training stations of the London General Surgical Skills Programme (median postgraduate years of surgical training 5 (range 5–6) years). These trainees had never performed WLE independently. Ten subjects, classified as novices, were either foundation trainees with 4 months' previous experience assisting on WLE procedures (2) or were practising as core trainees in surgical specialties (8). None of the subjects in the novice category had any previous experience of performing WLE procedures, either in part or in total, as first operator.

Face and content validation

Anatomical structures

Table 2 provides a summary of respondents' feedback regarding anatomical realism of the simulator from simulated skin to ribcage. Skin, subcutaneous fat, the nipple–areola complex appearance and breast cancer pathology were all deemed to be realistic (median score 4·0 or above on a 5-point scale). The lowest median realism scores were given to the lobules, ducts and Cooper's ligaments.

Tissue handling

Feedback regarding the tissue-handling characteristics of the model are summarized in *Table 2*. All simulated tissue

Table 2 Results of face and content questionnaires assessing model realism, suitability for training, and assessment of technical skills in wide local excision

	Score
Anatomical layer of interest	
Skin	4·0 (4–5)
Subcutaneous fat	5·0 (4–5)
Nipple–areola complex	4·0 (4–5)
Lobules	3·0 (3–5)
Ducts	3·0 (2–4)
Cooper's ligaments	3·0 (2–4)
Pectoralis muscle	3·0 (2–5)
Ribs	3·0 (1–4)
Breast cancer pathology	4·0 (3–4)
Overall realism	4·0 (4–4)
Tissue-handling parameters	
Overall	4·0 (4–5)
Skin	5·0 (4–5)
Subcutaneous fat handling	4·0 (4–5)
Tumour excision	4·0 (4–4)
Skin closure	4·0 (3–4·5)
Instrument handling	4·0 (4–5)
Suitability for training	
Useful learning tool	5·0 (4–5)
Simulates real WLE	5·0 (4–5)
Model lacks components to teach WLE	2·0 (1–2·8)
Model useful for training for real WLE	4·0 (4–5)
Suitability for assessment (trainers only)	
Hand movements simulate those of real WLE	4·5 (4–5)
Tissues handle as per real WLE	3·0 (3–4)
Model provides information on trainee competence	4·0 (4–4)
Model can assess trainee performance	4·0 (4–4)
If performs well, likely to allow real WLE	3·5 (2·8–4)
Competence in model = competence in real WLE	4·0 (3–4)

Values are median (i.q.r.). WLE, wide local excision.

layers that required manipulation were observed to be realistic (median score 4·0 or above).

Suitability for training

Table 2 shows that the model was considered to be a useful learning tool, effectively simulating real WLE, and it was believed that practice on the model would be useful for real performance. On average, respondents disagreed with the statement that the model lacked essential components to teach WLE skills effectively (median 2·0, equivalent to 'disagree').

Suitability for assessment

The results of expert feedback on the suitability of the simulator for assessment are also summarized in *Table 2*. Consultants considered that competence on the model probably translated to competence in real WLE, and that valid information was provided on trainee competence at WLE (median 4·0 or above).

Procedure-specific global rating scores

Some 33 videos were viewed and scored independently by two expert surgeons. *Table S1* (supporting information) provides median (i.q.r.) scores of procedure-specific global rating scales for each experience group, subclassified according to the operative phase of WLE. For each phase of the operative procedure, statistically significant differences were observed between groups of surgeons with varying operative experience. Critically, the highest median performance scores were observed for the expert group and the lowest for novices.

Table S2 (supporting information) summarizes the results of pairwise comparisons in procedure-specific global rating scores. For each operative phase, expert scores were significantly superior to those of both trainees and novices. However, not all pairwise comparisons were significant; the performance of trainees and novices could be distinguished only for certain substeps. Specifically, scores for exposure, flap development, glandular remodelling, skin closure and final product assessment differentiated all experience groups (*Fig. 5*); this is demonstrated in *Video S1* (supporting information). *Tables S3* and *S4* (supporting information) demonstrate excellent inter-rater reliability ($\alpha > 0{\cdot}80$, $P < 0{\cdot}050$), and moderate inter-rater agreement ($\kappa = 0{\cdot}132-0{\cdot}361$) and agreement on each operative phase ($\alpha > 0{\cdot}80$, $P < 0{\cdot}050$).

Margins of clearance and the Oncoplastic Deviation Score

The ODS varied significantly with operator expertise ($P = 0{\cdot}023$, Kruskal–Wallis test). Experts received a significantly lower median (i.q.r.) ODS than either trainees or novices: experts 30·0 (30–120), trainees (specialty trainees) 175·0 (55–310), novices (core trainees) 165·0 (47·5–530) ($P = 0{\cdot}012$, experts *versus* trainees; $P = 0{\cdot}028$, experts *versus* novices, Mann–Whitney U test). There was no significant difference in ODS between trainees and novices ($P = 0{\cdot}974$, Mann–Whitney U test). However, margin measurement data showed that, on average, trainees resected simulated tumours with wider margins of clearance and novices with narrower margins of clearance. Compared with experts (median 9·0 mm), wider margins were observed amongst specialty trainees (median 12·0 mm) and narrower margins amongst core trainees (median 7·1 mm) ($P = 0{\cdot}001$). Therefore, both specialty and core groups are appropriately penalized by the ODS.

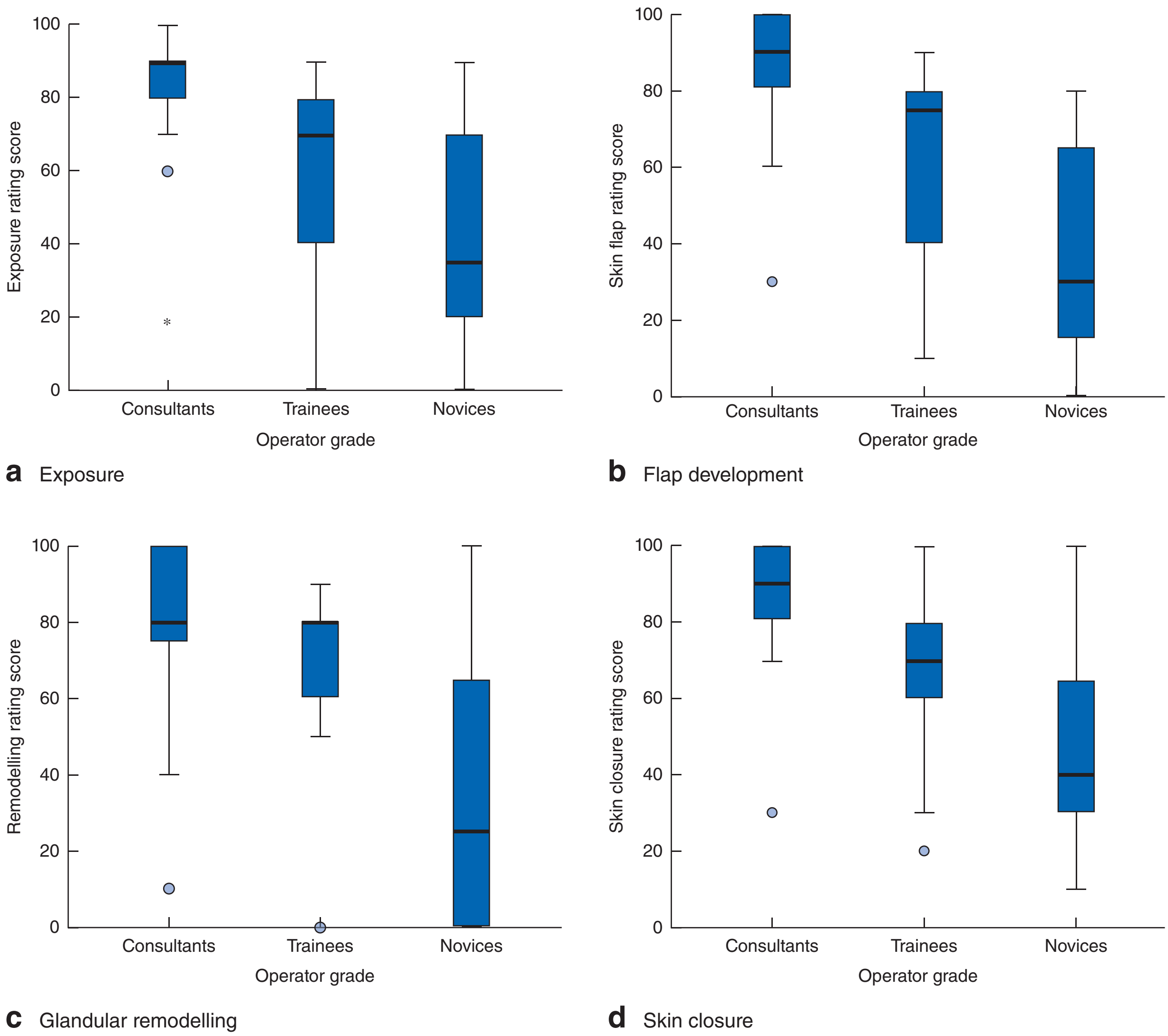

Fig. 5 Box plots comparing procedure-specific global rating scores between operators for **a** exposure, **b** flap development, **c** glandular remodelling and **d** skin closure. Median values, i.q.r. and ranges, excluding outliers (◦) and extreme values (*), are denoted by horizontal bars, boxes and error bars respectively

Macroscopic margin status

Although the number of specimens with macroscopically exposed 'tumour' was proportionally greater for novices (80 per cent) compared with trainees (50 per cent) and experts (45 per cent), no statistically significant differences were observed ($\chi^2 = 0{\cdot}22$, $P = 0{\cdot}203$).

Specimen weights

Specimens were observed to be lighter on average in WLEs conducted by novice surgeons *versus* trainees and consultants. However, there was no statistically significant relationship between median (i.q.r.) specimen weight and operator grade (experts, 38·0 (26·0–51·0) g, specialty trainees, 39·4 (33·4–55·9) g, core trainees, 31·0 (18·9–41·0) g; $P = 0{\cdot}172$).

Discussion

Skills training in the operating theatre is undeniably superior to other forms of training. However, several factors, including the dilution of exposure to core oncological procedures as a result of working time restrictions[9], shifts to competency-based curricula[10] and the potential for bias in assessments by trainers[10], have led to increasing interest in simulation. Moreover, with the publication of surgeon-specific outcomes[11], the public are becoming increasingly aware of variability associated with procedural learning curves, and training opportunities outside the operating room may help to reduce such performance variability. Indeed, simulation training has gained popularity as a valuable training tool that can expose trainees to a safe, stress-free environment, and high-fidelity simulation may provide learning opportunities to practise and improve skills as well as help to assess performance objectively[10]. Here, a synthetic breast simulator has been developed to practise and assess skills in WLE in the surgical treatment of breast cancer.

Although simulators for practising axillary procedures such as sentinel node biopsy have been described previously[7], this is the first simulator to have been developed for the purpose of training and assessing technical skills on the gland itself, and is also the first breast simulator to undergo rigorous validation testing. The model was found to have good overall anatomical realism, and the simulated tissues were found to handle in a similar fashion to real breast tissues. Of the assessment methods employed, procedure-specific global rating scales were found to be construct valid and to have excellent inter-rater reliability, sufficient for high-stakes examinations[9]. End-product assessments in the form of a margin width deviation score were able to differentiate surgeons based on high- and low-volume operative experience. Finally, the model was considered to be a useful training tool, and expert surgeons agreed that the model and assessment methods employed were suitable for trainee evaluation in WLE.

Regarding the null hypothesis that the assessment methods could not distinguish surgeons of differing experience and grade, this has been disproved and is therefore rejected. Specifically, significant differences were observed between operators in performance scores on each phase of the WLE procedure. This notwithstanding, only certain steps differentiated between all operators based on operative experience (experts *versus* trainees, trainees *versus* novices, and experts *versus* novices). Thus, should the WLE simulator and this system of performance scoring be adopted for trainee assessment, it would be recommended to restrict performance scoring to the steps of exposure, flap development, glandular remodelling, skin closure and overall procedural quality, as only these steps differentiated performance between all three experience groups. In general, the results add to a large body of data regarding the validity and reliability of expert ratings of trainee performance using retrospective video analysis[12–16]. Importantly, the present study demonstrates not only that operators can be distinguished based on performance scoring, but also that independent and blinded assessments have excellent reliability and interobserver agreement. Moreover, procedure-specific global rating scales such as the one employed here may have distinct advantage over generic global ratings such as Objective Structured Assessments of Technical Skills (OSATS) style assessments, as they may provide improved feedback to learners to focus their efforts better on a given aspect of the procedure[12,13].

The fundamental challenges in accepting this method of assessment are time and costs. Similar technical skills assessments conducted by Martin and colleagues[17] required 48 examiners to assess 20 surgical trainees for 3 h each. For the present evaluation, two expert assessors were required to review and score every video independently. Given that performance scoring required about 15 min per video, approximately 16·5 h of expert consultant surgeon time was required to complete the evaluation. Hence, rolling out such an assessment would require significant investment in terms of human resources, aside from the manufacturing time and costs for simulator reproduction. From this perspective, end-product assessments such as the ODS may be a useful surrogate for quality in WLE. Compared with video ratings, the ODS calculation is quick to execute, certainly objective (it cannot easily be manipulated to score more points), and could even be conducted in the absence of an expert consultant assessor. However, it must be acknowledged that the participants were not instructed to aim for a given macroscopic margin, and indeed the target resection width may vary depending upon local unit margin policy. This notwithstanding, certain patterns observed during these end-product assessments warrant further discussion.

Interestingly, wider margins were observed amongst trainees compared with novices and expert surgeons. Theoretically, the narrower margins of clearance amongst novices *versus* trainees may reflect their inexperience in terms of their three-dimensional perception of tumour dimension(s), tumour location and the required macroscopic margin of clearance. Supporting this theory, it is interesting to note that median specimen weights were lightest in the novice cohort, although this was not statistically significant, and minimum margins of clearance were most likely to be found amongst novices. Trainees, on the other hand, were perhaps more focused on oncological control, which manifested as wider margins. By comparison, experts, arguably more accustomed to balancing both oncological control with cosmesis, sought to remove only as much tissue as absolutely necessary to achieve oncological control whilst also considering the cosmetic outcome. Similarly, this may account for improved 'remodelling' scores in experts, as the size of defect to approximate may have been smaller and hence easier to remodel than defects created by novices and trainees. In this regard, data comparing short-term oncological outcomes between high- and low-volume breast surgeons, or between specialist breast surgeons and general surgeons, might help to contextualize these hypotheses.

Although no systematic evaluations have been undertaken to compare outcomes amongst trainees and expert breast consultants, there have been a few studies exploring volume–outcome relationships and investigations comparing specialists with non-specialists[18,19]. General surgeons have been found to perform WLE with a higher frequency of transected positive margins, close margins (defined by authors as less than 1 mm) and need for re-excision than specialist breast surgeons[19]. Comparison of the performance, including margin width and gross margin positivity, between general surgeons and those with breast specialist training using simulation(s) such as those described here may be useful in certifying and recertifying oncoplastic surgeons. Indeed, similar differences in performance have been observed between high- and low-volume surgeons. For example, Lovrics and co-workers[18] observed that crude positive margin rates were significantly higher for median- and low-volume surgeons than for high-volume surgeons. Taken together, these data may help to explain the narrower margin widths observed amongst novice breast surgeons in the present study. In this study the frequency of macroscopic margin positivity was not influenced by operator experience, but it was noted that experts more often recognized a grossly involved margin, prompting an immediate margin re-excision (percentage further shave excision when gross margin positive: experts, 3 of 4; trainees, 1 of 5; novices, 1 of 7). In the trial by Zork *et al.*[19], specimen weights were found to be lighter on average amongst specialist breast surgeons, although differences were not statistically significant. In contrast, the present authors observed lighter specimens amongst novice breast surgeons. As cosmetic outcome depends more on the volume of breast tissue resected than on weight *per se*[20,21], a method could perhaps be developed to relate the specimen weight to the 'preoperative' volume of the breast simulator.

Finally, the results of participant questionnaires clearly demonstrated overall content realism and good tissue-handling properties of the simulation. Expert surgeons agreed that the model has all the necessary components to assess trainee performance in WLE, that competence in simulated WLE was likely to reflect competence in real WLE, and that the model provides valid information on trainee performance. It is hoped that, in a manner similar to the sentinel lymph node biopsy trainer developed by Keshtgar and colleagues[7], which was employed to train consultants as part of the NEW START training programme, breast specialty trainees will now be able to benefit from repeated safe practice and technical skills training using the new validated oncoplastic breast simulator.

Further work will focus on the potential impact of simulation training in reducing learning curves and improving real performance during oncoplastic procedures, as has been demonstrated recently for endovascular[22] and cystourethroscopic[23] skills. In particular, it would valuable to demonstrate that, with repeated practice on the simulator, trainees learn to optimize margin control, removing sufficient tissue to obtain clear margins but not excessively excising healthy breast tissue.

Acknowledgements

This work was supported by educational grants from Imperial College Healthcare NHS Trust's Breast Cancer Charity, Professor Dudley Sinnett, the Pink Ribbon Foundation, and the Association of Breast Surgery. The authors thank S. Marchington and M. Roy at the London Deanery Skills Programme headquarters for their help in organizing simulation training and assessment sessions.
Disclosure: The authors declare no conflict of interest.

References

1 Down SK, Pereira JH, Leinster S, Simpson A. Training the oncoplastic breast surgeon – current and future perspectives. *Gland Surg* 2013; **2**: 126–127.
2 Zendejas B, Cook DA, Bingener J, Huebner M, Dunn WF, Sarr MG *et al*. Simulation-based mastery learning improves patient outcomes in laparoscopic inguinal hernia repair: a randomized controlled trial. *Ann Surg* 2011; **254**: 502–509.
3 Crochet P, Aggarwal R, Dubb SS, Ziprin P, Rajaretnam N, Grantcharov T *et al*. Deliberate practice on a virtual reality laparoscopic simulator enhances the quality of surgical technical skills. *Ann Surg* 2011; **253**: 1216–1222.
4 Gallagher AG, Ritter EM, Champion H, Higgins G, Fried MP, Moses G *et al.* Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005; **241**: 364–372.
5 Black SA, Nestel DF, Kneebone RL, Wolfe JH. Assessment of surgical competence at carotid endarterectomy under local anaesthesia in a simulated operating theatre. *Br J Surg* 2010; **97**: 511–516.
6 Willaert WI, Aggarwal R, Van Herzeele I, Plessers M, Stroobant N, Nestel D *et al.* Role of patient-specific virtual

reality rehearsal in carotid artery stenting. *Br J Surg* 2012; **99**: 1304–1313.
7 Keshtgar MR, Chicken DW, Waddington WA, Raven W, Ell PJ. A training simulator for sentinel node biopsy in breast cancer: a new standard. *Eur J Surg Oncol* 2005; **31**: 134–140.
8 Cocker DM, Mavroveli S, Leff DR, Hanna G. Do United Kingdom breast surgeons feel training needs updating? *Eur J Surg Oncol* 2012; **38**: 1132–1133.
9 Bates T Cecil E, Greene I. The effect of the EWTD on training in general surgery. *Ann R Coll Surg Engl* 2007; **89**(Suppl): 106–109.
10 Beard JD. Assessment of surgical skills of trainees in the UK. *Ann R Coll Surg Engl* 2008; **90**: 282–285.
11 Healthcare Quality Improvement Partnership. *Consultant Outcomes Publication*. http://www.hqip.org.uk/consultant-outcomes-publication/ [accessed 28 June 2015].
12 Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. *Ann Surg* 2008; **247**: 372–379.
13 Nimmons GL, Chang KE, Funk GF, Shonka DC, Pagedar NA. Validation of a task-specific scoring system for a microvascular surgery simulation model. *Laryngoscope* 2012; **122**: 2164–2168.
14 Glarner CE, McDonald RJ, Smith AB, Leverson GE, Peyre S, Pugh CM *et al.* Utilizing a novel tool for the comprehensive assessment of resident operative performance. *J Surg Educ* 2013; **70**: 813–820.
15 Hopmans CJ, den Hoed PT, van der Laan L, van der Harst E, van der Elst M, Mannaerts GH *et al*. Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. *Surgery* 2014; **156**: 1078–1088.
16 Hance J, Aggarwal R, Stanbridge R, Blauth C, Munz Y, Darzi A *et al.* Objective assessment of technical skills in cardiac surgery. *Eur J Cardiothorac Surg* 2005; **28**: 157–162.
17 Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C *et al.* Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 1997; **84**: 273–278.
18 Lovrics PJ, Cornacchi SD, Farrokhyar F, Garnett A, Chen V, Franic S *et al.* Technical factors, surgeon case volume and positive margin rates after breast conservation surgery for early-stage breast cancer. *Can J Surg* 2010; **53**: 305–312.
19 Zork NM, Komenaka IK, Pennington RE Jr, Bowling MW, Norton LE, Clare SE *et al*. The effect of dedicated breast surgeons on the short-term outcomes in breast cancer. *Ann Surg* 2008; **248**: 280–285.
20 Cochrane RA, Valasiadou P, Wilson AR, Al-Ghazal SK, Macmillan RD. Cosmesis and satisfaction after breast-conserving surgery correlates with the percentage of breast volume excised. *Br J Surg* 2003; **90**: 1505–1509.
21 Chan SW, Cheung PS, Lam SH. Cosmetic outcome and percentage of breast volume excision in oncoplastic breast conserving surgery. *World J Surg* 2010; **34**: 1447–1452.
22 Hseino H, Nugent E, Lee MJ, Hill AD, Neary P, Tierney S *et al.* Skills transfer after proficiency-based simulation training in superficial femoral artery angioplasty. *Simul Healthc* 2012; **7**: 274–281.
23 Schout BM, Ananias HJ, Bemelmans BL, d'Ancona FC, Muijtjens AM, Dolmans VE *et al*. Transfer of cysto-urethroscopy skills from a virtual-reality simulator to the operating room: a randomized controlled trial. *BJU Int* 2010; **106**: 226–231.

Supporting information

Additional supporting information may be found in the online version of this article:

Video S1 Simulator for training and assessment of technical skills in oncoplastic wide local excision (avi/wmv file)

Table S1 Procedure-specific global rating scales by operator grade, subcategorized according to the operative phase of the simulated wide local excision procedure (Word document)

Table S2 Pairwise comparisons in technical skills, subcategorized according to the operative phase of the simulated wide local excision procedure (Word document)

Table S3 Cronbach α analysis of inter-rater reliability, subcategorized according to the operative phase of the wide local excision procedure (Word document)

Table S4 Cohen's κ coefficient for inter-rater agreement, subcategorized according to the operative phase of the wide local excision procedure (Word document)