Plastic Surgeon Expertise in Predicting Breast Reconstruction Outcomes for Patient Decision Analysis

Clement S. Sun, MS*†
Gregory P. Reece, MD†
Melissa A. Crosby, MD†
Michelle C. Fingeret, PhD†‡
Roman J. Skoracki, MD†
Mark T. Villa, MD†
Matthew M. Hanasono, MD†
Donald P. Baumann, MD†
David W. Chang, MD†
Scott B. Cantor, PhD§
Mia K. Markey, PhD*¶

Background: Decision analysis offers a framework that may help breast cancer patients make good breast reconstruction decisions. A requirement for this type of analysis is information about the possibility of outcomes occurring in the form of probabilities. The purpose of this study was to determine if plastic surgeons are good sources of probability information, both individually and as a group, when data are limited.
Methods: Seven plastic surgeons were provided with pertinent medical information and preoperative photographs of patients and were asked to assign probabilities to predict number of revisions, complications, and final aesthetic outcome using a questionnaire designed for the study. Logarithmic strictly proper scoring was used to evaluate the surgeons' abilities to predict breast reconstruction outcomes. Surgeons' responses were analyzed for calibration and confidence in their answers.
Results: As individuals, there was variation in surgeons' ability to predict outcomes. For each prediction category, a different surgeon was more accurate. As a group, surgeons possessed knowledge of future events despite not being well calibrated in their probability assessments. Prediction accuracy for the group was up to 6-fold greater than that of the best individual.
Conclusions: The use of individual plastic surgeon–elicited probability information is not encouraged unless the individual's prediction skill has been evaluated. In the absence of this information, a group consensus on the probability of outcomes is preferred. Without a large evidence base for calculating probabilities, estimates assessed from a group of plastic surgeons may be acceptable for purposes of breast reconstruction decision analysis. (*Plast Reconstr Surg Glob Open 2013;1:e78; doi: 10.1097/GOX.0000000000000010; Published online 26 November 2013.*)

Given that a woman can be a candidate for multiple medically appropriate methods of breast reconstruction, making a decision on which method to pursue can be a difficult one.[1] Normative decision-making is a framework that could be used to help women make difficult decisions about breast reconstruction. Normative decision-making is an exhaustive, iterative process that involves identifying alternatives, obtaining information about the uncertainty of the outcomes, and clarifying prefer-

ences and values.[2–4] For the case of breast reconstruction, the alternatives, that is, different reconstruction procedures, are very well understood.[1,5,6] However, information about the uncertainties of the outcomes (eg, number of revisions needed, chances of experiencing a complication, or the final aesthetic result) is more difficult to obtain because large quantities of data may not exist for uncommon procedures or rare patient profiles. Clarifying preferences and values about breast reconstruction is also challenging[7–9]; however, the focus of this study is on the difficulty of estimating the probabilities of reconstruction outcomes. Such probabilities may be employed by future computational decision support systems to aid in patient decision-making and may make use of surgeon predictions obtained before system deployment.

In the clinical decision-making literature, it is suggested that, in the absence of large quantities of data, probabilities of outcomes can be estimated by experts,[10–12] that is, plastic surgeons in the case of breast reconstruction. Of course, estimating probabilities about breast reconstruction outcomes is difficult because there are numerous variables involved. Moreover, to provide the information needed for decision support, plastic surgeons must be able to estimate outcome probabilities in general, not simply for their own patients. However, we are unaware of any prior studies that address the validity of expert-elicited probabilities about breast reconstruction outcomes.

The clinical impact of this study is in the context of a future computational decision support system for shared breast reconstruction decision-making. Such systems require patient-specific probability information that surgeons may provide. The goal of this study was to investigate to what extent plastic surgeons can predict breast reconstruction outcomes. The purpose of the system is to inform both the patient and the surgeon of risk and help them make better decisions regarding breast reconstruction.

MATERIALS AND METHODS

Participants

Of 19 faculty members in the Department of Plastic Surgery at The University of Texas MD Anderson

Cancer Center, 7 were willing and able to submit completed questionnaires. These surgeons perform 21–74 breast reconstruction surgeries per year and possess 6–23 years of experience performing breast reconstruction.

The study utilized photographs and health records of 10 women aged 21 or older who underwent breast reconstruction between January 1, 2004, and March 30, 2012, while participating in an ongoing prospective study at MD Anderson. All eligible patients had both *prereconstruction* and *postreconstruction* photographs on file in the study database. Prereconstruction was defined as not having begun the breast reconstruction process. If a patient had previous mastectomy or breast conservative therapy, so long as it preceded reconstruction, their status was considered prereconstruction. Postreconstruction photographs were defined by the presence of both breast mounds after breast reconstructive surgery. If the patient completed breast mound reconstruction but declined additional reconstructive procedures, nipple reconstruction or areola micropigmentation was not a requirement. Of 180 patients available in the study database, only 10 met the eligibility criteria and all 10 were included in this study.

A Coolpix 8400 (Nikon, Melville, N.Y.) or EOS Rebel T1i (Canon U.S.A., Lake Success, N.Y.) digital camera was used to acquire digital photographs of patients posing in the standard 5 views, that is, anterior-posterior, left-lateral, right-lateral, left-oblique, and right-oblique, as recommended in the photographic standards set forth by the American Society of Plastic Surgeons for patients undergoing a transverse rectus abdominis myocutaneous flap reconstruction.[13] We obtained institutional review board approval from The University of Texas MD Anderson Cancer Center to conduct this research, which included a waiver of informed consent to obtain the patient information.

Multiple-choice Questionnaire

A multiple-choice questionnaire was prepared, with the help of clinical experts G.P.R. and M.A.C. (Fig. 1),[14] who were blinded from the answers. For each patient, we provided prereconstruction photographs of the standard 5 views for visual information. Pertinent information from the medical records included age at the time the photographs were taken, diagnosis, race, ethnicity, body mass index, history of radiation therapy, history of chemotherapy, smoking status, comorbidities that may affect breast reconstruction, the ultimate method of reconstruction selected by the patient, and the
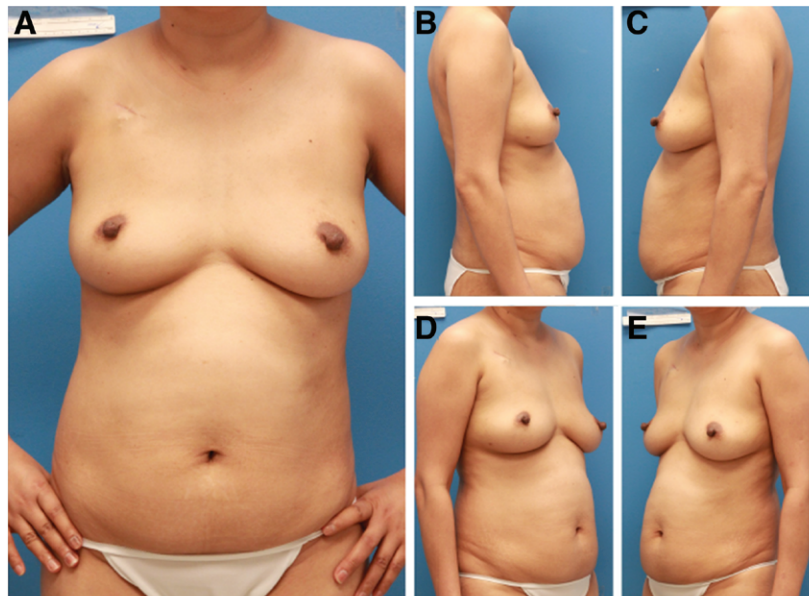
F Patient 165 is a 42-year-old, non-smoker, non-Hispanic Asian female diagnosed with invasive ductal carcinoma of the left breast. She will be treated with a nipple-sparring mastectomy, IOLM, and a sentinel node biopsy. Her BMI is 22.5. She has no relevant co-morbidities. She has had chemotherapy (12 weeks of herceptin and paclitaxel and 4 cycles of FEC) but no radiation therapy.

She has elected to undergo **immediate unilateral TRAM reconstruction** of the left breast.

The reconstruction will be performed by Dr. ▮▮▮▮ who performs approximately 75 breast reconstruction cases a year and has 5 years of experiences practicing this form of breast reconstruction.

Please use photographs and patient history to make the following predictions.

28) Including nipple reconstruction and any contralateral breast procedure(s), how many revisions will she have? (A revision is any procedure that improves the reconstructive aesthetic outcome or required to correct complications.)

    A. None
    B. 1 or 2
    C. 3 to 6
    D. 7 or more

29) What is the worst complication she will have?

    A. None
    B. Minor
    C. Major
    D. Life-threatening

30) What will your final overall impression of her breast (i.e. the aesthetic outcome) be?

    A. Excellent
    B. Good
    C. Fair
    D. Poor

Fig. 1. A sample patient from the breast reconstruction outcome question-naire. Five prereconstruction images of the A, anterior-posterior, B and C, left and right lateral, and D and E, left and right oblique views of the patient were provided. F, In addition, a brief pertinent medical history was provided detail-ing the patient's age, smoking status, race, ethnicity, diagnosis, treatments, and relevant comorbidities. The ultimate reconstruction method and timing are provided along with the operating plastic surgeon's years of experience and approximate breast reconstruction cases performed per year. Three questions are asked regarding expected revisions, complications, and final aesthetic out-come. Each question has 4 possible responses.

experience of the plastic surgeon who performed the reconstruction in terms of years of practice and cases performed per year.

Three types of prediction questions were posed regarding (1) the ultimate number of revisions, (2) the worst complication experienced by the patient,

and (3) the final overall (aesthetic) impression of the reconstructed breasts. Each question had 4 responses (Table 1). The surgeons were asked to assign a probability to each response with the sum of the probabilities equal to 100%.

Questionnaire Answer Key

The actual answers for the number of revisions and worst complication were retrieved from the medical record. The final overall (aesthetic) impression was not recorded in the medical record; therefore, the postreconstruction 5 standard view photographs were shown to 4 clinical experts (R.J.S., M.T.V., M.M.H., and M.A.C.) after all questionnaires were submitted, who then independently and subjectively rated the final appearance on a positively oriented 0-to-10 Likert-like scale with 10 as the best reconstructed appearance and 0 the worst. The 4 clinical experts demonstrated outstanding agreement in their aesthetics ratings, with an intraclass correlation coefficient of 0.82 ($P < 0.0001$). The average of their ratings was used as the correct response.

Measuring Predictive Skill and Confidence

Plastic surgeon expertise in predicting breast reconstruction outcomes was assessed from the questionnaire using a *strictly proper logarithmic scoring function.*[15] The scoring function takes the surgeons' responses and produces a score, which is maximized by being honest (ie, expressing a probability that one truly believes).[16,17] The logarithmic scoring function is favored because it adheres to the likelihood principle where the only thing that matters is the likelihood of events that occur, not the likelihood of events that do not occur.[18] In addition, higher probabilities assigned to the correct response always result in a higher score.[17] Thus, the resulting score can be treated as a measure of the plastic surgeon's ability to accurately predict breast reconstruction outcomes.

The scoring function was designed with the constraints that perfect prediction of reconstruction outcomes would correspond to a maximum total score of 100 and complete uncertainty about the reconstruction outcomes would yield a score of zero. The logarithmic scoring function is written as:

$$Score = \frac{100}{number\ of\ questions} + \frac{100 \times \ln(correct\ probability)}{number\ of\ questions \times \ln(possible\ responses)} \quad (1)$$

where the total number of questions was 30 and there were 4 possible responses. In the event that the probabilities did not sum to one, the surgeons' probabilities were normalized. A positive score indicates that the plastic surgeon is able to accurately predict some aspects of breast reconstruction outcomes. A score of zero represents a benchmark of no ability to predict breast reconstruction outcomes, that is, the equivalent of a layperson submitting a blank questionnaire. A negative score indicates poor prediction ability or potential harm resulting from inaccurate predictions of breast reconstruction outcomes. The

Table 1. Definitions for Responses to Breast Reconstruction Outcome Questions

Question Type	Response	Definition
Final overall impression (aesthetic outcome)	Excellent	9 or 10 on a scale from 0 to 10 (a score of 10 would be considered the perfect reconstruction)
	Good	7 or 8 on a scale from 0 to 10
	Fair	4, 5, or 6 on a scale from 0 to 10
	Poor	0, 1, 2, or 3 on a scale from 0 to 10 (a score of 0 would be considered the worst reconstruction imaginable)
Complications	None	No complications
	Minor	Requires local wound care
	Major	Requires an invasive procedure (eg, large amounts of mastectomy skin necrosis, hematoma, seroma, expander or implant rupture/leakage, implant migration, extrusion, displacement, peri-implant calcification, breast asymmetry, and infection around implant)
	Life-threatening	Threatens the patient's life (eg, toxic shock syndrome, pneumothorax, deep venous thrombosis, pulmonary embolus, pneumonia, respiratory arrest, prolonged ventilation, >3 U blood loss and need for transfusion, severe sepsis, stroke, and myocardial infarct)
Revisions	None	No procedures to improve the reconstructive aesthetic outcome or correct complications
	1–2	One or two procedures to improve the reconstructive aesthetic outcome or correct complications
	3–6	Three to as many as 6 procedures to improve the reconstructive aesthetic outcome or correct complications
	>7	Seven or more procedures to improve the reconstructive aesthetic outcome or correct complications

maximum score for each question is 3.333 and the minimum is negative infinity. Assigning a 0% probability to any response was discouraged because if 0% is assigned to what is actually the true outcome, the score will be negative infinity.

We also assessed the *confidence* or self-perceived ability of each plastic surgeon in making predictions. Confidence may be measured by the Shannon entropy (*H*), which is calculated from the probabilities across all responses for each question where

$$H(\mathbf{p}) = \sum_i p_i \ln p_i \qquad (2)$$

and where **p** is the set of probabilities assigned to the responses of a question.[19] As the entropy approaches zero, one can say that the plastic surgeon is more confident in their predictions. As the Shannon entropy approaches the natural log of 0.25 or approximately −1.39, one can say the plastic surgeon is less confident in their predictions.

Plastic Surgeon Group Calibration

With 7 participating surgeons, each assigning a probability to 4 responses for each of 30 questions, we examined the calibration of the surgeons as a group using 840 probability assessments. In other words, when the surgeons said something would happen *x* percent of the time, how often did it really happen? We determined a calibration curve by creating 10 equal-sized bins for the surgeons' probability assessments from 0% to 100% and plotting the frequency of correct assessments.

Combining Plastic Surgeons' Predictions

A simple approach to combining the predictions of different surgeons would be to calculate the average of their individual predictions. Calculating the average treats the surgeons as if they are equally skilled in making predictions. However, if we know that the different surgeons are not all equally skilled at making predic-

tions, then alternatively we can calculate a *weighted* average in which the amount that each surgeon contributes to the average is scaled by a *weight* that may be viewed as a "probability of prediction expertise."[14]

We use the Roberts' method[20] to calculate the weight of each surgeon based on their prediction ability in relation to their peers. The final weight of surgeon *i*, w_i, over all *N* predictions is represented by:

$$w_i = \frac{w^0 \prod_{k=1}^{N} p_i^k}{\sum_{j=1}^{M}\left(w^0 \prod_{k=1}^{N} p_j^k\right)}, \; w^0 = \frac{1}{M} \qquad (3)$$

where p_i is the probability assigned to the correct response and the denominator is the sum of the product of correct probabilities and weights for all *M* surgeons. The group score based on the combined surgeons' predictions can then be calculated from the sum of the weighted, correct probabilities.

RESULTS

The demographic, surgical, and clinical characteristics of the patient sample are summarized in Table 2. Regarding final aesthetic outcomes, individual ratings by plastic surgeons ranged from a perfect 10 to as low as 2. The mean individual rating was 5.7 with a standard deviation of 2.1.

We de-identified the surgeon results in 2 ways because surgeons could be identified by the number of *years of experience* performing breast reconstruction and number of *cases performed per year* (Table 3).

Individually, the plastic surgeons demonstrated incomplete knowledge with an average overall score of −9.59. However, a few earned a positive score; the best scored 5.63. The lowest score was −32.59 points. Overall, the plastic surgeons were better at predicting the number of revisions with an average score of 7.74. The score range in the revisions category was

Table 2. Patient Participant Demographics

Patient	Age	Race	Hispanic	BMI	Chemotherapy	Radiation Therapy	Reconstruction Method and Timing
1	47	African American	No	29.9	Yes	Yes	Delayed unilateral TRAM
2	54	White	No	25.9	Yes	No	Delayed unilateral TRAM
3	54	White	No	22.4	No	No	Immediate bilateral implant
4	68	White	No	33.5	Yes	No	Immediate unilateral implant
5	43	White	Yes	20.6	No	No	Delayed unilateral LatDM
6	39	White	No	25.5	Yes	Yes	Delayed unilateral TRAM
7	53	White	No	32.9	Yes	Yes	Delayed unilateral TRAM
8	58	White	No	30.2	Yes	Yes	Delayed unilateral implant and TRAM
9	40	Asian	No	26.1	Yes	No	Immediate bilateral TRAM
10	42	Asian	No	22.5	Yes	No	Immediate unilateral TRAM
Mean	49.8	—	10% Yes	27.0	80% Yes	40% Yes	—

BMI, body mass index; LatDM, latissimus dorsi myocutaneous; TRAM, transverse rectus abdominis muscle.

Table 3. Plastic Surgeon Demographics, Ordered by Years of Experience Performing Breast Reconstruction

Surgeon	Years of Experience	Cases Per Year	Receive Feedback
A	6	53	No
B	6	57	Yes
C	6	74	No
D	7	21	Yes
E	8	34	Yes
F	17	69	Yes
G	23	38	Yes
Mean	10.1	49.4	71.4% Yes

between 17.69 and −3.74. In predicting the worst complication, they did not perform well with a mean score of −6.45, a maximum of 12.99, and a minimum of −26.84. Performance was worst for predicting the aesthetic outcome with an average score of −10.88, a maximum of 0.30, and a minimum of −22.76 (Table 4). Surgeons' knowledge scores were plotted against their confidence (Fig. 2). There was no correlation between surgeon experience and confidence. However, we did observe a significant negative correlation between aesthetic outcome score and confidence ($R^2 = −0.9017$, $P = 0.0055$).

As a group, surgeons were not well calibrated in assessing the probability of future events (Fig. 3). They tended to underestimate the probability of unlikely events when they assigned a probability between 0% and 30% and overestimated between the 31% and 100% range. The frequency of high-probability assessments was small; there were only 2 assessments in the 91–100% range.

Different surgeons were best able to predict the different types of outcomes. For predicting revisions, surgeon 7 was most accurate and, consequently, their predictions were highly weighted (weight = 0.87) in the weighted average. By contrast, surgeon 6 was most accurate at predicting complications (weight = 0.99), whereas surgeon 2 was most accurate at predicting the aesthetic outcome (weight = 0.83). We

took the average of the correct probabilities across all questions and compared it to the weighted consensus. The weighted consensus yielded higher scores in the revisions and complications category and overall. Notably, if the best scoring surgeons from each category combined their scores, it would not surpass the weighted consensus. As the best individual overall score was 5.63, the weighted consensus produced a respectable score (Table 5).

DISCUSSION

Interestingly, there may be different expertise domains in outcome prediction in breast reconstruction surgery concerning, at the very least, the revisions, complications, and aesthetic outcomes. No one surgeon may possess the ability to accurately predict across all 3 outcome categories, and no individual can be assumed reliable.

As a group, however, the plastic surgeons in this study performed surprisingly well, predicting reconstruction outcomes with some accuracy. The equal-weighted consensus yielded a positive score of more than 18. One might think that a score of 18.74 out of 100 is a poor score, but be reminded of the difficulty of making predictions. A score of 100 corresponds to perfect prediction, which is not realistic for real-world tasks. Predicting breast reconstruction outcomes is particularly challenging because numerous factors must be taken into consideration (ie, whether a patient will have postoperative radiation therapy) Furthermore, the weighted consensus yielded an even better score at 32.94, which is quite reasonable. This suggests that, as a group, plastic surgeons may be approached for their combined probabilities.

The surgeons did not perform well in predicting aesthetic outcome. Individually and as a group, this was the worst prediction category, though the group score was one order of magnitude greater than the individual average. On examining surgeons' confidence, we found a significant negative correlation between the score and surgeons' confidence with

Table 4. Summary of Surgeon Prediction Scores and Confidence

Surgeon	Overall		Revisions		Complications		Surgeon Rated Aesthetic Outcome	
	Score	Entropy	Score	Entropy	Score	Entropy	Score	Entropy
1	−18.42	−0.74	7.74	−0.78	−8.28	−0.68	−17.88	−0.76
2	1.00	−0.99	8.84	−0.95	−8.14	−0.96	0.30	−1.04
3	5.63	−0.99	10.29	−0.99	2.77	−0.96	−7.44	−1.03
4	−32.59	−0.77	11.13	−0.60	−26.84	−0.88	−16.88	−0.82
5	−9.59	−0.95	2.23	−0.89	−6.74	−0.95	−5.08	−1.02
6	2.83	−0.81	−3.74	−0.75	12.99	−0.78	−6.42	−0.90
7	−15.97	−0.76	17.69	−0.70	−10.91	−0.77	−22.76	−0.79
Mean	−9.59	−0.86	7.74	−0.81	−6.45	−0.86	−10.88	−0.91

Surgeons were ordered randomly. Overall, on average, surgeons were not able to accurately predict outcomes. However, in distinct categories, particularly in revisions, they did display some prediction ability. Each category also had a different high-scoring surgeon.
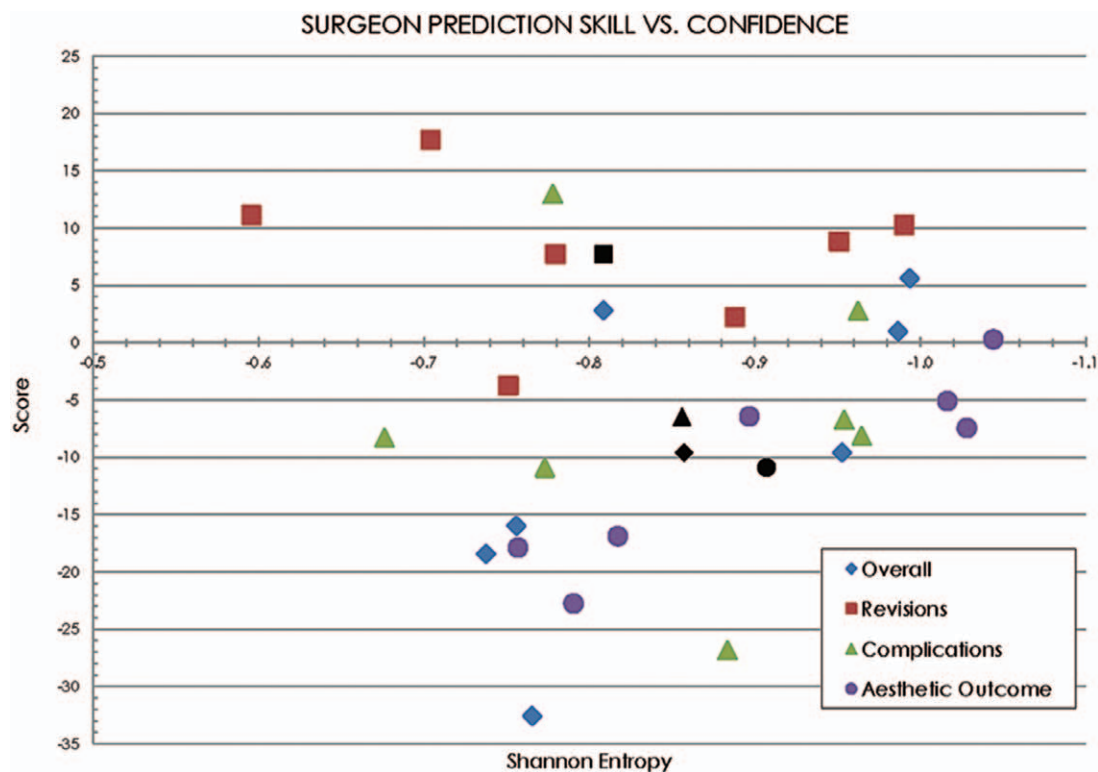
Fig. 2. Surgeons' prediction scores were plotted against confidence. The black markers represent the means of the corresponding colored markers. Scores above zero indicate increasing prediction ability and the farther from zero entropy, the less confident a surgeon was in his or her probability assessments. Overall, surgeons displayed poor predictive ability, especially in predicting aesthetic outcome. However, they did remarkably well predicting the number of expected revisions. Note that the highest scorers were not the most confident.

regard to the aesthetic outcome. There is evidence that probability exercises and feedback may reduce this overconfidence bias,[21] improving scores in turn. However, these results are not surprising, as "aesthetics" is a poorly defined concept and understood vaguely. On examining the distribution of probabilities across all responses, we noticed that the surgeons tended to be optimistic of the final aesthetic result: overestimating the probability of better results when poorer results actually occurred.

Regarding calibration, the results are not surprising. People in general, even the brightest scientists or physicians, have limited experience with probability assessment.[22] Previous studies[23–25] have concluded that physician-group probabilities are not well calibrated, with a similar tendency to underestimate the probability of less likely events and vice versa. Thus, our results are consistent with what would be expected from the prior literature given that this was a novel prediction task and that surgeons are subject to the same cognitive biases as are all people. "One such bias in probability assessment is anchoring and adjustment whereby people start with a reference (the "anchor," eg, a 20% chance of a fair aesthetic outcome) and

make incremental adjustments to reach an estimate (which is often insufficient, eg, adjusting to 22% when the adjustment should have been more)."[26]

The goal of this study was only to examine the prediction of reconstruction outcomes *in general*, as opposed to each surgeon's prediction of results for their own patients. Although that would also be an interesting question, it is not possible to retrospectively perform such a study on each surgeon's own patients because of prior knowledge. Furthermore, such a study would be of limited value for generalizable decision support that is non–surgeon-specific.

We did not observe any correlation between the surgeon's ability to predict reconstruction outcomes and the surgeons' experience or between confidence and experience, but the sample size was too limited to draw strong conclusions on this point. For instance, there were 2 tight groupings in years of experience: 1 group with between 6 and 8 years comprised 5 surgeons and another group with between 17 and 23 years comprised 2 surgeons.

The influence of patient factors, such as reconstruction timing, laterality, and race, may play a role in the ability of surgeons to make predictions. In-
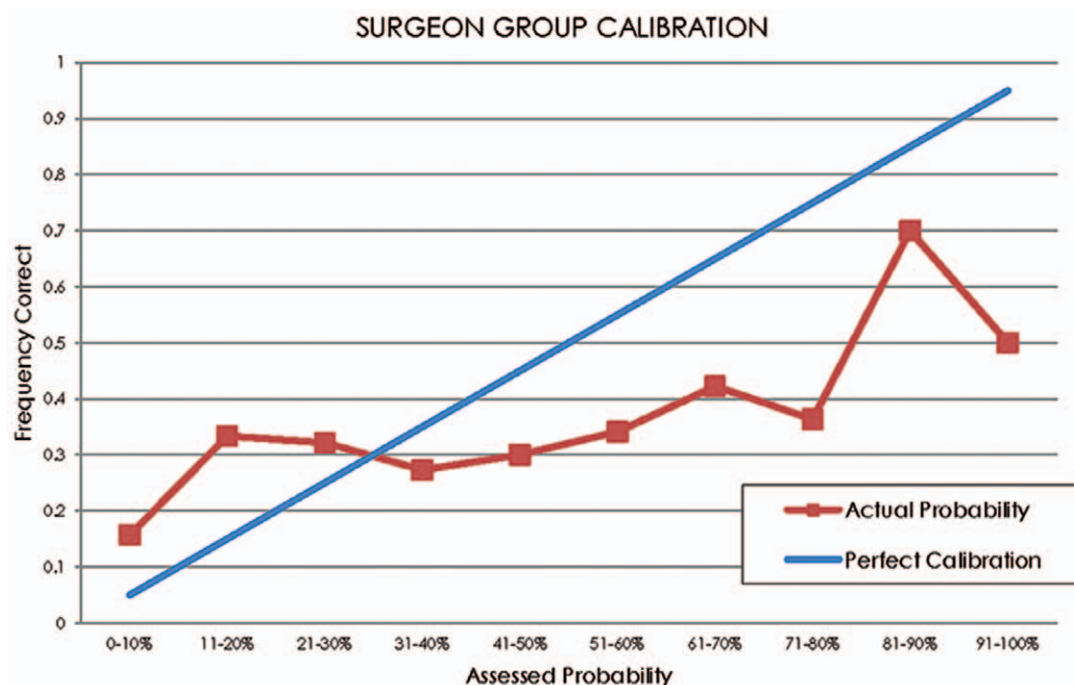
Fig. 3. Surgeon group calibration curve over 840 assessments.

deed, we found several significant variables and others that trend toward significance.

The questionnaire had inherent limitations in that the surgeons were only given superficial exposure to patients on which to form their predictions. However, such information, in the form of photographs and short histories, represents the most data reasonably available for making predictions. It would be infeasible to have multiple surgeons consult the same patient for a prospective study. Although the sample sizes may appear small, they are comparable to similar studies.[14,23–25] It is not likely that including additional patients or surgeons would change the overall findings.

CONCLUSIONS

We found that plastic surgeons as a group, and certain individuals therein, demonstrated the ability to accurately predict some breast reconstruction outcomes: the number of revisions, the worst complication, and the aesthetic results. They performed remarkably well at predicting the number of revisions, but less well with the other categories, particularly aesthetic outcome. The plastic surgeons were not equally able to predict across the outcome categories. Nor were the plastic surgeons, as a group, calibrated in making probability assessments. They tended to underestimate the probability of low likelihood events and overestimate the probability of more likely events. Their exact proficiency may be masked by limitations of the study including its retrospective nature, small sample size, the novelty of the task at hand, vulnerability to certain cognitive biases, and imprecise nature of quantifying aesthetic outcome.

We concluded that several surgeons should be consulted for probabilities rather than relying on 1 expert alone. The latter strategy is not preferred when predicting complications and aesthetic outcome as most surgeons were not able to accurately predict reconstruction outcomes. Ascertaining

Table 5. Comparing the Prediction Scores by Best Surgeon, Equal-weighted Average Consensus, and Weighted Consensus

	Best Surgeon Overall	Best Surgeon by Category	Average Consensus	Weighted Consensus
Revisions	10.29	17.68	14.68	18.36
Complications	2.77	12.99	1.76	13.02
Aesthetic outcome	−7.44	0.30	2.35	1.56
Overall	5.63	30.97	18.79	32.94

The weighted consensus yielded a respectable overall score higher than the equal-weighted average consensus and sum of best surgeon scores from each prediction category.

8

the weight of each surgeon's opinion may improve group prediction performance. Nonetheless, plastic surgeons' predictions in the form of expert-elicited probabilities may be used for breast reconstruction patient decision analysis.

Mia K. Markey, PhD
Department of Biomedical Engineering
The University of Texas at Austin
107 W. Dean Keeton
Stop C0800
BME 3.314
Austin
TX 78712-1801
E-mail: mia.markey@utexas.edu

REFERENCES

1. Crosby MA. *Reshaping You.* Houston, TX: The University of Texas M. D. Anderson Cancer Center Patient Education Office; 2010.
2. Howard RA. Microrisks for medical decision analysis. *Int J Technol Assess Health Care.* 1989;5:357–370.
3. Clemen RT. *Making Hard Decisions: An Introduction to Decision Analysis.* Boston, MA: Duxbury; 1997.
4. Davis Sears E, Chung KC. Decision analysis in plastic surgery: a primer. *Plast Reconstr Surg.* 2010;126:1373–1380.
5. Sun CS, Wang D, Lee J, et al. Towards a decision basis of breast reconstruction: defining the alternatives. *Poster presented at: American Medical Informatics Association Annual Symposium.* Washington, DC, October 25, 2011.
6. Bostwick J III. *Plastic and Reconstructive Breast Surgery.* St. Louis, MO: Quality Medical Publishing; 1999.
7. Keeney RL, Raiffa H. *Decisions With Multiple Objectives: Preferences and Value Tradeoffs.* New York, NY: John Wiley & Sons; 1976.
8. Keeney RL. *Value-Focused Thinking.* Cambridge, MA: Harvard University Press; 1992.
9. Begum S, Grunfeld EA, Ho-Asjoe M, et al. An exploration of patient decision-making for autologous breast reconstructive surgery following a mastectomy. *Patient Educ Couns.* 2011;84:105–110.
10. Weinstein MC, Fineburg HV. *Clinical Decision Analysis.* Philadelphia, PA: W.B. Sanders Company; 1980.
11. Hunink MGM. *Decision Making in Health and Medicine.* New York, NY: Cambridge University Press; 2001.
12. Sox HC, Blatt MA, Higgins MC, et al. *Medical Decision Making.* Philadelphia, PA: The American College of Physicians; 2007.
13. American Society of Plastic Surgeons. *Photographic Standards in Plastic Surgery.* Available at: http://trebol.sirexmedica.com/download/Photographic_Standards.pdf. Accessed July 1, 2011.
14. Bickel JE. Scoring rules and decision analysis education. *Decis Anal.* 2010;7:346–357.
15. Bickel JE. Some comparisons between quadratic, spherical, and logarithmic scoring rules. *Decis Anal.* 2007;4:49–65.
16. Winkler RL. Scoring rules and the evaluation of probability assessors. *J Am Stat Assoc.* 1969;64:1073–1078.
17. Shuford EH Jr, Albert A, Massengill HE. Admissible probability measurement procedures. *Psychometrika* 1966;31:125–145.
18. Lindley DV. The philosophy of statistics. *Statistician* 2000;49:293–337.
19. Shannon CE. A mathematical theory of communication. *Bell Syst Tech J.* 1948;27:379–423.
20. Roberts HV. Probabilistic prediction. *J Am Stat Assoc.* 1965;60:50–62.
21. Alpert M, Raiffa H. A progress report on the training of probability assessors. In: Kahneman D, Slovic P, Tversky A, eds. *Judgement Under Uncertainty: Heuristics and Biases.* New York, NY: Cambridge University Press; 1982:294–305.
22. Hora S. Eliciting probabilities from experts. In: Edwards W, Miles RF Jr, von Winterfeldt D, eds. *Advances in Decision Analysis: From Foundations to Applications.* New York, NY: Cambridge University Press; 2007:129–153.
23. Kharbanda AB, Fishman SJ, Bachur RG. Comparison of pediatric emergency physicians' and surgeons' evaluation and diagnosis of appendicitis. *Acad Emerg Med.* 2008;15:119–125.
24. Minne L, DeJonge E, Abu-Hanna A. Repeated prognosis in the intensive care: how well do physicians and temporal models perform? In: *Proceedings of the 13th Conference on Artificial Intelligence in Medicine.* Bled, Slovenia, July 2–6, 2011.
25. Christensen-Szalanski JJ, Bushyhead JB. Physician's use of probabilistic information in a real clinical setting. *J Exp Psychol Hum Percept Perform.* 1981;7:928–935.
26. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974;185:1124–1131.