Abstract

Background: Patient outcomes in aesthetic breast surgery are highly dependent on breast measurements used in preoperative planning.

Objectives: The purpose of this study is to determine the reliability of anthropometric breast measurements.

Methods: Four raters measured 28 women using 7 measurements: sternal notch to nipple distance (Sn-N), nipple to midline (N-M), nipple to infra-mammary-fold distance under maximal stretch (N-IMF), breast base width (BW), soft tissue pinch thickness of the upper pole (STPT:UP), STPT at the infra-mammary fold (STPT:IMF), and anterior pull skin stretch (APSS). Reliability was assessed using intra-class correlation coefficients (ICCs).

Results: Inter-rater reliability was excellent for Sn-N, N-M, and BW (ICC = 0.94, 0.90, and 0.76, respectively) and was good for N-IMF (ICC = 0.70). The STPT:UP, STPT:IMF, and APSS measurements were not reliable between raters (ICC < 0.2). Intra-rater reliability was excellent for Sn-N, N-M, and BW for all raters (all ICC > 0.75). The N-IMF intra-rater reliability was excellent in senior raters (ICC > 0.75) and good in junior raters (ICC > 0.6). The STPT:UP, STPT:IMF, and APSS measurements showed fair or poor reliability for most raters (ICC < 0.6).

Conclusions: The Sn-N, N-M, and BW measurements are very reliable. Dynamic measurements including APSS, STPT:UP, and STPT:IMF are unreliable. N-IMF is the only reliable dynamic measurement, and its reliability improves with increasing clinical experience. The variable reliability of preoperative measurements must be considered in the planning of aesthetic breast surgery.

Level of Evidence: 4

Breast augmentation is one of the most-frequently performed cosmetic surgical procedures in the world. The goal of breast augmentation is an aesthetically pleasing breast with minimal risk of complications and reoperation. According to the 2011 US Food and Drug Administration Update on the Safety of Silicone Gel-Filled Implants, reoperation following primary breast augmentation is above 20%. Poor preoperative planning is an important factor contributing to the high rates of complications and reoperations.

In an attempt to optimize outcomes and minimize reoperation rates in breast augmentation, industry stakeholders and plastic surgeons have developed algorithmic approaches to guide appropriate preoperative planning. These approaches are based on the objective evaluation of the breast and include characteristics of the breast size, shape, and soft-tissue coverage overlying the parenchyma. These quantitative assessments are thought to be important in describing the dimensions of the breast, quantifying soft-tissue characteristics, and identifying asymmetries in preoperative planning. This approach has been termed tissue-based planning. With tissue-based planning, many surgeons utilize breast measurements to define the essential tissue characteristics of the envelope, parenchyma, dimensions, and fill distribution, which may affect short and long-term outcomes.

From the Division of Plastic and Reconstructive Surgery, Department of Surgery, University of Toronto, Toronto, Ontario, Canada.

Corresponding Author: Dr Kathryn V. Isaac, Division of Plastic and Reconstructive Surgery, Department of Surgery, University of Toronto, 149 College Street, 5th Floor, Toronto, ON M5T 1P5, Canada. E-mail: kathryn.isaac@mail.utoronto.ca
results in breast augmentation. The breast measurements are the foundation upon which critical preoperative surgical decisions are made, namely the implant size, the implant dimensions, the implant fill, the implant surface, and the choice of surgical pocket. The implementation of tissue-based planning results in the reduction of reoperation rates and improved patient outcomes. Furthermore, quantitative methods for the assessment of breast shape and volume are considered a requirement for obtaining a pleasing aesthetic result in breast-augmentation surgery.

Any system for the selection of an implant based on patient measurements must be reproducible. For tissue-based planning to predictably and successfully reduce adverse outcomes, it must be transferable to plastic surgeons performing breast augmentation. Measurements required for the objective evaluation of the breast must be reliable and precise.

Reliability describes how consistently values are measured. The reliability of a measurement is an important quality but is distinct from accuracy, which defines how well the values we measure reflect the true value. There is no gold standard to determine the true or accurate value of a breast measurement. However, the reliability of a measurement can be determined by comparing the results obtained by different raters. The reliability of a measurement is critical for its utility and application. To our knowledge, there has been no investigation into the reliability of breast measurements. A high degree of reliability by a single rater and between raters is necessary if surgeons are to base critical preoperative decisions on these measurements. The objective of this study is to determine the degree of reliability with commonly performed anthropometric breast measurements in the preoperative planning of aesthetic breast surgery.

**METHODS**

After obtaining ethics approval from the North York General Hospital Institutional Review Board (REB#13-0268), 28 women were recruited through advertisements and self-referral. Informed consent was obtained from all participants. The study period for advertisement, recruitment, and data collection was approximately 6 months in duration from June 2014 to December 2014. The study was conducted in accordance with the guidelines set forth in the Declaration of Helsinki. In an attempt to simulate a population of patients requesting primary breast augmentation, the following criteria were utilized. Inclusion criteria included English-speaking women between 18 and 45 years of age with a self-reported breast size of A or B cup. Exclusion criteria included obesity (body mass index [BMI] > 30), previous breast surgery, current pregnancy or lactation in the past 12 months, and pseudoptosis. All patients completed a questionnaire including patient demographics and committed to 2 sessions prior to beginning the study (a blank copy of the questionnaire is available as Supplemental Material at www.aestheticsurgeryjournal.com). Financial compensation for time was provided for participation in the study; participants were required to attend both sessions and received the compensation at the completion of the second session.

Four raters (2 plastic surgeons and 2 senior plastic-surgery residents) were utilized in this study, and all raters had substantial experience with breast-measurement techniques through their practice or formal residency training in breast surgery. The 2 senior raters have been in practice for 10 and 20 years and have practices that evaluate more than 300 breast patients per year. The 2 junior raters are plastic-surgery residents in their fourth and fifth years of training. They have each logged a minimum of 60 breast cases per year of training. The senior raters were both men and the junior raters were one woman and one man.

Prior to commencing, each rater was provided a standard measuring tape and a caliper and underwent a review session with the senior author to ensure standardization of measurement techniques. The raters measured each breast of the 28 participants on 2 separate occasions, separated by at least 2 weeks. Seven anthropometric measurements were included in each assessment with standardized technique (Table 1): sternal notch to nipple distance (Sn-N), nipple to midline (N-M), nipple to inframammary-fold distance under maximal stretch (N-IMF), breast base width (BW), soft-tissue pinch thickness of the upper pole (STPT:UP), soft-tissue pinch thickness at the inframammary fold (STPT:IMF), and anterior pull skin stretch (APSS) (Figure 1).

Raters were blinded to the participants’ personal clinical information and demographics. The raters were aware that their measurements were being analyzed for reliability and were blinded to the measurements of other raters and their

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Method of Obtaining Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sn-N</td>
<td>Patient standing, using measuring tape</td>
</tr>
<tr>
<td>N-M</td>
<td>Patient standing, using measuring tape</td>
</tr>
<tr>
<td>BW</td>
<td>Patient standing with arms abducted at 90 degrees, using caliper from medial edge to lateral edge of breast parenchyma</td>
</tr>
<tr>
<td>N-IMF</td>
<td>Patient supine; using measuring tape, a mark was made in the IMF and then the tissue was placed under maximal stretch</td>
</tr>
<tr>
<td>STPT:UP</td>
<td>Patient standing with arms at side, using caliper with pinch of skin and subcutaneous tissue superior to the breast parenchyma</td>
</tr>
<tr>
<td>STPT:IMF</td>
<td>Patient standing with arms at side, using caliper with pinch of skin and subcutaneous tissue at the IMF</td>
</tr>
<tr>
<td>APSS</td>
<td>Patient standing with arms at side, using caliper with anterior pull of periareolar skin to maximal stretch just below the nipple</td>
</tr>
</tbody>
</table>

APSS, anterior pull skin stretch; BW, breast base width; N-IMF, nipple to inframammary-fold distance under maximal stretch; N-M, nipple to midline; Sn-N, sternal notch to nipple distance; STPT:IMF, soft-tissue pinch thickness at the inframammary fold; STPT:UP, soft-tissue pinch thickness of the upper pole.
own previous measurements of a participant. Specifically, they were not permitted to review their recorded measurements of a participant prior to the next replicate observation or for comparison with measurements obtained by another rater. Each rater performed and recorded their measurements independently.

### Statistical Analysis

Inter-rater and intra-rater reliability was assessed using intraclass correlation coefficients (ICCs). Inter-rater reliability defines the consistency in the measurements obtained by different raters using the same instrument to assess the same patients. Intra-rater reliability (also known as test-retest) defines the consistency in measurements obtained by a single rater using the same instrument to assess the same patients repeatedly at different times. The sample size required to test a reliability hypothesis at the 5% significance level with 80% power is 25 patients. Because it is generally accepted that inter-rater reliability will be lower than intra-rater reliability, we have chosen the minimal acceptable level to reflect this. Our minimal acceptable level (criterion value) defined as “almost perfect” inter-rater and intra-rater reliability was selected to calculate the sample size required. The qualitative ratings of agreement are based on ICC values. For the inter-rater reliability analysis, a 2-way mixed model, ICC (3,1), was applied and calculated for absolute agreement. For the intra-rater reliability analysis, a 1-way random model, ICC (1,1) was applied and calculated for consistency. The quality of reliability is defined by the ICC values; ICC values less than 0.40 denote poor reliability; ICC values between 0.40 and 0.59 denote fair reliability; ICC values between 0.60 and 0.74 denote good reliability; ICC values between 0.75 and 1.0 denote excellent reliability.10

### RESULTS

A total of 28 women were eligible for participation through advertisement and self-referral to our study. All volunteers met the study criteria except one participant who was excluded for her absence at the second measurement session (Figure 2). The average age of the 28 women was 23.1 years (range, 18-35 years) with an average BMI of 20.7 kg/m² (range, 18.3-26.6 kg/m²). All women were nulliparous. To replicate a breast augmentation cohort, all volunteers were A- or B-cup breast size.

Each rater measured each breast of the 28 participants to obtain the 7 anthropometric breast measurements. The mean values for the measurements performed by all raters on all participants for each measurement are listed in Table 2. The measurements were analyzed for both inter-rater and intra-rater reliability. Inter-rater reliability was excellent for the Sn-N, N-M, and BW measurements (ICC = 0.94, ICC = 0.90, and ICC = 0.76, respectively) and was within a good range for the N-IMF measurement (ICC = 0.70). The STPT:UP, STPT:IMF, and measurements were not reliable (all ICC < 0.2), suggesting a significant amount of measurement error (Table 3).

Intra-rater reliability was excellent for Sn-N, N-M, and BW for all 4 raters (all ICC > 0.75). The N-IMF intra-rater
reliability was excellent in senior raters (ICC > 0.75) and was good in junior raters (ICC > 0.6). The STPT:UP, STPT: IMF, and APSS measurements showed fair or poor reliability for all raters (ICC < 0.6), regardless of the seniority of the rater. The intra-rater reliability for all measurements improved with increasing seniority of the rater (Table 4).

**DISCUSSION**

Breast measurements are routinely performed and recorded when planning aesthetic and reconstructive breast procedures. These measurements have been used to communicate strategies for patient, procedure, and implant selection as well as surgical technique. Measurements are also used for other purposes, including the development of classification systems and to enable surgeons to compare and contrast their outcomes with those of other surgeons.

Although much has been published on the utility and benefits of objective measurements in breast surgery, very little information exists in the literature regarding their reliability. Outcome studies have used objective breast measurements to guide implant selection. Recommendations for surgical techniques, such as lowering the inframammary fold (IMF), are often made with adjustments of as little as 1 mm. It is reasonable to wonder whether the surgeon can measure with enough accuracy and reliability to make these minor adjustments.

Tissue-based planning is recognized as an important component in obtaining successful outcomes in breast-implant surgery. Multiple measurement systems have been described by various authors and are based on a series of static and dynamic measurements of the breast. Static measurements are taken between 2 anatomic landmarks.
Static Measurements

In tissue-based planning, implant selection is based on both the quality of the tissue and the dimensions of the breast. The footprint of the breast, representing 2 of the 3 dimensions, is defined by the base width (BW) and height. The BW is defined as the measure of the breast parenchyma from the medial edge where the breast begins as an upward slope to its lateral border. The BW is often cited as a critical parameter in objectively defining the footprint of the breast and selecting an appropriate implant. Thus, the importance of an accurate BW value in preoperative planning cannot be overstated. There are several variables that can affect the accuracy of the BW measurement. These include patient positioning, arm position, location on the breast the measure is performed, and the tool used to measure the BW. In our study, the BW measure was highly reliable across all raters when measured using a caliper, with the patient standing, arms at 90 degrees, and measuring from the medial breast to the most prominent area of the lateral border.

The Sn-N measurement describes the position of the nipple on the breast mound relative to the chest wall along the cranial-caudal axis and is measured along the skin of the breast as it curves from the sternum to nipple. The nipple position is a defining feature of the ideal breast. An attractive breast will have a 45:55 percent ratio of upper pole to lower pole, with the nipple meridian distinguishing the limits of the upper and lower poles. Therefore, the position of the nipple on the breast mound is a critical parameter for achieving an aesthetically pleasing outcome. The Sn-N measure is used both preoperatively for planning as well as intraoperatively to assess for symmetry. Reduction, augmentation, or reshaping of the breast during surgery can affect the nipple position. It is common to perform intraoperative assessment of the Sn-N distance to assess these changes and insure a proper balance between the nipple and the underlying breast mound. Based on our observed results, the Sn-N value is a reliable and reproducible measure of nipple position.

The N-M measurement describes the position of the nipple relative to the midline. In the horizontal plane, it objectively identifies asymmetric nipple positions or nipples that are medial or lateral to the breast meridian. This is an important measure for patient education. When adjusting the nipple position in reduction or mastopexy procedures, this measure can be used to plan for horizontal adjustments. In augmentation, this measure, taken in conjunction with BW and intermammary distance, allows the surgeon to plan for ideal positioning of the implant and provides a framework for informing the patient about expected postoperative nipple placement on the new breast mound. We found the N-M distance measure to be highly reliable, independent of rater experience.

Dynamic Measurements

Dynamic measurements require manipulation of the breast gland and skin in order to record a measurement. They are more difficult to perform and are subject to greater variability in measurement technique. Overall, the dynamic measurements performed in this study demonstrated poor reliability.

N-IMF was the most reliable measurement in this category. This measurement is important not only in determining the appropriate surgical procedure for the patient but also in implant selection and setting the IMF location when performing breast augmentation. The reliability of this measurement will be based on the ability of the surgeon to determine exactly where the IMF exists and to be consistent
on how much stretch is applied to the breast while obtaining measurements. We have demonstrated good inter-rater reliability with the N-IMF measurement. Intra-rater reliability was excellent for more senior raters and good for junior raters, demonstrating the importance of experience in refining measurement technique.

The compliance of the breast is an important feature when planning aesthetic breast surgery. This can be determined either subjectively or objectively. Many surgeons use a subjective assessment, essentially deciding that a breast is either tight, average, or loose. The APSS is an objective measure designed to assess compliance. Although a prestudy session to review measurement technique was completed, the raters in this study found this measurement to be difficult to perform. Both inter-rater and intra-rater reliability was poor, demonstrating limited use for this measurement. It would be useful to further study the agreement between a subjective assessment of tissue compliance in comparison to objective measures such as APSS.

The distribution of tissue within the breast is an important consideration in planning maximal soft-tissue coverage and distribution over a breast implant. Tissue cover may determine the ideal choice for implant pocket or the degree of pectoral-muscle release that is desirable along the IMF. It is also helpful in educating patients about the likelihood for implant palpability, visibility, and rippling. This study demonstrated poor reliability for both the STPT:UP and STPT: IMF measurements. The authors strongly support the need for assessment of the quantity of soft-tissue cover in various areas of the breast. At a minimum, decision making should be based on subjective assessments. It would be important to improve standardization and accuracy of existing objective measures and strive to develop better objective measures of soft-tissue cover.

Although, as a whole, the dynamic measurements prove to be unreliable measures, the features they attempt to measure are nonetheless important when planning aesthetic breast surgery. It is possible that individual surgeons measuring consistently over time will be able to improve upon the reliability of their measurements. A combination of these objective measurements along with a subjective assessment of the skin envelope should be considered as part of the assessment of the breast.

Limitations

The study and rater populations of interest were selected to maximize the generalizability of the data and conclusions. Although the sample size required was met, a larger number of participants and raters from different geographical locations would be more representative of the general patient population and trained plastic surgeons.

Raters were blinded to the recorded measurements; however, they were aware that their judgments would be compared to other raters. The raters’ behavior may have been altered because of their awareness of the study observation. This awareness would likely increase the raters’ diligence and precision during the quantitative assessment, resulting in good-quality data.

CONCLUSIONS

It is important to evaluate the reliability of anthropometric breast measurements that are widely used in the preoperative planning of aesthetic breast surgery. Basic measurements used in the objective evaluation of the breast are not all reliable. Static measurements, namely Sn-N, N-M, and BW, have excellent reliability. With the exception of N-IMF, dynamic measurements proved to be unreliable. Subjective and objective assessments of breast dimensions and tissue quality are critical when planning for aesthetic or reconstructive breast surgery. Further work is necessary to develop objective measures that can deliver reliable and reproducible results.

Supplementary Material

This article contains supplementary material located online at www.aestheticsurgeryjournal.com.

Disclosures

The authors declared no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Funding

The authors received no financial support for the research, authorship, and publication of this article.

REFERENCES


