Published in final edited form as:

Radiology. 2006 February ; 238(2): 425–437.

Breast Lesions: Evaluation with US Strain Imaging—Clinical Experience of Multiple Observers¹

Dawn M. Regner, MD, Gina K. Hesley, MD, Nicholas J. Hangiandreou, PhD, Marilyn J. Morton, DO, Michelle R. Nordland, RTRM, Duane D. Meixner, RDMS, Timothy J. Hall, PhD, Michael A. Farrell, MD, Jayawant N. Mandrekar, PhD, W. Scott Harmsen, MS, and J. William Charboneau, MD

Abstract

Purpose—To prospectively determine the accuracy of using an ultrasonographic (US) strain imaging technique known as lesion size comparison to differentiate benign from malignant breast lesions.

Materials and Methods—Institutional Review Board approval and patient informed consent were obtained for this HIPPA-compliant study. US strain imaging was performed prospectively for 89 breast lesions in 88 patients. Lesions were imaged by using free-hand compression and a real-time strain imaging algorithm. Five observers obtained manual measurements of lesion height, width, and area from B-mode and strain images. By using these size measurements, individual observer and group performances were assessed by using the area under the receiver operating characteristic curve (A_z). The performance of a single size parameter versus that of a combination of size parameters was evaluated by using univariate and multivariate logistic regression.

Results—Group A_z values showed that width ratio and area ratio yielded the best results for differentiating benign and malignant breast lesions, and they were not statistically different from one another ($P = .499$). For the group, the performance of area and width, which was superior to that of height and aspect ratio, was statistically significant for all cases ($P = .011$) except for those that compared area with aspect ratio ($P = .118$). By using a group threshold of 1.04 for width ratio and 1.13 for area ratio, the sensitivity and specificity of the technique were 96% and 21%, respectively, for width and 96% and 24%, respectively, for area. The best observer achieved a sensitivity of 96% and a specificity of 61% by using the area ratio. For all but one observer, combined size parameters did not improve observer performance ($P = .258$). Significant interobserver performance variability was observed ($P = .001$).

Conclusion—Results suggest that US strain imaging has the potential to aid diagnosis of breast lesions. However, manually tracing lesion boundaries for size ratio differentiation in a busy clinical setting did not match the diagnostic performance levels previously reported. Focusing on

¹From the Department of Radiology, Mayo Clinic, 200 First St SW, Rochester, MN 55905. From the 2003 RSNA Annual Meeting.

Address correspondence to D.M.R. (e-mail: regner.dawn@mayo.edu).

Author contributions:

Guarantors of integrity of entire study, D.M.R., G.K.H., N.J.H., M.A.F.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; literature research, D.M.R., G.K.H., N.J.H., T.J.H., M.A.F.; clinical studies, D.M.R., M.J.M., M.R.N., D.D.M.; statistical analysis, D.M.R., N.J.H., J.N.M., W.S.H.; and manuscript editing, D.M.R., G.K.H., N.J.H., M.J.M., D.D.M., T.J.H., M.A.F., W.S.H., J.W.C.

See Materials and Methods for pertinent disclosures.

Supported in part by a grant from Siemens Medical Solutions, Ultrasound Division, Issaquah, Wash.

T.J.H. supported by grant DAMD17-00-1-0596 from the U.S. Army Medical Research and Materiel Command.

measurements of lesion width, along with additional observer training or automated processes, may yield a suitable method for routine clinical application.

Abbreviation

A_z = area under the receiver operating characteristic curve

The observation that benign and malignant breast lesions have inherently different firmness has long been used by clinicians during palpation of the breast; harder and less mobile lesions are considered more likely to be malignant (1). Several different imaging methods, including ultrasonographic (US) strain imaging (also known as US elastography), have been developed to measure the relative stiffness of lesions in contrast to the surrounding tissue (2–17). US strain imaging may allow the differentiation of benign from malignant solid masses on the basis of differences in stiffness.

US strain images are produced by comparing the US echo data obtained prior to and after slight axial compression of the breast. These data are used to determine the tissue displacement at each location in the breast as a result of compression. Strain is computed as the rate of change in axial tissue displacement as a function of depth. Strain images are produced when the relative differences in tissue motion at each location in the breast are calculated and displayed. Harder areas of the breast (ie, areas with less tissue displacement during compression) appear darker on strain images, and softer areas of the breast (ie, areas with more tissue displacement during compression) appear brighter. Preliminary work does not support the use of direct quantification of lesion strain as an accurate means of differentiating benign from malignant solid lesions (6,11,14,17,18).

Peer-reviewed work by Garra et al (6) and others has shown that discrepancies between the size of breast lesions on B-mode and strain images may be a promising way to distinguish benign from malignant lesions; this method is known as the lesion size comparison technique. Researchers have found that malignant lesions tend to be larger on strain images than on corresponding B-mode images, potentially because of the surrounding desmoplastic reaction that accompanies most malignancies (6,11,17,19). Thus, the purpose of our study was to prospectively determine the accuracy of using a US strain imaging technique known as lesion size comparison to differentiate benign from malignant breast lesions.

Materials and Methods

Siemens Medical Solutions, Ultrasound Division, Issaquah, Washington, provided equipment and financial support for this study. The authors had control of the data and information submitted for publication.

Patient Population

Approval for this study was obtained from the Institutional Review Board, and informed consent was obtained from the enrolled patients. The study was compliant with the Health Insurance Portability and Accountability Act. From February 6, 2002, to August 6, 2003, US strain imaging was performed prospectively for 102 consecutive breast masses in 101 patients (mean age, 58 years; median age, 58.5 years; age range, 20–85 years). Candidates were recruited from a clinical pool of patients who were known to have lesions that were visible at US and that warranted biopsy. The first four recruited patients were excluded as training patients, seven recruited patients were excluded because of technical difficulties with radiofrequency echo data transfer, and two patients were excluded because biopsy results showed both benign and malignant histologic features in the same biopsy sample. Thus, 89 lesions in 88 women were included in the final data analysis. Lesion diameters ranged from

0.3 to 3.0 cm, with a mean lesion diameter of 1.3 cm. For 73 of 88 lesions, biopsy was performed after strain imaging on the same day (mean, 41.5 minutes; range, 3–245 minutes). For 12 of 88 lesions, biopsy was performed at a later date (mean, 17 days; range, 1–112 days). For three of 88 lesions, biopsy was performed at a previous date (1, 2, and 6 years previously). For these three lesions, the five observers were blinded to the fact that a previous biopsy had been performed, as well as to the previous biopsy results.

US Strain Imaging Technique

A mammography technologist with 10 years of experience in breast US imaged all lesions at 7.3 MHz by using a real-time strain imaging algorithm that was developed by Hall et al (17); this algorithm was implemented by using a standard commercial US system (Elegra scanner and 7.5L40 linear array transducer; Siemens Medical Solutions). By using a freehand compression technique, the technologist manually applied slight axial compression to the lesions with the US probe. B-mode and strain images were reconstructed off line by using the same radiofrequency echo data that were acquired during the US examination. Images were displayed in a side-by-side format within the individual frames that were contained within a cine-loop sequence of approximately 100 frames. Off-line processing of the stored radiofrequency data was performed by using custom software which utilized the same algorithm that exists on the US scanner, except a smaller correlation window was used (16 samples instead of 24 samples) resulting in superior strain image spatial resolution. Also, radiofrequency data were stored at a greater frame rate than were those displayed in real time with the US scanner; thus, off-line temporal resolution was superior to that displayed in real time with the scanner (17).

B-Mode and Strain Image Evaluation

Five observers (designated as observers A–E) with different characteristics (Table 1) individually produced measurements of lesion size, including height and width, and traced the area of the lesion off line on reconstructed B-mode and corresponding strain images by using a mouse. Observers made their measurements on a single representative image frame that was personally chosen as best (designated as “own image”) from the cine-loop sequence of about 100 frames. At a later date, the measurements, including height, width, and traced area, were repeated by each observer on the unmarked standardized image frames (designated as “reference image”) that were chosen as best by the technologist who performed the examinations (observer D). Prior to obtaining study measurements, all of the observers met two to three times to receive specific training both in choosing the optimal strain image from the cine-loop sequence and in making consistent lesion size measurements. All observers practiced breast lesion measurements on multiple sample B-mode and strain images during these training sessions. All observers were comfortable with the technique at the end of the training sessions. During the study, observers were asked to report any concerns they had regarding difficulties in producing measurements, as well as any anecdotal observations they believed to be important.

Data and Statistical Analysis

Size ratios were computed for each of the four size parameters (ie, height, width, area, and aspect [width-to-height] ratio) by dividing the strain image measurements by the B-mode image measurements. Ratios were calculated for each individual observer and for the group as a whole. Group ratios were obtained by averaging the individual observer B-mode and strain image measurements and by calculating ratios from these averaged measurements. The own image size ratios were analyzed by using receiver operating characteristic methods for each observer and for the group. Because images from the same patient were used multiple times, the correlated areas under the receiver operating characteristic curve (A_z values) were compared

to determine which size ratio was best suited for differentiating benign from malignant lesions; comparisons between these values were made by using a method proposed by DeLong et al (20). The ratio threshold value that achieved at least 95% sensitivity in differentiating malignant from benign lesions while optimizing specificity was then determined for each individual observer and for the group. This 95% sensitivity threshold is likely the most valuable parameter for breast imaging because the desire is to minimize the chances of missing a malignancy.

Univariate and multivariate logistic regression analyses (21) were performed, and a comparison of the c statistic values that were generated from these analyses was performed by using the method proposed by DeLong et al (20) to determine if the combined size parameters were superior to any one size parameter alone in predicting benignity or malignancy. The A_z value, threshold, and sensitivity and specificity values for own image observer data were compared to see if there was significant interobserver performance variation. Own image and reference image A_z values for width and area were compared by using the method proposed by DeLong et al (20), and the ranges of own image and reference image sensitivity and specificity were compared in an attempt to isolate the most important sources of any interobserver performance variation. A subset analysis was performed by using receiver operating characteristic methods of only those lesions that were judged by observer A to have well-defined borders on conventional B-mode US images to determine if the performance of the lesion size comparison technique was dependent on the conspicuity of lesion borders at B-mode imaging.

Finally, a comparison was made between the results presented by Hall et al (22) and those obtained in the current study by using the same US strain imaging algorithm and analysis approach. A P value of less than .05 was considered to indicate a statistically significant difference for all analyses. All of the statistical analyses were performed by using a commercially available software program (SAS, version 8.0; SAS Institute, Cary, NC).

Observers A, B, and C participated in a preliminary study (19) (performed 4–6 months earlier) during which measurements were made of the first 50 of 89 lesions that were included in the current study. Although the three observers involved in the preliminary study remeasured the first 50 lesions for the current study, none of the observers believed that the initial analysis provided any advantage or bias other than increased practice. To support this, an analysis was performed to compare the A_z values of the first 50 lesions of the current study (ie, those that were previously measured as part of the preliminary study) with those of the last 39 lesions of the current study (ie, those that were not previously measured) for observers A, B, and C.

Results

All lesions were grouped as benign ($n = 38$) or malignant ($n = 51$) according to surgical or needle biopsy results (Table 2). Figures 1–3 show examples of the appearance of benign and malignant lesions, as well as the size measurements made by each of the observers. Figures 4–9 show additional examples of various benign and malignant lesions that were observed during the study.

A_z Value, Threshold, Sensitivity, and Specificity

Own image A_z values for the five observers and for the group as a whole are shown in Table 3. For four of five observers and for the entire group, width and area ratios yielded the best results (ie, the largest A_z values) for differentiating benign from malignant breast lesions. Observer E was the exception in that, for this observer, width and aspect ratios yielded the best results. For the entire group, the superior performance of width and area measurements when compared with height and aspect ratio was statistically significant for all cases ($P \leq .011$), except for the comparison of area and aspect ratio, which was not significant ($P = .118$). For the five individual observers, the A_z values for width and area, when compared with those of

height and aspect ratio in every combination, resulted in 13 of 20 comparisons in which width and area yielded significantly better performance ($P \leq .032$). In six of 20 comparisons, width and area were superior to height and aspect ratio, but the differences were not statistically significant ($P \geq .054$). In one of 20 comparisons (ie, for observer E), aspect ratio was superior to area, but the difference was not significant ($P = .311$). Width and area were statistically equivalent to one another ($P = .499$ for the group, $P \geq .421$ for observers A–D, and $P = .069$ for observer E).

Table 4 shows the own image threshold data that was computed for each observer (designated as “observer threshold”) and for the entire group (designated as “group threshold”) for both width and area ratios. The optimal group ratio threshold was 1.04 for width and 1.13 for area. These thresholds yielded a sensitivity and specificity of 96% and 21%, respectively, for width and of 96% and 24%, respectively, for area. Figures 10 and 11 show individual case-by-case data in graphic format.

To determine if a combination of multiple size parameters was superior to any single size parameter in predicting if a lesion was benign or malignant, univariate and multivariate logistic regression analyses were performed (Table 5). Several size parameter models (ie, single size parameters and combined size parameters) were assessed on the basis of the c statistic, which is equivalent to A_z value (21), and were compared by using the method proposed by DeLong et al (20). Models with a larger c statistic represent enhanced observer performance in differentiating benign from malignant breast lesions. In Table 5, either traced area or width alone was an important discriminating factor of malignancy for four of five observers by using univariate logistic models. Also in Table 5, only a few combinations of size parameters yielded slightly higher A_z values than width or area alone; however, for four of five observers, the differences were not statistically significant ($P \geq .258$). Observer E was the exception in that, for two of eight comparisons, a significant advantage was seen when combined parameters were used ($P \leq .027$). Overall, this analysis demonstrates that there is no overall improvement in the discriminating capabilities of multivariate models, as is shown in Table 5.

Assessment of Interobserver Variation

As seen in Table 3, individual observer A_z values ranged from 0.74 to 0.90 for width ratios and from 0.67 to 0.92 for area ratios. As seen in Table 4, individual observer thresholds for own image measurements ranged from 0.28 to 1.04 for width and from 0.90 to 1.24 for area. Applying the group threshold to each observer’s own image width data resulted in observer sensitivities ranging from 86% to 96% and specificities ranging from 21% to 61%. Applying the group threshold to each observer’s own image area data resulted in observer sensitivities ranging from 82% to 98% and specificities ranging from 16% to 71%. Testing all pairs of individual observers against one another showed some statistically significant differences between observers at the 95% confidence level ($P < .001$ for the group) (Table 6). Specifically, observer A performed significantly better than all other observers when the area parameter was used ($P \leq .01$) and significantly better than all but one other observer when the width parameter was used ($P \leq .01$ for the remaining observers). Observers B and C performed significantly better than observer E when the area parameter was used ($P \leq .01$). In addition, observers B and C tended to perform better than observer D for both width and area and better than observer E for width; this difference, however, was not statistically significant ($P \geq .07$).

Analysis of Interobserver Variation

After the US data are acquired, there are at least two potential sources of interobserver variation when the lesion size comparison technique is used—that is, variation in choosing the optimal strain image from the cine-loop sequence and variation in measuring the lesion size parameter on the strain and corresponding B-mode images within the chosen frame. To determine which

of these sources contributed most to our significant observer variation, all observers made measurements on unmarked standard images. To remove the potential interobserver variation that is inherent in choosing the image, these unmarked images were chosen by the same technologist who performed the examinations and were designated as the reference images. Applying the group threshold to each observer's reference image width data resulted in observer sensitivities ranging from 90% to 100% and specificities ranging from 11% to 42% (Table 4). Applying the group threshold to each observer's reference image area data resulted in observer sensitivities ranging from 90% to 98% and specificities ranging from 16% to 50% (Table 4). For width, individual reference image A_z values ranged from 0.71 to 0.87 (Table 3), and for area, individual observer reference image A_z values ranged from 0.66 to 0.85 (Table 3). By definition, the own image and reference image A_z values for observer D were equivalent. For the remaining four observers, comparison of own image and reference image A_z values for width and area demonstrated a statistical reduction in observer performance in only one of eight comparisons ($P = .025$ for area, observer C). In three other instances, however, marginal P values were obtained ($.057 \leq P \leq .084$).

Subset Analysis of Breast Lesions with Well-defined Borders at B-Mode US

During the analysis, several observers reported that, in certain cases, measuring lesions on the conventional B-mode display was particularly difficult (Fig 3). This might lead to a decrease in measurement accuracy and an increase in interobserver (and possibly intraobserver) variability. Observer A reviewed the cases and determined that a total of 60 of 89 lesions had well-defined borders on B-mode images. A subset analysis was performed for measurements obtained in these 60 lesions (31 malignant and 29 benign). Group A_z values for this subset were 0.81 for both width and area (Table 3). These A_z values were slightly lower than the group A_z values for the complete data set, which were 0.85 and 0.83 for width and area, respectively (Table 3). In only one case did the subset produce a somewhat superior A_z value (observer E, area). Also, for the group and for all but one observer, the sensitivity and specificity of width and area ratio thresholds were not substantially improved when only the subset of lesions with well-defined borders at B-mode imaging was considered.

Comparison with Previous Results

In a previous study by Hall et al (22), a series of 169 breast lesions were evaluated by using the same strain imaging technique as was used in our study. A single observer who had extensive experience with the lesion size comparison technique picked the optimal strain image from the cine-loop sequence and obtained area measurements. By using area as the size parameter and 1.2 as the optimal area ratio threshold, a sensitivity of 100% and a specificity of 75.4% in characterizing breast lesions were obtained in this series of 169 lesions. When this 1.2 threshold for area ratio was applied to the own image ratio data obtained in our current study, our group sensitivities and specificities were 94% and 29%, respectively. When the 1.2 threshold for area ratio was applied to the own image ratio data of observer A (our best performer), sensitivity and specificity were 86% and 82%, respectively.

Analysis of the First 50 Lesions

To show that the preliminary study did not provide observers A, B, and C with any advantage or bias other than increased practice, an analysis was performed to compare A_z values of the first 50 lesions of the current study (ie, those that were previously measured as part of the preliminary study) with those of the last 39 lesions of the current study (ie, those that were not previously measured). This analysis showed that, in general, the A_z values were not significantly different between the first 50 lesions and the last 39 lesions for observers A, B, and C.

Observations Reported by Observers

All observers reported that choosing the optimal frame from the cine-loop sequence and making size measurements took much more time than expected and more time than might be available in a typical busy breast imaging practice. Particularly, observers thought that choosing the optimal image was the most time-consuming step. Also, it was a consensus observation that observers A and B had much more time available in a laboratory environment to make measurements than did the other observers, who performed measurements under considerable time constraints during the course of a busy clinical day.

Several important anecdotal observations that were unrelated to lesion size comparison were reported by the observers during the study. The lesion was sometimes much more conspicuous on strain images than on B-mode images, which aided in determining that a lesion was, in fact, real (Fig 3). In many cases, there was little contrast between the breast lesion and surrounding tissue on the strain images, thereby making the lesion inconspicuous. This finding seemed to correlate with benignity (Fig 9). These observations are consistent with those obtained by Garra et al (6). We also observed that repetitive conspicuity of lesions on US strain images throughout the course of the cine-loop sequence was correlated with increased probability of malignancy. Finally, in at least two patients, strain imaging aided in localizing the optimal site of biopsy. For example, in one patient, the initial biopsy passes, which were made by using only the B-mode image as a guide, yielded a suboptimal sample. US strain imaging was subsequently performed on the lesion, and the firmest area of the lesion was identified as superior to the initial biopsy site. Multiple biopsies that were performed through the area that was identified at strain imaging as being the firmest yielded several good core biopsy results that proved to be adequate for histologic diagnosis. These observations suggest that the use of strain imaging with means of interpretation other than the comparison of lesion dimensions may have clinical utility.

Discussion

In our study, low specificity values were obtained when using the lesion size comparison technique along with our best parameters (ie, width and area) and an optimal ratio threshold to characterize breast lesions. For each observer and for the group as a whole (except for observer E), width and area yielded similar results. A combination of size parameters did not produce better results than either width or area alone. The use of width measurements may be beneficial in a busy breast imaging practice because width is an easier and quicker measurement to obtain than traced area.

One of the main limitations that we found was the statistically significant performance differences between observers. On the basis of A_z values, differences in performance were most directly related to the amount of time available to make the lesion size measurements (laboratory vs clinical environment) and, to a lesser extent, the amount of prior experience that each observer had with the lesion size comparison technique; differences in performance were not correlated with general clinical breast imaging experience. For example, although no formal time data were collected, all observers agreed that, compared with the other three observers, observers A and B had much more available time in a laboratory setting to make the measurements. This conferred a performance advantage (ie, a larger A_z value) for observers A and B. This suggests that manually selecting a frame from the cine-loop sequence, tracing the lesion boundaries, and measuring the lesion width and height may require more time than is routinely available in typical clinical practice; this supports the consensus impression that the current technique required too much time. Development of an automated version of the technique that could choose the optimal frame from the cine-loop sequence and make measurements of the lesion border could improve the speed of the technique.

Even though observers D and E have the most clinical breast imaging experience, observers A, B, and C performed better (in some cases significantly better). We believe this is in part because of their prior experience with and training in the lesion size comparison technique as part of a prior preliminary study. A comparison between A_z values for the first 50 lesions and those for the second 38 lesions showed that the enhanced performance of observers A, B, and C was not the result of previous experience in specifically measuring the first 50 lesions. Rather, the enhanced performance of observers A, B, and C was a function of increased general experience and practice with the lesion size comparison technique that was derived from the preliminary study. This suggests that even highly experienced mammographers and breast US technologists (eg, observers E and D) who have specific training in the technique will potentially need more extensive training than that which was offered prior to this study, as well as more practice and experience in the current lesion size comparison technique in order to use the technique effectively.

Having proved that significant inter-observer variability exists with this technique, an attempt to isolate the main source of this variability was made by comparing the group and observer own image and reference image data. After US data have been acquired, the two potential sources of variability include choosing the optimal frame from the cine-loop sequence and making measurements of the lesion in that chosen frame. If a reduction in the performance range was observed for the reference image data when compared with the wide performance range for the own image data, this would suggest that most of the variation was introduced during the choosing of the optimal frame. Likewise, if no reduction was seen, this would indicate that little or no variation was introduced during the choosing of the frame, and thus, the bulk of variability occurred during measuring. Results showed that there was only a slight decrease in the performance range of reference images (and thus in interobserver variability) compared with the observer performance range of own images (particularly for area) when comparing observer sensitivity and specificity data.

When comparing observer own image A_z values with reference image A_z values, a statistically significant reduction in performance occurred for only one of eight comparisons; in three cases, however, the P values were marginal. This demonstrates that most of the interobserver variability resulted from the measurement of lesion size, although frame selection was still important.

Finally, subjectively, one can see from Figures 2 and 3 that the five different observers sometimes had different interpretations of lesion borders, and thus, different size ratios were obtained. Identifying that the differences in lesion measurement contributed most to observer variability is helpful. It suggests that the most concentrated training for breast imagers who are learning to use the lesion size comparison technique should be on the actual measuring of lesion size on B-mode and strain images. It also suggests that automated border detection and measurement may decrease variability and hopefully lead to higher sensitivity and specificity values.

Several observers reported that some of the lesions were difficult to measure on the B-mode display because of ill-defined borders. This may, in part, be the result of the relatively low-frequency US probe (7.3 MHz) that was used in the study. To determine if B-mode lesion border conspicuity affected the performance of the technique, a subset analysis of only those lesions that had well-defined margins on B-mode images was performed. On the basis of the results of the analysis, which included lower A_z values for the subset, it appears that the conspicuity of lesions on B-mode images is not a complicating factor when performing these measurements and analysis. This was unexpected because it was believed that lesions with ill-defined borders on B-mode images might have more inaccurate measurements (and thus more inaccurate ratios) than the well-defined lesions owing to the difficulty in perceiving lesion

borders. It was postulated that including only those lesions with well-defined borders in the analysis would increase the performance of the technique. The results, however, are contrary to this postulation.

One potential explanation for this unexpected finding is that most of the ill-defined lesions (20 of the 29 lesions) had malignant histologic characteristics. In looking at these malignant lesions, such lesions tended to be obviously larger on the US strain images than they were on the corresponding B-mode images. Thus, an inaccurate measurement obtained from B-mode images potentially becomes unimportant because the resultant size ratio would still be well above the threshold ratio for determining malignancy, and thus, the technique would still be able to accurately characterize most of these lesions as malignant. Taking these lesions that are accurately characterized by the technique out of the complete data set would actually result in lower A_z values and no improvement in the sensitivity and specificity values. This analysis may also suggest that the low-frequency US probe that was used in the study probably did not contribute to the low specificity values that were obtained.

In the previous presentation by Hall et al (22), an optimal area ratio threshold of 1.2, a sensitivity of 100%, and a specificity of 75.4% in characterizing breast lesions were obtained in a series of 169 lesions. This threshold is similar to our group area threshold of 1.13, but our group sensitivity was 96% and specificity was only 24%. Also, when the threshold of 1.2 for area ratio was applied to our own image ratio data, our group sensitivities and specificities were 94% and 29%, respectively. Our best performer (observer A) obtained a sensitivity of 96% and a specificity of 61% by using an area ratio threshold of 1.01, which is closer to the sensitivity and specificity values obtained by Hall et al (22).

When using the area threshold of 1.2, however, we found that the sensitivity and specificity for observer A were only 86% and 82%, respectively. This shows that “tuning” a ratio threshold to an individual observer can produce good results but that finding a single generalizable threshold may be difficult. The differences in our group performance compared with the results of Hall et al are likely related to the fact that, in the study by Hall et al, the results were obtained by a single observer who had both extensive experience with this technique and more time available to make the measurements, which were performed in the laboratory as opposed to a typical busy clinical environment. This conclusion is supported by the fact that the results of observer A, who had prior experience with the technique and who had more available time to make measurements, better approached the results of Hall et al (although not with the same area threshold) than did the results of the group as a whole. Again, this supports the conclusion that more extensive training and experience with the technique are vital to use the lesion size comparison technique in an acceptable way. Most importantly, this suggests that the sensitivity and specificity of the technique are directly related to the time available to perform the measurements and that fitting the technique into a typical busy clinical practice may pose problems.

We identified two main limitations in our study. First, strain images were generated by a single technologist, and thus, variation that was introduced during the acquisition of the strain images was not tested. Second, although all observers agreed that the observers who obtained measurements in the clinical environment had much less time and that the technique took much more time than anticipated, formal time records of how long each observer took to pick the optimal image and to generate the size measurements were not documented. Future studies could address these issues.

In conclusion, US strain imaging has the potential to aid in diagnosis of breast lesions; however, manually traced lesion borders for size ratio differentiation did not result in acceptable performance, though improvements were noted. The main limitations we identified with this

technique include low specificity for some observers and marked interobserver variation (mainly in lesion size measurement). The consensus of the participants was that the extensive time required in choosing the optimal frame from the cine-loop sequence and in tracing the lesion boundaries in both B-mode and strain images could make the routine application of this technique in a typical busy clinical breast imaging practice difficult. These findings, however, do not preclude the possibility that alternative approaches to using this general method would be more successful. For example, manual measurements of lesion width alone might be performed in an acceptably short time, and this simpler measurement, coupled with more extensive user training and experience, may result in better observer performance and improved observer agreement. The development of an automated process that could choose the optimal frame from the cine-loop and obtain lesion size measurements might also decrease variability and increase the ease and speed with which the technique can be applied.

References

1. Anderson, WAD. Pathology. St Louis, Mo: Mosby; 1953.
2. Lerner RM, Huang SR, Parker KJ. Sonoelasticity images derived from ultrasound signals in mechanically vibrated tissues. *Ultrasound Med Biol* 1990;16:231–239. [PubMed: 1694603]
3. Ophir J, Cespedes EI, Ponnekanti H, Yazdi Y, Li X. Elastography: a quantitative method for imaging the elasticity of biological tissues. *Ultrason Imaging* 1991;13:111–134. [PubMed: 1858217]
4. O'Donnell M, Skovoroda A, Shapo B, Emelianov S. Internal displacement and strain imaging using ultrasonic speckle tracking. *IEEE Trans Ultrason Ferroelectr Freq Control* 1994;41:314–325.
5. Skovoroda AR, Emelianov Y, O'Donnell M. Tissue elasticity reconstruction based on ultrasonic displacement and strain images. *IEEE Trans Ultrason Ferroelectr Freq Control* 1995;42:747–765.
6. Garra BS, Cespedes EI, Ophir J, et al. Elastography of breast lesions: initial clinical results. *Radiology* 1997;202:79–86. [PubMed: 8988195]
7. Ophir J, Kallel F, Varghese T, et al. Elastography: a systems approach. *Int J Imaging Syst Tech* 1997;8:89–103.
8. Chaturvedi P, Insana MF, Hall TJ. 2-D compounding for noise reduction in strain imaging. *IEEE Trans Ultrason Ferroelectr Freq Control* 1998;45:179–191.
9. Fatemi M, Greenleaf JF. Application of radiation force in noncontact measurement of the elastic parameters. *Ultrason Imaging* 1999;21:147–154. [PubMed: 10485567]
10. Lubinski MA, Emelianov Y, O'Donnell M. Adaptive strain estimation using retrospective processing. *IEEE Trans Ultrason Ferroelectr Freq Control* 1999;46:97–107.
11. Ophir J, Alam S, Garra B, et al. Elastography: ultrasonic estimation and imaging of the elastic properties of tissues. *Proc Inst Mech Eng [H]* 1999;213(3):203–233.
12. Pesavento A, Lorenz A, Ermet H. System for real-time elastography. *Elect Lett* 1999;35:941–942.
13. Doyley MM, Bamber JC, Fuechsel F, Bush NL. A freehand elastographic imaging approach for clinical breast imaging: system development and performance evaluation. *Ultrasound Med Biol* 2001;27:1347–1357. [PubMed: 11731048]
14. Hiltawsky KM, Druger M, Starke C, Heuser L, Ermet H, Jensen A. Freehand ultrasound elastography of breast lesions: clinical results. *Ultrasound Med Biol* 2001;27:1461–1469. [PubMed: 11750744]
15. Nightingale KR, Palmeri ML, Nightingale RW, Trahey GE. On the feasibility of remote palpation using acoustic radiation force. *J Acoust Soc Am* 2001;110:625–634. [PubMed: 11508987]
16. Varghese T, Zagzebski JA, Lee FT Jr. Elastographic imaging of thermal lesions in the liver in vivo following radiofrequency ablation: preliminary results. *Ultrasound Med Biol* 2002;28(11–12):1467–1473. [PubMed: 12498942]
17. Hall TJ, Zhu Y, Spalding CS. In vivo real-time freehand palpation imaging. *Ultrasound Med Biol* 2003;29:427–435. [PubMed: 12706194]
18. Hangiandreou NJ, Meixner DM, Hesley GK, et al. Ultrasound strain image data obtained in breast masses: preliminary quantitative analysis [abstr]. *Ultrasound Med Biol* 2003;29:S178.
19. Regner DM, Hesley GK, Hangiandreou NJ, et al. Initial clinical experience with ultrasound strain imaging for the evaluation of breast masses [abstr]. *Radiology* 2003;229:P213.

20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845. [PubMed: 3203132]
21. Hosmer, DW.; Lemeshow, S. *Applied logistic regression*. New York, NY: Wiley; 2000.
22. Hall TJ, Svensson W, Von Behren P, et al. Lesion size ratio for differentiating breast masses. *IEEE International Ultrasonics Symposium Book of Abstracts* 2003:87–88.

Figure 1.

Side-by-side paired displays of B-mode (left image of each pair) and strain (right image of each pair) US images of benign fibroadenoma. Unmarked reference images (upper left pair) along with images showing measurements of height (green line), width (yellow line), and traced area (blue line) are displayed for each observer (A–E). The resulting size ratios for width (WR) and area (AR) are noted. By using the group area and width threshold ratios presented in Table 4, strain imaging was used to correctly predict lesion benignity (true-negative finding).

Figure 2.

Side-by-side paired displays of B-mode (left image of each pair) and strain (right image of each pair) US images of invasive ductal carcinoma. Unmarked reference images (upper left pair) along with images showing measurements of height (green line), width (yellow line), and traced area (blue line) are displayed for each observer (A–E). The resulting size ratios for width (WR) and area (AR) are noted. Lesion has somewhat indistinct borders on B-mode images and demonstrates that marked interobserver measurement variability can result with this technique. By using the group width and area threshold ratios (Table 4), strain imaging was used to correctly predict lesion malignancy (true-positive finding).

Figure 3.

Side-by-side paired displays of B-mode (left image of each pair) and strain (right image of each pair) US images of benign fat necrosis. Unmarked reference images (upper left pair) along with images showing measurements of height (green line), width (yellow line), and traced area (blue line) are displayed for each observer (A–E). The resulting size ratios for width (WR) and area (AR) are noted. Images illustrate one of the potential problems with this technique in that, for some lesions, the lesion is difficult, if not impossible, to distinguish from the surrounding breast tissue on B-mode images. This likely makes measurements of lesion size on B-mode images inaccurate, and as shown here, different observers will vary widely in their interpretation of border and size measurements. In this example, only the size ratios for observer A correctly predicted lesion benignity. This lesion also demonstrates how strain imaging can confirm the presence of subtle lesions, which are much more conspicuous at strain imaging.

Figure 4. Side-by-side paired display of B-mode (left) and strain (right) US images of fibrocystic change. By using the group width and area thresholds presented in Table 4, strain imaging was used to incorrectly predict lesion malignancy (false-positive finding).

Figure 5.

Side-by-side paired display of B-mode (left) and strain (right) US images of invasive ductal carcinoma. Lesion appears much bigger on strain image than on corresponding B-mode image. By using the group area and width threshold ratios presented in Table 4, strain imaging was used to correctly predict lesion malignancy (true-positive finding).

Figure 6. Side-by-side paired display of B-mode (left) and strain (right) US images of invasive ductal carcinoma. By using the group width and area threshold ratios presented in Table 4, strain imaging was used to correctly predict lesion malignancy (true-positive finding).

Figure 7. Side-by-side paired display of B-mode (left) and strain (right) US images of invasive lobular carcinoma. By using the group width and area threshold ratios presented in Table 4, strain imaging was used to correctly predict lesion malignancy (true-positive finding).

Figure 8. Side-by-side paired display of B-mode (left) and strain (right) US images of invasive ductal carcinoma. Lesion width and area are similar on strain and B-mode images. By using the group width and area threshold ratios presented in Table 4, strain imaging was used to incorrectly predict lesion benignity (false-negative finding).

Figure 9. Side-by-side paired display of B-mode (left) and strain (right) US images of benign stromal fibrosis in patient with a history of breast cancer. Anecdotally, strain images that show little or no contrast between the lesion and the surrounding breast tissue, as is seen in this patient, seem to be a reliable indicator of benignity.

Figure 10.

Graph of own image group width ratios plotted for each lesion in the study. Horizontal line represents optimal group threshold ratio of 1.04. Above this threshold, malignant lesions were classified as true-positive findings, and below this threshold, benign lesions were classified as true-negative findings. Although malignant lesions tend to have larger width ratios, one can see that there is considerable overlap near the threshold line, which is reflected in low specificity values.

Figure 11.

Graph of own image group area ratios plotted for each lesion in the study. Horizontal line represents optimal group threshold ratio of 1.13. Above this threshold, malignant lesions were classified as true-positive findings, and below this threshold, benign lesions were classified as true-negative findings. Although malignant lesions tend to have larger area ratios, as is the case with width, one can see that there is considerable overlap near the threshold line, which is reflected in low specificity values.

Table 1

Observer Characteristics

Observer	Degree	Title	Clinical Experience in Breast US	Experience in US Strain Imaging	Measurement Environment*
A	MD	Radiology resident	1 mo	Yes	Laboratory
B	RDMS	Ultrasoundographer	None	Yes	Laboratory
C	MD	Board-certified radiologist	5 yrs	Yes	Clinical
D	RTRM	Breast US technologist	10 yrs	No	Clinical
E	MD	Board-certified radiologist	15 yrs	No	Clinical

* The laboratory measurement environment refers to an environment in which observers made measurements in a laboratory, with fewer distractions and much more time available to dedicate to lesion measurements. The clinical measurement environment refers to a more hectic environment in which observers made measurements under considerable time constraints owing to concomitant clinical duties.

Table 2**Histologic Characteristics of Breast Lesions**

Characteristic	No. of Lesions
Benign (<i>n</i> = 38)	
Fibroadenoma	16
Fibrocystic change	6
Stromal fibrosis	6
Abscess	2
Intraductal papilloma	2
Fat necrosis	2
Pseudoangiomatous stromal hyperplasia	1
Hematoma	1
Radial scar	1
Complex cyst	1
Malignant (<i>n</i> = 51)	
Invasive ductal carcinoma	39
Invasive lobular carcinoma	6
Mixed ductal and lobular carcinoma	3
Invasive tubular carcinoma	2
Ductal carcinoma in situ	1

Table 3
A_z Values for Complete Data Set and Subset of Lesions with Well-defined Borders

Parameter	Complete Data Set for Own Image			Complete Data Set for Reference Image		Well-defined Subset for Own Image*	
	Height	Width	Area	Aspect Ratio	Width	Width	Area
Entire group	0.71	0.85	0.83	0.74	0.77	0.81	0.81
Observer A	0.72	0.90	0.92	0.72	0.87	0.90	0.92
Observer B	0.73	0.83	0.82	0.70	0.73	0.80	0.81
Observer C	0.71	0.79	0.79	0.63 [†]	0.75	0.76	0.74
Observer D	0.62 [†]	0.75 [†]	0.77	0.68 [†]	0.75	0.68	0.72
Observer E	0.47 [†]	0.74	0.67	0.74	0.71	0.75	0.71

Note.—Standard errors for all *A_z* values ranged from 0.03 to 0.06.

* Subset includes data generated for lesions with well-defined borders at B-mode imaging only.

[†] Values were not found to be statistically significant at the 5% level.

Table 4
Ratio Thresholds with Sensitivities and Specificities for Each Observer and the Entire Group

Parameter	Width			Area		
	Ratio Threshold Value	Sensitivity (%)	Specificity (%)	Ratio Threshold Value	Sensitivity (%)	Specificity (%)
Entire group						
Own image						
Group threshold	1.04	96 (87, 100)*	21 (10, 37)*	1.13	96 (87, 100)*	24 (11, 40)*
Reference image						
Group threshold	1.03	96 (87, 100)*	18 (8, 34)*	1.17	96 (87, 100)*	21 (10, 37)*
Observer A						
Own image						
Observer threshold	0.98	96	42	1.01	96	61
Group threshold	1.04	92	61	1.13	88	71
Reference image						
Observer threshold	1.13	96	68	1.05	96	40
Group threshold	1.03	100	42	1.17	90	50
Observer B						
Own image						
Observer threshold	0.90	96	24	0.90	96	22
Group threshold	1.04	86	46	1.13	82	51
Reference image						
Observer threshold	1.01	96	37	1.06	96	18
Group threshold	1.03	94	37	1.17	94	24
Observer C						
Own image						
Observer threshold	1.04	96	29	1.08	96	21
Group threshold	1.04	96	29	1.13	94	21
Reference image						
Observer threshold	0.98	96	16	1.33	96	29
Group threshold	1.03	94	21	1.17	96	16
Observer D						
Own image						
Observer threshold	0.28	96	3	1.24	96	29
Group threshold	1.04	90	21	1.13	98	16
Reference image						
Observer threshold	0.28	96	3	1.24	96	29
Group threshold	1.03	90	16	1.17	98	18
Observer E						
Own image						
Observer threshold	0.97	96	29	0.96	96	18
Group threshold	1.04	88	34	1.13	86	32
Reference image						
Observer threshold	0.96	96	11	1.01	96	13
Group threshold	1.03	90	11	1.17	90	21

Note.—The optimal group and observer ratio thresholds for width and area were determined by obtaining a sensitivity of at least 95% while maximizing specificity for own image and reference image data. The group threshold was then applied to each observer's own image and reference image data to obtain a second set of sensitivities and specificities, which allowed for comparison of individual observer performance.

* Numbers in parentheses are 95% confidence intervals.

Table 5

Univariate and Multivariate Logistic Regression Analyses

Observer	Univariate <i>c</i> Statistic *			Multivariate <i>c</i> Statistic			
	Height	Width	Area	Aspect Ratio	Height and Aspect Ratio	Width and Aspect Ratio	Height and Aspect Ratio
A	0.72	0.90	0.92	0.72	0.90	0.92	0.92
B	0.73	0.83	0.82	0.70	0.84	0.83	0.85
C	0.71	0.79	0.79	0.63	0.75	0.79	0.80
D	0.62	0.75	0.77	0.68	0.70	0.74	0.79
E	0.47	0.74	0.67	0.74	0.76	0.76	0.78
							0.92

Note.—For all *c* statistics, standard errors ranged from 0.03 to 0.06.

* Univariate *c* statistics are equivalent to A_z values, which appear in Table 3.

Table 6

Comparison of A_z Values and P Values for Pairs of Observers

Observer	A_z Value *	P Value			
		Observer B	Observer C	Observer D	Observer E
Width					
A	0.90 (0.04)	.08	.01	<.01	<.001
B	0.83 (0.03)	...	.39	.24	.08
C	0.79 (0.04)	...	...	.58	.32
D	0.75 (0.05)	...	...	...	.74
E	0.74 (0.05)	...	...	...	...
Area					
A	0.92 (0.03)	.01	<.01	.01	<.001
B	0.82 (0.04)	...	.58	.41	<.01
C	0.79 (0.04)	...	...	.68	.01
D	0.77 (0.04)	...	...	...	.07
E	0.67 (0.05)	...	...	...	...

Note.—For any given pair, a P value of less than .05 represents a statistically significant difference in observer performance at the 95% confidence level. The overall P value for the entire group of observers was less than .001 for both width and area.

* Numbers in parentheses represent standard errors.