Differentiating Between Malignant and Benign Breast Masses: Factors Limiting Sonoelastographic Strain Ratio

Ziel: Ermittlung des Stellenwerts der semiquantitativen Strain Ratio (SR) in der Dignitätsbeurteilung von Mammatumoren unter Berücksichtigung limitierender Faktoren


Ergebnisse: Die 116 malignen Tumoren wiesen eine signifikant größere mittlere SR auf als die 108 benignen Tumoren (3,04 ± 0,9 vs. 1,91 ± 0,75; p < 0,001). Die ROC-Analyse ergab für die SR einen AUC(95 % CI 0,777; 0,888), für den ES 0,869 (95 % CI 0,822; 0,917) und für die B-Mode-Sonografie 0,822 (95 % CI 0,768; 0,876). Unter Nutzung eines Cut-off-Wertes von 2,0 weist die SR eine Sensitivität von 90,7 % und eine Spezifität von 58,2 % auf (ES 87,9 % und 73,1 %, B-Mode-Sonografie 97,4 % und 42,6 %). Die Multivarianzanalyse zum Einfluss von Brustdichte, Tumorgröße und Hautabstand zeigt eine negative Korrelation zwischen Hautabstand und Genauigkeit der SR.

Schlussfolgerung: Beide Elastografieverfahren zeigen im Vergleich zur B-Mode-Sonografie eine höhere Spezifität bei etwas geringerer Sensitivität. Bei oberflächlich gelegenen Tumoren ist die SR limitiert verwertbar.

Abstract

Purpose: We compared strain ratio vs. qualitative elastography for the further differentiation of focal breast lesions, with special focus on limiting factors.

Materials and Methods: 215 patients with 224 histologically proven breast masses (116 malignant, 108 benign) were prospectively examined using a high-end ultrasound system (Philips iU22) with serial elastography function. B-mode scans and available mammograms were reviewed according to the BI-RADS classification, raw elastogram data was analyzed qualitatively using the Tsukuba score and semiquantitatively by calculating the strain ratio (fat to lesion ratio). For diagnostic performance, Receiver Operating Characteristic (ROC) curve analysis was obtained. A sub-group analysis regarding breast density, lesion size, lesion depth and histological subtypes was performed.

Results: Mean strain ratio values were 3.04 ± 0.9 for malignant and 1.91 ± 0.75 for benign lesions (p < 0.001). The areas under the ROC curve values were 0.832 (95 % CI 0.777; 0.888) for strain ratio, 0.869 (95 % CI 0.822; 0.917) for Tsukuba score, 0.822 (95 % CI 0.768; 0.876) for B-mode ultrasound and 0.853 (95 % CI 0.799; 0.907) for mammography. Sensitivity, specificity, positive predictive value and negative predictive value of the strain ratio were 90.7 %, 58.2 %, 70.3 % and 85.1 %, when a cutoff point of 2.0 was used. Only lesion depth ≤4 mm was associated with diagnostic failure in the multivariate analysis of factors influencing accuracy, whereas no significant correlation between breast density and lesion size and the accuracy of the strain ratio could be found.

Conclusion: The addition of strain ratio to B-mode ultrasound increases specificity without loss of sensitivity in differentiating between malignant and benign breast tumors. Strain ratio measurements should not be carried out on tumors with a lesion depth ≤4 mm.
Introduction

Over the last 20 years, sonoelastography has developed from a technically complex examination method to one that is simple to carry out and reproduce, and that can be integrated into clinical examination procedures [1]. Various manufacturers of ultrasound devices have integrated elastography as a standard feature, and its use in a range of clinical settings is the subject of numerous current studies [2–5]. Previous studies have provided evidence for the benefits of sonoelastography as an additional procedure in breast ultrasound [6–11]. A 5-stage elasticity score has been established by Ueno et al. as the standard in elastography evaluation [12]. However, the use of this score requires a certain amount of experience; the interobserver and intraobserver variability is considerable [13]. The calculation of the strain ratio is a semi-quantitative measurement that indicates the average deformation of a lesion in relation to the deformation in a reference region (usually the lateral, surrounding fatty tissue) [14]. The measurement results in a metric value with no specified dimensions. More recent studies have shown a significant difference between the average strain ratio of benign and malignant breast tumors [6, 15–18].

The aim of our study was to determine the value of strain ratio to discriminate between malignant and benign breast tumors in comparison to the qualitative 5-stage elasticity score and conventional B-mode ultrasound. Secondary study aim was to investigate factors that limit the use of sonoelastography.

Materials and Methods

Patients

Between January 2010 and October 2010, a total of 215 patients with 224 breast masses were prospectively enrolled in the study. B-mode ultrasound was carried out on all patients, and mammograms were available for 190 patients. Mammography was not carried out on 25 patients due to their young age. After informed consent, real-time elastographic examination was carried out and the data sets were saved. Final diagnoses were made on the basis of histological findings. The local institutional review board approved the study.

Data Acquisition/Imaging Methods

Conventional ultrasound images and real-time elastographic data sets were obtained from three very experienced ultrasound examiners using a 12.5-MHz linear transducer (Philips IU22, Bothell, WA, USA). The principles of ultrasound real-time elastography are well-known [12, 19, 20]. The sonoelastographic function of the Philips software is based on tissue compression during thoracic respiratory excursions. In order to obtain appropriate images, the transducer has to be applied with only the light pressure necessary to maintain contact with the skin. The results are displayed as color-coded images, with blue representing hard tissue and red to green representing soft tissue on a continuous scale (Fig. 1). In addition, for semi-quantitative evaluation of the strain ratio, a 60-second cineloop was obtained during elastographic examination. The strain ratio describes the fat to lesion ratio, indicating the stiffness of a lesion. It is obtained by setting a rectangular Region of Interest (ROI) to cover the subcutaneous fatty tissue and a second ROI to cover the lesion. The Philips iU22 software package features the possibility of retrospectively analyzing the strain ratio (known as „deformation quotient“). The cineloop can be used to create a deformation quotient curve, which can in turn be used to calculate the mean and maximum deformation quotients (Fig. 2).

All acquired data were reviewed by an experienced ultrasound examiner. The B-mode images were evaluated according to the Breast Imaging Reporting and Data System (BI-RADS) classification. BI-RADS categories 1 to 3 were assumed to be benign, and BI-RADS categories 4 and 5 were taken as malignant. The elastographic data sets were analyzed qualitatively using the Tsukuba elastography score. This is a color-coded 5-stage score with score 1 revealing the whole lesion in green and score 5 revealing the lesion and its surrounding tissue in blue. Lesions categorized as score 1, 2 or 3 are assumed to be benign, while lesions categorized as score 4 or 5 are assumed to be malignant. Further details can be found elsewhere [12].

In order to carry out a semi-quantitative elastographic evaluation, the strain ratio of the lesions was obtained. According to the user manual, the most appropriate single image should be selected for further analysis. Since we observed significant intraobserver variability, the following protocol was used: All measurements (setting the ROIs) were carried out three times. A mean was calculated from the three results for mean deformation quotient as well as maximum deformation quotient. This procedure
was carried out with a single image, with a sequence of 20 images, and with the entire cineloop (60 images).

Statistics
Statistical analysis was carried out in collaboration with an independent statistician (G.K.) using the SPSS 19.0 software package (IBM Ehningen, Germany). First, descriptive statistics were computed for variables of interest. The statistics computed included mean, median, minimum, maximum and standard deviations of continuous variables, frequencies and percentage frequencies of categorical factors. Additionally, 95% confidence intervals are presented. Furthermore, comparison of the strain ratio means between benign and malignant masses was realized by using Student t test. In order to further evaluate the performance of strain ratio in the differentiation between benign and malignant lesions, a Receiver Operating Characteristic (ROC) analysis was carried out to compare the area under the curve (AUC). The best results were obtained using the mean strain ratio of the sequence image analysis, so these data were used for further evaluation. The diagnostic sensitivity, specificity, positive and negative predictive values and accuracy were calculated for the qualitative elasticity score and the strain ratio in comparison with mammography and conventional ultrasound. Finally, we compared the diagnostic properties of the strain ratio according to lesion size, distance from the skin (lesion depth), breast density and histological subtypes. All p-values resulted from two-sided statistical tests and p ≤ 0.05 was considered to be significant.

Results

Study population
The patients’ median age was 56 years, with a range of 12 to 90 years. Of the 224 lesions, 116 (51.8%) were malignant and 108 (48.2%) were benign. The lesion size ranged from 4 to 110 mm with a median of 15 mm. There were 125 (55.8%) palpable masses among the study population. Concerning breast density, 10% revealed fatty tissue, 47% scattered fibroglandular tissue, 35% heterogeneously dense tissue and 8% extremely dense tissue.

Strain ratio
The mean strain ratio for malignant breast tumors was 3.04 ± 0.9 (mean ± standard deviation SD, range 0.86 to 5.43) and was significantly higher than for benign breast tumors (1.91 ± 0.75, range 0.31 to 3.91) (Fig. 3). Fig. 4 shows the diagnostic performance of the strain ratio using the ROC curve. At a cutoff of ≤2.0 for benign tumors and > 2.0 for malignant tumors, sensitivity was 90.7% (95% CI 83.6%; 94.8%), specificity was 59.2% (95% CI 48.3%; 67.4%), the positive predictive value (PPV) was 70.3%
(95% CI 62.2%; 77.3%), the negative predictive value was 85.1% (95% CI 74.7%; 91.7%) and accuracy was 75.1% (95% CI 69.2; 81).

Diagnostic Performance in Comparison to Mammography, B-mode Ultrasound and Qualitative SonoeLASTography (5-point scale)

Table 1 shows the sensitivity, specificity, PPV, NPV and accuracy in comparison to conventional methods (mammography and B-mode ultrasound) and to qualitative sonoelastography using a 5-point scale (1–3 benign, 4–5 malignant). Although the sensitivity of elastography is lower than for B-mode ultrasound, it provides higher specificity. At 80.8%, qualitative sonoelastography provides the greatest accuracy.

If we consider the sub-group with BIRADS 4 in B-mode ultrasound, this results in a pretest probability for a malignant tumor of 48%; after including sonoelastography for scale 1–3/strain ratio ≤ 2.0, this results in a posttest probability of 25%/25.8%; and for scale 4–5/strain ratio > 2.0, this results in a posttest probability of 67.2%/56.9%.

Strain Ratio Depending on Tumor Size, Distance from Skin, Breast Density and Histological Diagnosis

The extent to which the predictive accuracy of strain ratio depends on external factors was investigated using bivariate correlation analysis. The analysis shows a statistically significant negative correlation between the distance of the tumor from the surface of the skin and the correct estimation of whether the tumor is malignant or benign (p = 0.038). If we compare the group of tumors located close to the surface (skin distance ≤ 4 mm) with deeper lesions (skin distance > 4 mm), the latter group displays considerably higher sensitivity (94.2 vs. 75%) and specificity (61.5 vs. 48.4%).

Breast density and tumor size did not influence the diagnostic performance of the strain ratio.

Fig. 5 shows the average strain ratio of the various histological diagnoses. It is shown that for malignancies, the average strain ratio is considerably higher than for benign histologies. Particularly, this applies to invasive lobular breast carcinomas (mean 3.69; range 2.62–5.43; n = 10), which are often difficult to judge using mammography and sonography. Of the benign tumors, fatty necrosis (mean 2.78; range 1.67–3.62; n = 7) and radial scars (mean 2.46; range 1.22–3.36; n = 4) show a high strain ratio. The diagnostic accuracy of the various methods (mammography, B-mode ultrasound, 5-point scale sonoelastography and strain ratio) in the different histological categories is shown in Table 2. Of the malignant tumors, mucinous breast carcinoma is not identified as malignant by sonoelastography (5-point scale) (0/3) and is identified as malignant by strain ratio in 2/3 cases. In identifying cysts as benign, sonoelastography outperforms B-mode ultrasound (5-point scale elastography 9/11, strain ratio 10/11, B-mode ultrasound 5/11). In all examination methods, fat necrosis and radial scars represent a diagnostic challenge.
Discussion

In the presented study, the raw data sets of the elastograms saved in DICOM format were retrospectively blinded and evaluated, and the strain ratio was calculated for each data set. However, integration into the current examination process does not seem to involve any problems and requires an additional time of approximately 2–3 minutes. This corresponds to the experiences of other working groups [21].

Previous studies have shown the potential of elastography for differentiating between benign and malignant breast tumors [6, 8, 16, 23]. Our study reveals a significant difference in the average strain ratio of benign and malignant lesions. At a cutoff value of ≤ 2.0, both the strain ratio and the 5-point score demonstrate a higher specificity and higher PPV than B-mode ultrasound and mammography, albeit with a slightly lower sensitivity. In terms of accuracy, 5-point score elastography was superior to the other examination methods. Other studies have demonstrated an increase in NPV with additional use of elastography [7, 18]. However, at an NPV of 85% for both elastography methods, histological findings for tumors located close to the surface of the skin. To our knowledge, these findings have not yet been described. In a multi-center study by Wojcinski et al., the following cal entities. A good differentiation of malignant versus benign findings is achieved for invasive ductal carcinomas and invasive lobular carcinomas. The latter could be a valuable addition, as invasive lobular carcinoma is often insufficiently diagnosed when using conventional methods (mammography, B-mode ultrasound). For mucinous carcinomas, strain ratio outperforms 5-point score elastography. However, the results must be interpreted with caution due to the low number of cases (n = 3). In contrast to the study by Mori [25], our own data supports that the elastography score is not suitable for differentiating between mucinous breast carcinomas and fibroadenomas.

In terms of false positive findings, the histological finding categories of fat necrosis and radial scars appear to be problematic. While fat necrosis is reliably diagnosed using mammography, radial scars remain a diagnostic challenge that, ultimately, can only be verified as benign by means of histological methods.

There are limitations in the retrospective design of the study and in the evaluation of the elastography data by just one investigator. It is not possible to provide data on interobserver variability. In addition, it is difficult to compare the results directly with other working groups as different cutoff values have been established for the strain ratio depending on the device [7, 16–18]. This means that, for the time being, strain ratio will continue to be implemented in the diagnostic algorithm for differentiating from 75.1 (69.2–80) to 71.9 (72–83.8). However, the results must be interpreted with caution due to the low number of cases (n = 3). In contrast to the study by Mori [25], our own data supports that the elastography score is not suitable for differentiating between mucinous breast carcinomas and fibroadenomas.

In terms of false positive findings, the histological finding categories of fat necrosis and radial scars appear to be problematic. While fat necrosis is reliably diagnosed using mammography, radial scars remain a diagnostic challenge that, ultimately, can only be verified as benign by means of histological methods.

There are limitations in the retrospective design of the study and in the evaluation of the elastography data by just one investigator. It is not possible to provide data on interobserver variability. In addition, it is difficult to compare the results directly with other working groups as different cutoff values have been established for the strain ratio depending on the device [7, 16–18]. This means that, for the time being, strain ratio will continue to be implemented in the diagnostic algorithm for differentiating from 75.1 (69.2–80) to 71.9 (72–83.8). However, the results must be interpreted with caution due to the low number of cases (n = 3). In contrast to the study by Mori [25], our own data supports that the elastography score is not suitable for differentiating between mucinous breast carcinomas and fibroadenomas.

In terms of false positive findings, the histological finding categories of fat necrosis and radial scars appear to be problematic. While fat necrosis is reliably diagnosed using mammography, radial scars remain a diagnostic challenge that, ultimately, can only be verified as benign by means of histological methods.

There are limitations in the retrospective design of the study and in the evaluation of the elastography data by just one investigator. It is not possible to provide data on interobserver variability. In addition, it is difficult to compare the results directly with other working groups as different cutoff values have been established for the strain ratio depending on the device [7, 16–18]. This means that, for the time being, strain ratio will continue to be implemented in the diagnostic algorithm for differentiating from 75.1 (69.2–80) to 71.9 (72–83.8). However, the results must be interpreted with caution due to the low number of cases (n = 3). In contrast to the study by Mori [25], our own data supports that the elastography score is not suitable for differentiating between mucinous breast carcinomas and fibroadenomas.

In terms of false positive findings, the histological finding categories of fat necrosis and radial scars appear to be problematic. While fat necrosis is reliably diagnosed using mammography, radial scars remain a diagnostic challenge that, ultimately, can only be verified as benign by means of histological methods.

There are limitations in the retrospective design of the study and in the evaluation of the elastography data by just one investigator. It is not possible to provide data on interobserver variability. In addition, it is difficult to compare the results directly with other working groups as different cutoff values have been established for the strain ratio depending on the device [7, 16–18]. This means that, for the time being, strain ratio will continue to be implemented in the diagnostic algorithm for differentiating from 75.1 (69.2–80) to 71.9 (72–83.8). However, the results must be interpreted with caution due to the low number of cases (n = 3). In contrast to the study by Mori [25], our own data supports that the elastography score is not suitable for differentiating between mucinous breast carcinomas and fibroadenomas.

In terms of false positive findings, the histological finding categories of fat necrosis and radial scars appear to be problematic. While fat necrosis is reliably diagnosed using mammography, radial scars remain a diagnostic challenge that, ultimately, can only be verified as benign by means of histological methods.

There are limitations in the retrospective design of the study and in the evaluation of the elastography data by just one investigator. It is not possible to provide data on interobserver variability. In addition, it is difficult to compare the results directly with other working groups as different cutoff values have been established for the strain ratio depending on the device [7, 16–18]. This means that, for the time being, strain ratio will continue to be implemented in the diagnostic algorithm for differentiating from 75.1 (69.2–80) to 71.9 (72–83.8). However, the results must be interpreted with caution due to the low number of cases (n = 3). In contrast to the study by Mori [25], our own data supports that the elastography score is not suitable for differentiating between mucinous breast carcinomas and fibroadenomas.

In terms of false positive findings, the histological finding categories of fat necrosis and radial scars appear to be problematic. While fat necrosis is reliably diagnosed using mammography, radial scars remain a diagnostic challenge that, ultimately, can only be verified as benign by means of histological methods.

There are limitations in the retrospective design of the study and in the evaluation of the elastography data by just one investigator. It is not possible to provide data on interobserver variability. In addition, it is difficult to compare the results directly with other working groups as different cutoff values have been established for the strain ratio depending on the device [7, 16–18]. This means that, for the time being, strain ratio will continue to be implemented in the diagnostic algorithm for differentiating from 75.1 (69.2–80) to 71.9 (72–83.8). However, the results must be interpreted with caution due to the low number of cases (n = 3). In contrast to the study by Mori [25], our own data supports that the elastography score is not suitable for differentiating between mucinous breast carcinomas and fibroadenomas.
breast lesions on a center and device-specific basis. No general recommendation can be provided due to a lack of standardization in the examination process, including in the software required for computation. „Shear wave ultrasound elastography“ facilitates quantitative measurement of tissue elasticity and represents an interesting development in this regard [23]. A recently published multi-center study by Berg et al. [26] showed that the addition of shear wave elastography to B-mode ultrasound improved specificity without reducing sensitivity. Moreover, shear wave elastography reveals a high intra- and interobserver reproducibility [27].

Acknowledgment

We thank Katrin Herbst (Philips Healthcare GmbH Hamburg, Germany) for providing a native speaker for the correction of our manuscript.

References

27 Cosgrove DO, Berg WA, Dore CJ et al. Shear wave elastography for breast masses is highly reproducible. Eur Radiol 2012; 22: 1023 – 1032