European Journal of Radiology 101 (2018) 157–161

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research Article

Breast strain elastograhphy: Observer variability in data acquisition and interpretation

YiJie Dong[1], Chun Zhou[1], JianQiao Zhou*, ZhiFang Yang, JinWen Zhang, WeiWei Zhan

Department of Ultrasound, Ruijin Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai 200025, China

ARTICLE INFO

Keywords:
Ultrasound
Elastography
Elasticity score
Breast lesion

ABSTRACT

Objective: To analyze the observer reproducibility of breast strain elastography in elasticity image acquisition and elasticity image interpretation.

Methods: This was an institutional ethics committee approved prospective study. One hundred twenty-four breast lesions in 118 women (mean age 45.39 ± 12.97 years old, range 21–77 years old) were examined with strain elastography by two blinded radiologists in turn. Three blinded observers separately reviewed and recorded the elasticity score of each lesion obtained by the two performers. The interobserver reproducibility of elasticity image acquisition between the two performers, the interobserver and intraobserver reproducibility of elasticity image interpretation among observers were evaluated. The diagnostic performance of strain elastography was compared between the two performers.

Results: Fifty-three lesions were malignant and 71 were benign. The interobserver kappa value was 0.438 for the elasticity score between the two performers. Between the three observers, the overall interobserver and intraobserver kappa value was 0.365 and 0.655, respectively. There was no significant difference of the area under the receiver operator characteristic curve (Az) value for the elasticity score between performer 1 and 2 (P = 0.143).

Conclusions: Our results suggested moderate interobserver reproducibility in breast strain elasticity image acquisition, poor interobserver and good intraobserver agreement in image interpretation.

1. Introduction

Real-time strain elastography is a noninvasive image technique that provides the stiffness information of a lesion[1]. Breast elastography has become a routine tool in addition to conventional ultrasound (US) and has been widely used in clinical practice. Studies showed that elastography has good diagnostic value in differentiating benign from malignant breast lesions, and could reduce the number of unnecessary breast biopsies [2,3]. As a result, guidelines and recommendations had been developed for the use of breast elastography[4,5], and elastography had been incorporated into the new edition of the Breast Imaging Reporting and Data System (BI-RADS) US classification[6].

In spite of the very good diagnostic performance, however, there is one main problem with breast strain elastography, i.e., its reproducibility, which may limit the usefulness of strain elastography in routine clinical practice. Up to now, there have been several reports on the reproducibility of breast elastography and yielded inconsistent results [7–17]. For interobserver reproducibility, some studies revealed that the agreement was moderate to substantial for the elasticity score [9,12–14], while others showed significant interobserver performance variability[8,10]. To the best of our knowledge, there was only one study that assessed the intraobserver reproducibility of strain elastography, in which the kappa value for intraobserver reproducibility was 0.720 for the elasticity score[9].

The observer variability of strain elastography may be caused by both elasticity image acquisition and elasticity image interpretation[8]. However, except for one study carried out by Yoon et al. [15], all previous reproducibility studies only demonstrated the observer reproducibility in elasticity image interpretation but not in elasticity image acquisition.

The aim of this study was to analyze the observer reproducibility of breast strain elastography in both elasticity image acquisition and elasticity image interpretation.

* Corresponding author.
E-mail address: zhousu30@126.com (J. Zhou).
[1] These authors contributed equally to this work and should be considered co-first authors.

https://doi.org/10.1016/j.ejrad.2018.02.025
Received 2 October 2017; Received in revised form 12 February 2018; Accepted 19 February 2018

2. Materials and methods

2.1. Patients

This study was approved by the ethics committee of our hospital, and informed consent was obtained from each patient. From August 2014 through October 2014, 136 breast lesions in 126 consecutive women with symptomatic and screen-detected masses underwent conventional and elastographic US. 12 lesions in 8 women were later excluded because of the following reasons: 5 lesions in 3 patients lack of complete US data, 4 lesions in 3 patients with simple breast cyst confirmed by conventional US, 3 lesions in 2 patients without histopathological proof. Finally, 124 breast lesions in 118 women (mean age 45.39 ± 12.97 years old, range 21–77 years old) were included in the final data analysis.

2.2. Image acquisition

Conventional US and elastographic examinations were performed using an Esaote MyLab 90 US scanner (Esaote SpA, Genoa, Italy) equipped with a 4–13 MHz linear probe. Whole breast US examination was first performed by one of 2 blinded radiologists (performer 1, J.Q.Z and performer 2, C.Z), each having 17 years of experience in breast US. Then, the elastographic examination of the target lesion was performed by the two radiologists in turn. Both of them underwent training session in breast elastography prior to the beginning of the study.

For elastography, the built-in elaXto software was activated and then the images were displayed in a split-screen mode with the conventional US image in the left and the same image with translucent color-coded elasticity features superimposed in the right. A region of interest (ROI) box was set to include the subcutaneous fat, breast tissue, and the pectoralis muscle with the target lesion in the center. Using a freehand technique, the breast was vertically compressed by the transducer under slight and constant pressure. The real-time elaXto-spring tool, displayed on the screen, was used to help the operator to obtain correct elastograms. One reproducible elastogram was selected and digitally recorded by the operator when the elaXto-spring turned green and lasted for at least 3 or 4 consecutive frames. In elastograms, the hardness of tissue is displayed in color tone using a scale from green (softest component), blue (intermediate component) to red (hardest component).

Both conventional and elastographic images were captured per case and saved on a hard disk.

2.3. Image analysis

The qualitative evaluations were performed by three blinded observers (Y.J.D, Z.F.Y, and J.W.Z) had more than 7 years' experience in breast US and more than 4 years' experience in breast elastography. The Tsukuba score system proposed by Itoh et al. [18] was used. A score of 1 indicated even strain for the entire lesion. A score of 2 indicated strain in most of the lesion. A score of 3 indicated strain at the periphery with sparing of the center of the lesion. A score of 4 indicated no strain in the entire lesion. A score of 5 indicated no strain in the entire lesion and the surrounding tissue.

In qualitative evaluation, first, the three observers separately reviewed and recorded the elasticity score of the static elastographic images of each lesion obtained by the two performers (J.Q.Z and C.Z). The data of the images obtained by the performer 1 were used for calculating the interobserver reproducibility of elasticity image interpretation among observers. Then, for calculating the interobserver reproducibility of elasticity image acquisition between the two performers, the observers together compared and discussed the elasticity score results obtained in the first step, and, using the principle of consensus, reached final elasticity score of each lesion obtained by each of the two performers (Fig. 1). These final elasticity scores were used to calculate the diagnostic performance of the five-point scoring system. Finally, with a delay of 6 months after the first evaluation, the three observers again separately reviewed and recorded the elasticity score of the static elastographic image of each lesion obtained by the performer 1 for calculating the intraobserver reproducibility of elasticity image interpretation.

2.4. Pathological examination

After elastographic examination, all 124 lesions were examined histologically on the resection or biopsy specimen. All specimens were subjected to Hematoxylin and Eosin, and Immunohistochemical staining. All diagnoses were made by an experienced pathologist with 29 years of experience in breast cancer analysis.

2.5. Statistical analysis

Statistical analyses were performed using SPSS version 17.0 (SPSS Inc, Chicago, IL). Receiver operator characteristic (ROC) analysis was performed using MedCalc for Windows, version 12.2.0.0 (MedCalc Software, Mariakerke, Belgium). The nonparametric Mann-Whitney *U* test was used to compare the elasticity score. The kappa value was used to calculate the agreement of elasticity score. The strength of agreement was set as the following: 0–0.20, slight agreement; 0.21–0.40, poor agreement; 0.41-0.60, moderate agreement; 0.61-0.80, good agreement; greater than 0.80, excellent agreement[19]. The areas under the ROC curves (Az) were calculated to compare the diagnostic performances. $P < 0.05$ was considered statistically significant.

3. Results

3.1. Pathological diagnoses

Histological analysis revealed that out of 124, 53 were malignant and 71 were benign. Final pathological diagnoses were shown in Table 1. Fibroadenoma was the commonest benign lesions, and invasive ductal carcinoma was the commonest malignant lesions.

3.2. Interobserver reproducibility between observers

The elasticity scores of the images obtained by the performer 1 evaluated by observers 1, 2, and 3 were 2.42 ± 0.97 (1–5), 3.12 ± 1.27 (1–5), and 2.77 ± 1.08 (1–5), respectively. The interobserver kappa value between observer 1 and observer 2 was 0.281, between observer 1 and observer 3 was 0.351, between observer 2 and observer 3 was 0.464, between the three observers was 0.365.

3.3. Intraobserver reproducibility of observers

For the elasticity scores of the images obtained by the performer 1, the intraobserver kappa values for observer 1, 2 and 3 were 0.699, 0.581, and 0.686, respectively. The overall intraobserver kappa value for observers was 0.655.

3.4. Interobserver reproducibility between performers

Between the two performers, the kappa value was 0.438 for elasticity score.

3.5. Diagnostic performance of strain elastography

Table 2 showed the diagnostic performance of strain elastography for each performer. For both performers, the mean elasticity score was significantly higher for malignant lesions than for benign ones ($P < 0.001$). There was no significant difference of the Az values for the elasticity score between performer 1 and 2 ($P = 0.143$).