Eur Radiol (2013) 23:1006–1014
DOI 10.1007/s00330-012-2682-0

BREAST

Impact of breast mass size on accuracy of ultrasound elastography vs. conventional B-mode ultrasound: a meta-analysis of individual participants

Gelareh Sadigh · Ruth C. Carlos · Colleen H. Neal · Sebastian Wojcinski · Ben A. Dwamena

Received: 27 April 2012 / Revised: 1 September 2012 / Accepted: 30 September 2012 / Published online: 20 October 2012

Abstract
Objectives To conduct an individual patient data meta-analysis comparing the diagnostic performance of ultrasound elastography (USE) versus B-mode ultrasound (USB) across size ranges of breast masses.
Methods An extensive literature search of PubMed and other medical/general purpose databases from inception through August 2011 was conducted. Corresponding authors of published studies that reported a direct comparison of the diagnostic performance of USE using the elasticity score versus USB for characterisation of focal breast masses were contacted for their original patient-level data set. Summary diagnostic performance measures were compared for each test within and across three mass size groups (<10 mm, 10–19 mm, and >19 mm).
Results The patient-level data sets were received from five studies, providing information on 1,412 breast masses. For breast masses <10 mm (n=543; 121 malignant), the sensitivity/specificity of USE and USB were 76 %/93 % and 95 %/68 %, respectively. For masses 10–19 mm of size (n=528; 247 malignant), sensitivity/specificity of USE and USB were 82 %/90 % and 95 %/67 %, respectively. For masses >19 mm of size (n=325; 162 malignant), sensitivity/specificity of USE and USB were 74 %/94 % and 97 %/55 %, respectively.
Conclusion Regardless of the mass size, USE has higher specificity and lower sensitivity compared to USB in characterising breast masses. The performance of each of these two tests does not vary significantly by mass size.
Key Points
- *Ultrasound elastography is increasingly used for breast lesions.*
- *Its diagnostic performance is not dependent on the size of the mass.*
- *Ultrasound elastography has higher specificity/lower sensitivity than B-mode ultrasound.*
- *Elastography is advised when B-mode results are equivocal.*

Keywords Breast mass · Breast ultrasound · Elastography · Malignancy · Size

Electronic supplementary material The online version of this article (doi:10.1007/s00330-012-2682-0) contains supplementary material, which is available to authorized users.

G. Sadigh (✉)
Department of Radiology, University of Michigan Medical Center, 1500 E. Medical Center Dr, B1 132H Taubman Center, Ann Arbor, MI 48109-5302, USA
e-mail: gelarehsadigh@yahoo.com

R. C. Carlos
Department of Radiology, Division of Magnetic Resonance Imaging, University of Michigan Medical Center, Ann Arbor, MI, USA

R. C. Carlos · C. H. Neal · B. A. Dwamena
VA Ann Arbor Healthcare System, Ann Arbor, MI, USA

C. H. Neal
Department of Radiology, Division of Breast Imaging, University of Michigan Medical Center, Ann Arbor, MI, USA

S. Wojcinski
Department of Obstetrics and Gynecology, Franziskus Hospital, Bielefeld, Germany

B. A. Dwamena
Department of Radiology, Division of Nuclear Medicine, University of Michigan Medical Center, Ann Arbor, MI, USA

Abbreviations

BI-RADS	Breast imaging reporting and data system
CI	Confidence interval
I^2	Inconsistency index
IPD	Individual patient data
LR	Likelihood ratio
USB	Breast B-mode ultrasound
USE	Breast ultrasound elastography
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS	Quality Assessment of Diagnostic Accuracy Studies

Introduction

Breast ultrasound elastography (USE) is a new breast imaging technique that may assist with further characterisation of breast masses as benign or malignant. USE estimates the level of change in size (strain) of a breast mass in response to external compression relative to its surrounding tissue [1–3]. The test output is a coloured image. Each colour can be translated into a relative level of tissue strain. The coloured image will be superimposed on a B-mode ultrasound (USB) image for better evaluation of strain distribution within the anatomical borders of the mass [4]. A 1–5-scale elasticity score is assigned to each breast mass based on its overall strain distribution, with the harder tissues (e.g. breast cancer) showing higher elasticity scores [4].

The diagnostic performance of USE using the elasticity score was previously reported in two meta-analyses, based on published aggregated data, with a sensitivity of 79 % to 83 % and a specificity of 84 % to 89 % [5, 6]. In direct comparison to USB, USE improves the specificity (from 70 % in USB to 88 % in USE) [5], while decreasing the sensitivity from 96 % in USB to 79 % in USE [5].

Several individual studies have reported that elasticity score may perform better than USB in the characterisation of small masses (<1 cm) [7–9]. However tissue stiffness (in USE) is an intrinsic material property and should not depend on the mass size [7]. On the other hand, small malignant breast masses may have benign morphology, making them challenging to diagnose by USB [8].

Individual-patient data (IPD) meta-analyses are the gold standard approach for conducting a meta-analysis. These meta-analyses require attaining patient-level data sets (source data) from individual studies, which facilitates accounting for the impact of patient-level variability within each study, as well as between study heterogeneity on diagnostic test accuracy [10]. However, IPD meta-analyses require more time and resources to perform compared to meta-analyses based on published aggregated data. Therefore, aggregated data meta-analyses, are more commonly conducted by researchers [11, 12].

We performed a meta-analysis of individual patient data of studies that reported a direct comparison of the elasticity score with the breast imaging reporting and data system (BI-RADS) [13] in differentiating breast masses. The aims of this study were (1) to compare the diagnostic performance of each of these tests (USE and USB) among three different size ranges of breast masses (e.g. <10 mm, 10–19 mm, and >19 mm) and (2) to evaluate the diagnostic performance of USE compared to USB within each of these three size ranges of breast masses.

Methods

The study was conducted and reported using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) recommendations [14].

Criteria for considering studies for this review

Types of studies

All analytical studies reporting a direct comparison of the elasticity score alone with BI-RADS in the differentiation of focal breast masses that were published in a full text were considered for eligibility. No language restriction was used.

Participants

Study participants were patients who were referred because of specific diagnostic queries such as breast symptoms (e.g. palpable breast mass, pain, nipple discharge), an abnormal clinical breast examination, USB or mammography, or intensified screening in high-risk populations. There was no age restriction for the study participants.

Index test

Breast ultrasound elastography (USE) was the index test used to differentiate benign from malignant breast masses. Only papers in which a 5-point scale elasticity score according to Itoh et al. [4] was calculated were included. Masses with an elasticity score of 4 and 5 were considered malignant, while the other elasticity scores were grouped as benign masses.

Comparator test

Conventional B-mode ultrasound (USB) was the comparator. USB images were reported according to BI-RADS

categories [13]. Masses with BI-RADS categories of 4 and 5 were considered malignant, while the other categories were grouped as benign masses.

Reference standards

Histopathological confirmation of breast mass (core biopsy or surgical biopsy) or cytological confirmation (fine-needle aspiration) with short-term follow-up is the reference standard.

Search methods for identification of studies

For the purpose of the current study, we used the search result of our prior meta-analysis [5]. Electronic searches of PubMed, EMBASE, ISI Web of Knowledge, and the Cochrane database from inception through to 22 August 2011 were performed without any constraints. We used relevant text words and Medical Subject Heading terms that included *breast* combined with *sonoelastography, elastosonography, elastography, elasticity imaging*, and *strain imaging*. Reference lists from identified studies were manually scanned to identify other relevant studies.

Data collection and analysis

Selection of studies

Two authors (G.S. and B.A.D.) independently conducted the literature search. A list of articles meeting the inclusion criteria based on abstracts was complied, and these articles were retrieved in full text. The same two authors independently reviewed the list of full texts for inclusion. Discrepancies were discussed and resolved upon agreement on a final set of studies. The original patient-level data set was requested by directly corresponding with authors of eligible studies.

Data extraction and management

Original patient-level data sets, received from the corresponding authors of eligible studies, included individual patient age, gender, breast mass dimensions, elasticity score, BI-RADS score, and pathology diagnosis.

Assessment of methodological quality

Two reviewers (G.S. and B.A.D.) independently assessed the methodological quality of included studies, using 11 items of the extensively validated Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist [15]. Disagreements were resolved by consensus.

Statistical analysis and data synthesis

The size of breast masses was categorised as less than 10 mm, 10–19 mm, and more than 19 mm. Summary sensitivity (true-positive fraction) and specificity (true-negative fraction) and 95 % confidence interval (CI) for USE and USB were calculated across three mass size ranges using the Bernoulli distribution for bivariate generalised linear mixed modelling (a random effects model) [16, 17]. The model was estimated using the adaptive quadrature method, which is available within the GLLAMM (generalised linear latent and mixed models) module [18, 19] in Stata software. Summary positive and negative likelihood ratios (LR) were calculated from the model estimates. LRs can be interpreted as follows: an LR of 0 excludes disease, an LR of infinity (∞) excludes normality, and an LR of 1 means no change in likelihood of disease. For the diagnostic information to have high probability of altering clinical management, a likelihood ratio greater than 10 or less than 0.1 would be required for a positive or negative test result, respectively. Moderate informational value can be achieved with likelihood ratios of 5–10 and 0.1–0.2; likelihood ratios of 2.0–5.0 and 0.2–0.5 indicate very little informational value [20]. Additionally, we assessed the performance of USE and USB in a subgroup of masses more than 29 mm.

Heterogeneity between studies was assessed by using the inconsistency index (I^2) statistics. I^2 values range between 0 % and 100 %, where 0 % indicates no observed heterogeneity and values greater than 50 % may be considered to indicate substantial heterogeneity [21]. We tried to explain the marked heterogeneity observed with the baseline modules by performing meta-regression analysis of the following variables: the diagnostic test used for characterisation of breast masses, mass size, and pathology.

Statistical significance was set at $p<0.05$. All analyses were conducted using Stata software, version 12 (Stata Corp., College Station, TX).

Results

The literature search from our previous study yielded 2,927 articles, of which 172 were reviewed in abstract; of these 51 were further reviewed in full text (Fig. 1). Of these, 36 studies with information on 6,779 masses were eligible for inclusion. Of the excluded studies, 13 did not fully meet the inclusion criteria, and 2 had used 4-point scale elasticity scores [22, 23], measured differently from the classification introduced by Itoh et al. [4]. After corresponding with the authors of the 36 eligible studies, five [7, 9, 24–26] with information on 1,412 masses provided the original patient-level data set (response rate of 14 % of studies, 21 % of

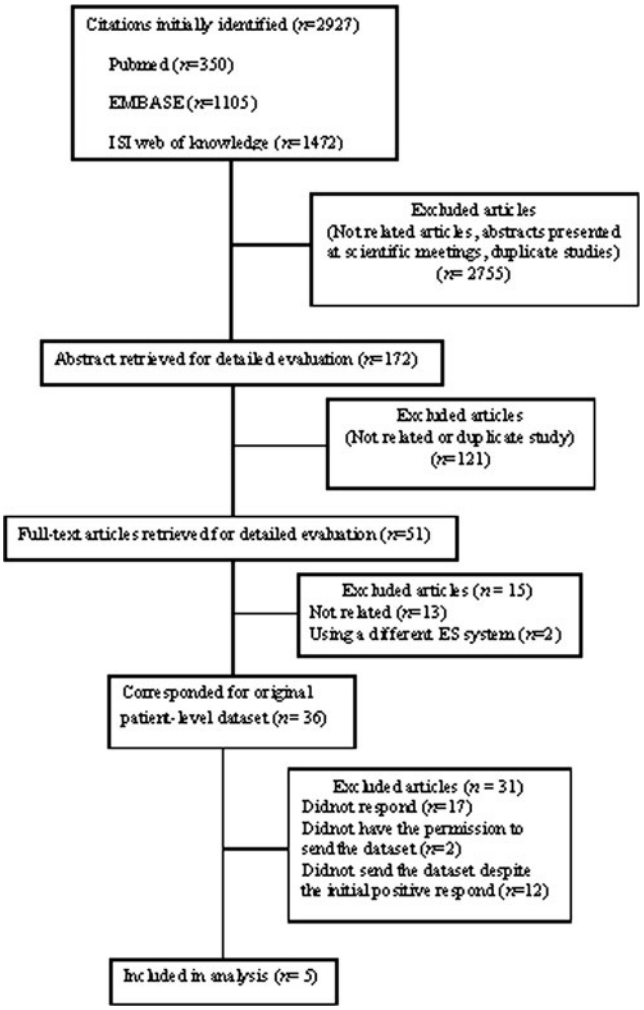

Fig. 1 Literature search and selection schema. *ES* elasticity score

masses). Of the remaining 31 studies, 17 authors did not respond, 2 did not have permission to send the patient-level data set, and 12 did not send the data set despite the initial positive response.

Characteristics of included studies

Included studies were published between January 2008 and February 2011 in peer-reviewed journals (Table 1). Overall, our analysis included information on 1,332 patients and 1,412 breast masses. All the patients were women with an age range of 11 to 94 years old. The mean age was 52±16. Of included masses, 536 (37.9 %) were malignant and 876 (62.1 %) were benign. The longest dimension of the breast masses ranged between 1 and 95 mm with a mean of 14.5±9.7 mm; 1,298 (92 %) of the masses were less than 30 mm. All studies used freehand compression elastography probes and the combined autocorrelation method [4].

Methodological quality of the included studies

All studies used appropriate reference standard(s) for verification, and all of them explained the withdrawals from the study. However, none of the studies clearly stated whether the reference standard interpretation was performed without the knowledge of index test results. Additionally, it was unclear in all studies whether relevant clinical data were available to radiologists when they interpreted the USE and USB images.

Diagnostic performance analyses

The overall test performance of USE and USB for 1,412 masses is summarised in Table 2. USE has a sensitivity and specificity of 78 % (95 % CI, 63 %–88 %) and 92 % (95 % CI, 85 %–96 %), respectively. The I^2 for heterogeneity was 98 % (95 % CI, 98 %–99 %). USB has a sensitivity of 96 % (95 % CI, 93 %–97 %) and a specificity of 65 % (95 % CI, 42 %–82 %). The I^2 for heterogeneity was 96 % (95 % CI, 94 %–98 %). Size was not reported in the source data set for 16 breast masses (0.01 %), which were excluded from the subgroup analyses.

Table 1 Characteristics of included studies and current meta-analysis

Study	Year	Country	No. of centres	No. of patients	Patient age, min–max (mean), years	No. of masses	Malignant masses, no. (% of all masses)	Mass long axis, min–max (mean), mm	Reference standard
Yerli et al. [26]	2011	Turkey	1	71	32–87 (50)	78	16 (20.5)	4–49 (15.3)	Histo
Lee et al. [9]	2011	South Korea	1	278	23–81 (47)	315	48 (15.2)	2.6–10 (6.8)	Histo
Ciurea et al. [25]	2011	Romania	1	92	19–82 (48)	101	43 (42.6)	5.5–60 (16.3)	Histo / cyto
Wojcinski et al. [24]	2010	Germany	3	779	11–94 (54)	779	360 (46.2)	1–83 (16.3)	Histo
Zhu et al. [7]	2008	China	1	112	17–83 (45)	139	69 (49.6)	6–95 (22.2)	Histo
Current study	2011	Meta-analysis	7	1,332	11–94 (52)	1,412	536 (37.9)	1–95 (14.7)	Histo / cyto

Histo histopathology; *Cyto* cytopathology

Table 2 Diagnostic performance of USE and USB across different size ranges of breast masses (95 % confidence interval in parentheses)

	Overall (n=1,412[a]; 536 malignant)		Mass size <10 mm (n=543[b]; 121 malignant)		Mass size 10–19 mm (n=528[b]; 247 malignant)		Mass size >19 mm (n=325[b]; 162 malignant)		Mass size >29 mm (n=114; 58 malignant)	
	USE	USB	USE	USB	USE	USB	USE	USB	USE	USB
Sensitivity	78 (63–88)	96 (93–97)	76 (62–91)	95 (92–98)	82 (72–93)	95 (93–97)	74 (60–89)	97 (95–99)	75 (61–85)	98 (94–99)
Specificity	92 (85–96)	65 (42–82)	93 (88–98)	68 (48–89)	90 (84–96)	67 (47–87)	94 (89–98)	55 (32–79)	91 (81–95)	64 (35–86)
Positive LR	9.76 (5.48–17.37)	2.69 (1.51–4.83)	11.05 (4.80–17.29)	2.99 (1.10–4.86)	8.49 (3.81–13.18)	2.88 (1.13–4.64)	11.73 (4.97–18.49)	2.18 (1.05–3.31)	7.60 (4.25–13.59)	2.73 (1.24–5.98)
Negative LR	0.23 (0.13–0.41)	0.07 (0.04–0.11)	0.25 (0.11–0.40)	0.07 (0.04–0.11)	0.19 (0.08–0.30)	0.07 (0.04–0.11)	0.27 (0.12–0.41)	0.05 (0.03–0.08)	0.27 (0.17–0.43)	0.03 (0.007–0.09)

USE ultrasound elastography; *USB* conventional B-mode ultrasound; *CI* confidence interval; *LR* likelihood ratio

[a] Overall sensitivity and specificity were evaluated in all 1,412 masses

[b] Size was not reported for 16 masses; therefore the sum of the number of masses in the size subgroup analyses (<10 mm, 10–19 mm, >19 mm) equals 1,396

Mass size of less than 10 mm

There were 543 breast masses less than 10 mm; 422 (77.7 %) were benign and the remaining 121 (22.3 %) masses were malignant. The summary test operating measures for USE were: sensitivity of 76 % (95 % CI, 62 %–91 %) and specificity of 93 % (95 % CI, 88 %–98 %) (Table 2; Fig. 2). The positive and negative likelihood ratios (LR) for USE were 11.05 (95 % CI, 4.80–17.29) and 0.25 (95 % CI, 0.11–0.40), respectively.

The summary test operating measures for USB were: sensitivity of 95 % (95 % CI, 92 %–98 %) and specificity of 68 % (95 % CI, 48 %–89 %) (Table 2; Fig. 2). There was a statistically significant difference between the sensitivity (p=0.01) and specificity (p=0.02) of USB vs. USE. The positive and negative LRs were 2.99 (95 % CI, 1.10–4.86) and 0.07 (95 % CI, 0.04–0.11), respectively.

Mass size of 10–19 mm

There were 528 breast masses with sizes of 10–19 mm; 281 (53.2 %) were benign and the remaining 247 (46.8 %) masses were malignant. The summary test operating measures for USE were: sensitivity of 82 % (95 % CI, 72 %–93 %), specificity of 90 % (95 % CI, 84 %–96 %), positive LR of 8.49 (95 % CI, 3.81–13.18) and negative LR of 0.19 (95 % CI, 0.08–0.30) (Table 2; Fig. 2).

The summary test operating measures for USB were: sensitivity of 95 % (95 % CI, 93 %–97 %) and specificity of 67 % (95 % CI, 47 %–87 %) (Table 2; Fig. 2). There was a statistically significant difference between the sensitivity (P=0.02) and specificity (P=0.03) of USB vs. USE. The positive and negative LRs were 2.88 (95 % CI, 1.13–4.64) and 0.07 (95 % CI, 0.04–0.11), respectively.

Mass size of more than 19 mm

There were 325 breast masses with size more than 19 mm; 163 (50.1 %) were benign and the remaining 162 (49.9 %) masses were malignant. The summary test operating measures for USE were: sensitivity of 74 % (95 % CI, 60 %–89 %), specificity of 94 % (95 % CI, 89 %–98 %), positive LR of 11.73 (4.97–18.49) and negative LR of 0.27 (0.12–0.41) (Table 2; Fig. 2).

The summary test operating measures for USB were: sensitivity of 97 % (95 % CI, 95 %–99 %) and specificity of 55 % (95 % CI, 32 %–79 %) (Table 2; Fig. 2). There was a statistically significant difference between the sensitivity (P=0.003) and specificity (P=0.002) of USB vs. USE. The positive and negative LRs were 2.18 (95 % CI, 1.05–3.31) and 0.05 (95%CI, 0.03–0.08), respectively.

Test performance of USE and USB for 114 masses (8 % of all masses) with size more than 29 mm was additionally

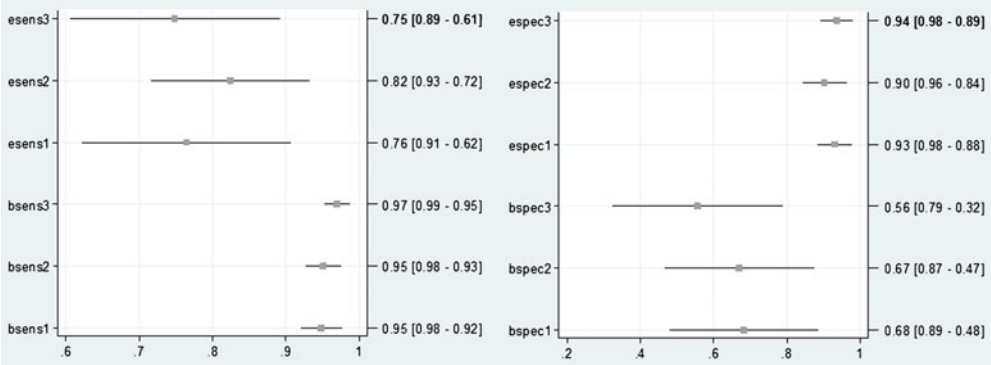

Fig. 2 Forest plot of sensitivity (*left*) and specificity (*right*) of ultrasound elastography and conventional B-mode ultrasound across three breast mass size groups. bsens1 = B-mode sensitivity for masses less than 10 mm; bsens2 = B-mode sensitivity for masses 10–19 mm of size; bsens3 = B-mode sensitivity for masses larger than 19 mm; esens1= elastography sensitivity for masses less than 10 mm; esens2 = elastography sensitivity for masses 10–19 mm of size; esens3 = elastography sensitivity for masses larger than 19 mm. bspec1 = B-mode specificity for masses less than 10 mm; bspec2 = B-mode specificity for masses 10–19 mm of size; bspec3 = B-mode specificity for masses larger than 19 mm; espec1 = elastography specificity for masses less than 10 mm; espec2 = elastography specificity for masses 10–19 mm of size; espec3 = elastography specificity for masses larger than 19 mm

assessed. Fifty-six (49.1 %) of these masses were benign, while the remaining 58 (50.9 %) masses were malignant. Seventy-four (65 %) of the masses were less than 40 mm. The summary test operating measures for USE were: sensitivity of 75 % (95 % CI, 61 %–85 %), specificity of 91 % (95 % CI, 81 %–95 %), positive LR of 7.60 (4.25–13.59) and negative LR of 0.27 (0.17–0.43) (Table 2). The summary test operating measures for USB were: sensitivity of 98 % (95 % CI, 94 %–99 %), specificity of 64 % (95 % CI, 35 %–86 %), positive LR of 2.73 (95 % CI, 1.24–5.98) and negative LR of 0.03 (95 % CI, 0.007–0.09) (Table 2).

Test performance across the three mass size groups

Comparing the sensitivity, specificity, and LRs of USE among three groups of masses with size of less than 10 mm, 10–19 mm, and more than 19 mm, we did not find any statistically significant differences ($P>0.05$). Figure 3 demonstrates the conditional probability graphs of USE and USB for three mass size groups. Our study results did not demonstrate any significant difference for USB performance (sensitivity, specificity, and LRs) across these three groups. Additionally, the results of subgroup analysis of 114 masses with size of more than 29 mm did not show any statistically significant differences between test performances of USE/USB compared to other mass sizes ($P>0.05$).

USE and USB test performances by mass size in each of the included studies are summarised in electronic Supplementary Table 1. Consistent with the results of our meta-analysis, the data from individual studies do not show any significant difference in USE and USB test performances by mass size.

Influences of potential confounders on sensitivity and specificity

Table 3 reports the influence of the potentially confounding variables, including the diagnostic test used for characterising the breast mass, mass size, and pathology, on the estimates of sensitivity and specificity. Test sensitivity significantly decreased when characterising benign pathologies (including papilloma, but not fibroadenoma) compared to fibroadenoma. Additionally, as expected, USB was significantly more sensitive and less specific than USE.

Discussion

The current comparative effectiveness study is an IPD meta-analysis of the diagnostic performance of USE and USB in 1,412 breast masses across three different breast mass size ranges (<10 mm, 10–19 mm, and >19 mm). IPD meta-analyses are considered the gold standard methodology for carrying out a meta-analysis compared to aggregated data meta-analysis because of their ability to model the within-study relationship between the patient-level covariates and diagnostic accuracy [10, 27]. Our study result demonstrated that the sensitivity and specificity of each of these techniques for characterising breast masses are not significantly different among masses with sizes of less than 10 mm, 10–19 mm, and more than 19 mm. Subgroup analysis of masses more than 29 mm demonstrated similar results. Comparison of USE versus USB within each of these size ranges demonstrates that USE has significantly higher specificity and lower sensitivity across all mass sizes.

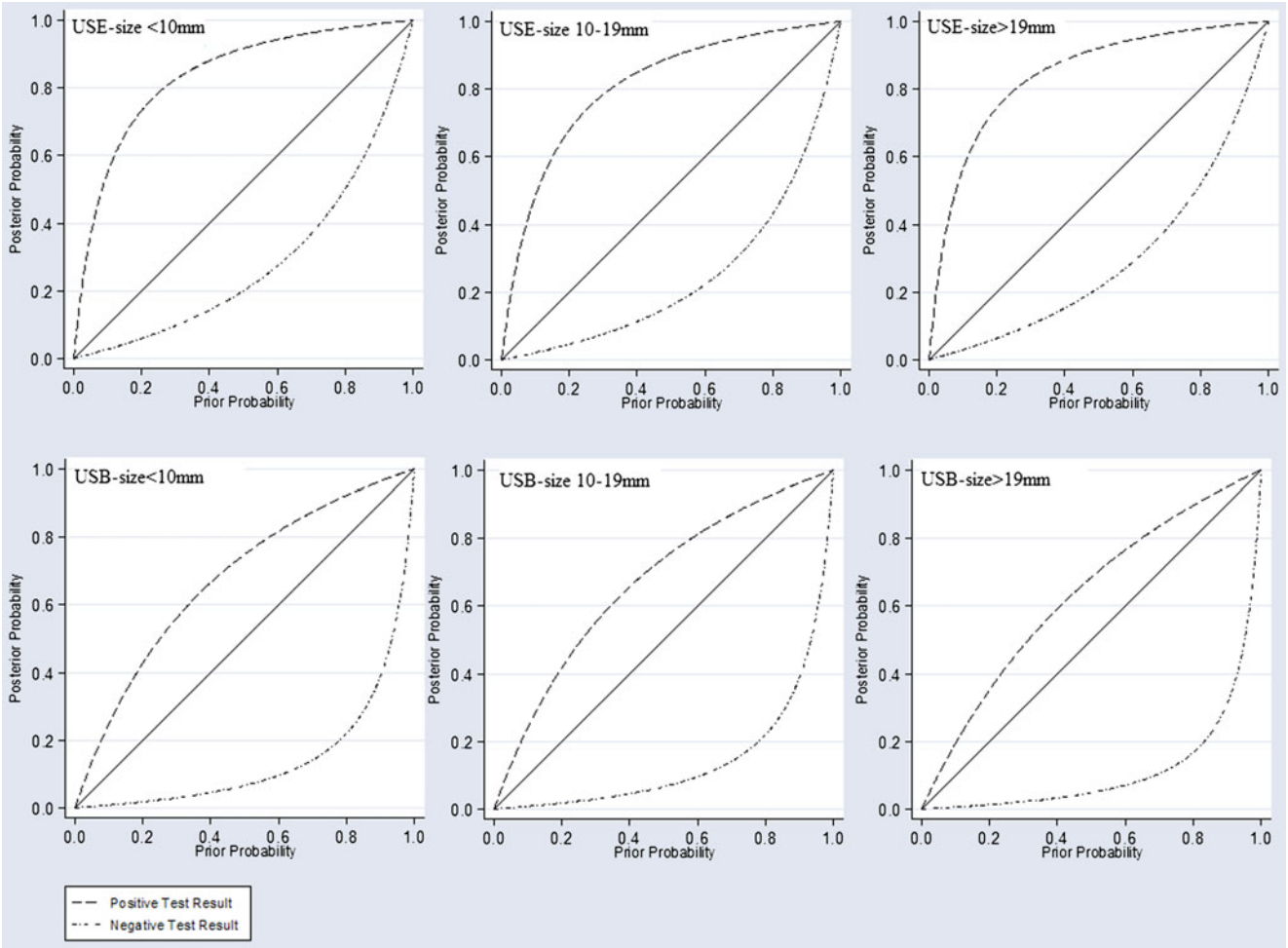

Fig. 3 Conditional probability graphs after a positive and negative test result for ultrasound elastography (USE) (*upper row*) and conventional B-mode ultrasound (USB) (*lower row*) in characterising breast masses with size of less than 10 mm (*left column*), 10–19 mm (*middle column*), and more than 19 mm (*right column*). The *horizontal axis* shows the pre-test probability of malignancy within a breast lesion, and the *vertical axis* shows the post-test probability of malignancy

Previous studies have reported controversial findings regarding the performance of USE in small masses. Giuseppetti et al. [28] and Regini et al. [29] demonstrated higher sensitivity and specificity for breast masses less than 20 mm compared to those larger than 20 mm. Inference of the confidence intervals from the above studies demonstrated no significant difference in test sensitivity between the two groups (Supplementary Table 2). Regini et al. [29] reported no statistically significant difference in sensitivity and specificity. Scaperrotta et al. and Zhi et al. [30, 31] reported higher sensitivity, lower specificity, and higher area under the curve of USE for masses less than 10 mm in size compared to masses 10–19 mm and masses >19 mm. However, regarding the inferred confidence intervals, neither study showed any significant difference for the diagnostic performance and area under the curve of USE among masses with different sizes (Supplementary Table 2). Zhu et al. [7] concluded that regardless of tumor size, USE is accurate in characterising breast masses. The current meta-analysis shows that although USE has the highest sensitivity (82 % vs. 76 % and 74 %) and lowest specificity (90 % vs. 93 % and 94 %) among breast masses 10–19 mm of size compared to masses <10 mm and >19 mm, the difference among the three groups is not significant.

In the current study, the summary estimates of sensitivity and specificity for USB do not significantly differ across the size ranges evaluated; however, USB has the highest sensitivity for masses larger than 19 mm compared to masses <10 mm and masses 10–19 mm (97 % vs. 95 % and 95 %) and the lowest specificity for the same size masses (55 % vs. 68 % and 67 %). This finding is not consistent with previous smaller individual reports on 256 and 88 breast masses that USB has lower sensitivity, specificity, and detection rates in the

Table 3 Estimates of ultrasound test performance according to the factors that influenced sensitivity and specificity

Variable	Sensitivity (95 % CI)	Specificity (95 % CI)
Diagnostic test		
USE (reference)	15 (5–16)	99 (99–100)
USB	62 (45–78)*	97 (95–99)*
Mass size		
<10 mm (reference)	81 (71–90)	85 (78–92)
10–19 mm	78 (67–88)	87 (81–93)
>19 mm	78 (37–89)	87 (81–93)
Mass pathology		
Fibroadenoma (reference)	85 (76–95)	85 (57–100)
Papilloma	54 (26–82)*	97 (89–100)
Other benign pathologies	76 (63–92)*	91 (74–100)
DCIS	92 (73–100)	75 (58–93)
IDC	77 (35–98)	91 (87–96)
ILC	76 (30–100)	92 (85–98)
Other malignant pathologies	89 (65–100)	81 (69–93)

CI confidence interval; *USE* ultrasound elastography; *USB* conventional B-mode ultrasound; *DICS* ductal carcinoma in situ; *IDC* invasive ductal carcinoma; *ILC* invasive lobular carcinoma

*P values <0.05

characterisation of small masses [32, 33] because of the lower frequency of characterising sonographic findings such as hypoechogenicity or acoustic transmission among small masses [32]. Our present meta-analysis includes a larger number of breast masses and accounts for the inter-patient and inter-study heterogeneity in the determination of test performance.

Comparing USE with USB within each size group of masses, the findings of the present study show that USE has significantly higher specificity and lower sensitivity for the characterisation of breast masses regardless of their size. This result is consistent with Lee et al.'s [9] report for small breast masses less than 1 cm. However, it is not consistent with Fu et al.'s study, which demonstrated USE has higher sensitivity for characterising small masses (<1 cm) when compared to USB (89 % vs. 67 %) [8].

Potential clinical utility of USE

Considering LRs for the two techniques and the conditional probability graphs (Fig. 3), USB with negative LRs of less than 0.1 is capable of almost excluding malignancy within a breast mass with a high probability regardless of the size. Therefore, a negative USB usually does not require any further diagnostic tests. On the other hand, USE is capable of confirming malignancy with a high probability within a breast mass, regardless of size. Therefore, biopsy of a positive USE will most probably reveal malignancy. Further, the profiles from the conditional probability graphs suggest that both B-mode and elastography should be used in conjunction to improve the overall specificity of US as a diagnostic technology.

Limitations

Completeness of acquired patient-level data from each of the individual studies was one of the limitations of this study. Many of the corresponding authors of individual studies either did not respond or did not send their source data. Therefore, our study was limited to data from five studies (7 centres, 5 countries), which may limit the power of the analysis or the generalisability of the findings. Our data set was limited to studies that used a 5-score elasticity score as the method for measuring strain. Patients included in the current study were either referred because of specific diagnostic queries or intensified screening in high-risk populations. Differentiation of benign from malignant masses based on USE in low-risk populations may require further investigations. Additionally, there was a high level of heterogeneity in the current study. This may be due to patient-level differences not captured in the data sets and technical dissimilarities when conducting the reported studies. For example, mass pathology was shown to be a potential factor influencing test sensitivity and specificity. None of the included five studies were conducted in United States; therefore, there might be practice variability between these studies and the studies conducted in the United States.

Implications

In summary, our study suggests that for differentiating between benign and malignant breast masses of any size, it is better to start investigation with conventional breast B-mode ultrasound and augment this with elastography as clinically needed. Considering the USB low negative likelihood ratio if the test result is negative, usually no further diagnostic test is required and the malignancy should be ruled out with a high probability. Lesions with high suspicion for malignancy (BI-RADS 4 or 5) should directly proceed to tissue sampling. If B-mode ultrasound findings are equivocal, such as a probably benign (BI-RADS 3) mass, augmenting the B-mode examination with elastography, with its higher specificity and positive likelihood ratio, is recommended.

Acknowledgment The authors would like to thank Dr. Anca Ileana Ciurea, Radiologist at the University of Medicine and Pharmacy Cluj-Napoca, Romania; Dr. Sung Hun Kim, Radiologist at the Catholic University of Korea, South Korea; Dr. Hasan Yerli, Radiologist at the Ankam University, Turkey; and Dr. Qing-Li Zhu, Radiologist at the Peking Union Medical College, China, for providing us with the patient-level source data from their published studies.

Ruth Carlos is a member of physician's advisory board of Philips. Gelareh Sadigh, Colleen Neal, Sebastian Wojcinski, and Ben Dwamena have no conflict of interest to declare.

References

1. Garra BS (2007) Imaging and estimation of tissue elasticity by ultrasound. Ultrasound Q 23:255–268
2. Wells PN, Liang HD (2011) Medical ultrasound: imaging of soft tissue strain and elasticity. J R Soc Interface 8:1521–1549
3. Sadigh G, Carlos RC, Neal CH, Dwamena BA (2012) Accuracy of quantitative ultrasound elastography for differentiation of malignant and benign breast abnormalities: a meta-analysis. Breast Cancer Res Treat 134:923–931
4. Itoh A, Ueno E, Tohno E et al (2006) Breast disease: clinical application of US elastography for diagnosis. Radiology 239:341–350
5. Sadigh G, Carlos RC, Neal CH, Dwamena BA (2012) Ultrasonographic differentiation of malignant from benign breast lesions: a meta-analytic comparison of elasticity and BIRADS scoring. Breast Cancer Res Treat 133:23–35
6. Gong X, Xu Q, Xu Z, Xiong P, Yan W, Chen Y (2011) Real-time elastography for the differentiation of benign and malignant breast lesions: a meta-analysis. Breast Cancer Res Treat 130:11–8
7. Zhu QL, Jiang YX, Liu JB et al (2008) Real-time ultrasound elastography: its potential role in assessment of breast lesions. Ultrasound Med Biol 34:1232–1238
8. Fu LN, Wang Y, Huang YH (2011) Value of ultrasound elastography in detecting small breast tumors. Chin Med J (Engl) 124:2384–2386
9. Lee JH, Kim SH, Kang BJ et al (2011) Role and clinical usefulness of elastography in small breast masses. Acad Radiol 18:74–80
10. Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR (2008) Meta-analysis of diagnostic test studies using individual patient data and aggregate data. Stat Med 27:6111–6136
11. Haines TP, Hill AM (2011) Inconsistent results in meta-analyses for the prevention of falls are found between study-level data and patient-level data. J Clin Epidemiol 64:154–162
12. Tugwell P, Knottnerus JA (2010) Advantages of individual patient data analysis in systematic reviews. J Clin Epidemiol 63:233–234
13. American College of Radiology (2009) Breast imaging reporting and data system (BI-RADS®) ultrasound. http://www.acr.org/SecondaryMainMenuCategories/quality_safety/BIRADSAtlas.aspx. Accessed October 12, 2010
14. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 151:264–269, W64
15. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 3:25
16. Chu H, Cole SR (2006) Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. J Clin Epidemiol 59:1331–1332, author reply 1332–1333
17. Arends LR, Hamza TH, van Houwelingen JC, Heijenbrok-Kal MH, Hunink MG, Stijnen T (2008) Bivariate random effects meta-analysis of ROC curves. Med Decis Mak 28:621–638
18. Rabe-Hesketh S, Skrondal A, Pickles A (2005) Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. J Econ 128:301–323
19. Rabe-Hesketh S, Skrondal A, Pickles A (2002) Reliable estimation of generalized linear mixed models using adaptive quadrature. Stata J 2:1–21
20. Cronin P, Dwamena BA, Kelly AM, Bernstein SJ, Carlos RC (2008) Solitary pulmonary nodules and masses: a meta-analysis of the diagnostic utility of alternative imaging tests. Eur Radiol 18:1840–1856
21. Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. BMJ 327:557–560
22. Fleury Ede F, Fleury JC, Oliveira VM, Rinaldi JF, Piato S, Roveda Junior D (2009) Proposal for the systematization of the elastographic study of mammary lesions through ultrasound scan. Rev Assoc Med Bras 55:192–196
23. Fleury Ede F, Fleury JC, Piato S, Roveda D Jr (2009) New elastographic classification of breast lesions during and after compression. Diagn Interv Radiol 15:96–103
24. Wojcinski S, Farrokh A, Weber S et al (2010) Multicenter study of ultrasound real-time tissue elastography in 779 cases for the assessment of breast lesions: improved diagnostic performance by combining the BI-RADS(R)-US classification system with sonoelastography. Ultraschall Med 31:484–491
25. Ciurea AI, Bolboaca SD, Ciortea CA, Botar-Jid C, Dudea SM (2011) The influence of technical factors on sonoelastographic assessment of solid breast nodules. Ultraschall Med 32:S27–S34
26. Yerli H, Yilmaz T, Kaskati T, Gulay H (2011) Qualitative and semiquantitative evaluations of solid breast lesions by sonoelastography. J Ultrasound Med 30:179–186
27. Lambert PC, Sutton AJ, Abrams KR, Jones DR (2002) A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. J Clin Epidemiol 55:86–94
28. Giuseppetti GM, Martegani A, Di Cioccio B, Baldassarre S (2005) Elastosonography in the diagnosis of the nodular breast lesions: preliminary report. Radiol Med 110:69–76
29. Regini E, Bagnera S, Tota D et al (2010) Role of sonoelastography in characterising breast nodules. Preliminary experience with 120 lesions. Radiol Med 115:551–562
30. Scaperrotta G, Ferranti C, Costa C et al (2008) Role of sonoelastography in non-palpable breast lesions. Eur Radiol 18:2381–2389
31. Zhi H, Xiao XY, Yang HY, Ou B, Wen YL, Luo BM (2010) Ultrasonic elastography in breast cancer diagnosis: strain ratio vs 5-point scale. Acad Radiol 17:1227–1233
32. Del Frate C, Bestagno A, Cerniato R et al (2006) Sonographic criteria for differentiation of benign and malignant solid breast lesions: size is of value. Radiol Med 111:783–796
33. Berg WA, Blume JD, Cormack JB, Mendelson EB (2006) Operator dependence of physician-performed whole-breast US: lesion detection and characterization. Radiology 241:355–365