Original Article | Breast Imaging

eISSN 2005-8330
https://doi.org/10.3348/kjr.2020.1227
Korean J Radiol 2021;22(8):1240-1252

Korean Journal of Radiology
KJR

Accuracy of Digital Breast Tomosynthesis for Detecting Breast Cancer in the Diagnostic Setting: A Systematic Review and Meta-Analysis

Min Jung Ko[1]*, Dong A Park[1]*, Sung Hyun Kim[1], Eun Sook Ko[2], Kyung Hwan Shin[3], Woosung Lim[4], Beom Seok Kwak[5], Jung Min Chang[6, 7]

[1]Division for Healthcare Technology Assessment Research, National Evidence-Based Healthcare Collaborating Agency, Seoul, Korea; [2]Department of Radiology, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea; Departments of [3]Radiation Oncology and [7]Radiology, Seoul National University College of Medicine, Seoul, Korea; [4]Department of Surgery, College of Medicine, Ewha Womans University, Seoul, Korea; [5]Department of Surgery, Dongguk University Ilsan Hospital, Goyang, Korea; [6]Department of Radiology, Seoul National University Hospital, Seoul, Korea

Objective: To compare the accuracy for detecting breast cancer in the diagnostic setting between the use of digital breast tomosynthesis (DBT), defined as DBT alone or combined DBT and digital mammography (DM), and the use of DM alone through a systematic review and meta-analysis.

Materials and Methods: Ovid-MEDLINE, Ovid-Embase, Cochrane Library and five Korean local databases were searched for articles published until March 25, 2020. We selected studies that reported diagnostic accuracy in women who were recalled after screening or symptomatic. Study quality was assessed using the Quality Assessment of Diagnostic Accuracy Studies-2 tool. A bivariate random effects model was used to estimate pooled sensitivity and specificity. We compared the diagnostic accuracy between DBT and DM alone using meta-regression and subgroup analyses by modality of intervention, country, existence of calcifications, breast density, Breast Imaging Reporting and Data System category threshold, study design, protocol for participant sampling, sample size, reason for diagnostic examination, and number of readers who interpreted the studies.

Results: Twenty studies (n = 44513) that compared DBT and DM alone were included. The pooled sensitivity and specificity were 0.90 (95% confidence interval [CI] 0.86–0.93) and 0.90 (95% CI 0.84–0.94), respectively, for DBT, which were higher than 0.76 (95% CI 0.68–0.83) and 0.83 (95% CI 0.73–0.89), respectively, for DM alone ($p < 0.001$). The area under the summary receiver operating characteristics curve was 0.95 (95% CI 0.93–0.97) for DBT and 0.86 (95% CI 0.82–0.88) for DM alone. The higher sensitivity and specificity of DBT than DM alone were consistently noted in most subgroup and meta-regression analyses.

Conclusion: Use of DBT was more accurate than DM alone for the diagnosis of breast cancer. Women with clinical symptoms or abnormal screening findings could be more effectively evaluated for breast cancer using DBT, which has a superior diagnostic performance compared to DM alone.

Keywords: *Breast cancer; Mammography; Breast tomosynthesis; Meta-analysis; Performance*

INTRODUCTION

Diagnostic mammography is widely used to assess potential abnormalities detected in screening mammography, to further evaluate patients who have signs or symptoms of breast disease, and for short-term follow-

Received: October 10, 2020 **Revised:** January 11, 2021 **Accepted:** January 17, 2021
This study was supported by National Evidence-based Healthcare Collaborating Agency (NECA) of Korea (NR-19-001-25).
*These authors contributed equally to this work.
Corresponding author: Jung Min Chang, MD, PhD, Department of Radiology, Seoul National University Hospital, 101 Daehak-ro, Jongno-gu, Seoul 03080, Korea.
• E-mail: imchangjm@gmail.com

Korean Journal of Radiology
KJR

up of patients with probable benign findings [1]. However, mammography has a limitation in visualizing overlapping dense fibroglandular breast tissue, which can ultimately reduce the conspicuity of breast cancers and make normal structures appear suspicious [2,3]. To solve the problem of inconclusive findings frequently noted in women with dense breasts, supplemental mammographic views are conventionally obtained with or without targeted breast ultrasonography [4,5].

Digital breast tomosynthesis (DBT) is currently used in clinical practice; compared with digital mammography (DM), it has increased the rate of breast cancer detection and decreased false-positive findings by reducing overlapping breast tissue. DBT allows for quasi-three-dimensional breast reconstruction, which permits the tissue visualization in these sections with subsequent resolution of overlying tissue. Consequently, key mammographic findings, such as masses, areas of architectural distortion, and asymmetries, are better discerned and characterized with greater confidence compared to conventional two-dimensional (2D) mammographic images [6,7]. A recent meta-analysis that evaluated studies comparing DBT and 2D mammography in a screening setting showed that the pooled incremental cancer detection rate for tomosynthesis was 1.6 cancers per 1000 screens [8]. Accumulating evidence from big-data analysis studies in the screening setting have also suggested the superiority of DBT in comparison with the current standard or synthetic DM [9,10].

DBT can be efficaciously used, not only in the screening setting, but also in various clinical diagnostic settings [11-13]. Studies have reported the diagnostic value of DBT to detect breast cancer in symptomatic patients or patients with suspicious mammographic findings; however, to our knowledge, a limited number of meta-analyses has compared the diagnostic accuracies of DBT and DM in a diagnostic setting across a wide range of disease presentations.

Therefore, this study was aimed at comparing the accuracy for detecting breast cancer in the diagnostic setting between DBT, defined as DBT alone or combined DBT and DM, and DM alone through a systematic review and meta-analysis. Our hypothesis was that DBT might offer superior diagnostic accuracy compared to DM alone.

MATERIALS AND METHODS

Search Strategy and Study Selection

This study followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses-Diagnostic Test Accuracy Statement [14]. We searched Ovid-MEDLINE, Ovid-Embase, Cochrane Library, and five Korean local databases for articles published until March 25, 2020. Supplementary Materials show the detailed search strategy and processes of the systematic review and meta-analysis. In addition, the bibliographies of relevant articles were also reviewed to identify additional publications.

Inclusion and Exclusion Criteria

The selection criteria were as follows: 1) breast imaging using both DM and DBT for breast symptoms, abnormal findings on screening examinations, or a need for follow-up imaging; 2) DBT (DBT alone or combined DBT and DM) compared to DM alone; 3) sufficient information to discriminate among true-positive, false-positive, false-negative, and true-negative results for breast cancer to determine the sensitivity and specificity; 4) reference standards of biopsy and/or imaging follow-up. The exclusion criteria were as follows: non-original articles; screening setting or participants solely composed of patients with breast cancer; single arm; non-comparative study; or fewer than 50 participants.

Data Extraction and Quality Assessment

We performed pilot data extraction for several studies to standardize the data extraction form and improve consistency between the reviewers. The two reviewers who conducted the study selection independently extracted data from the selected studies into a standardized form, including 1) study characteristics: authors, year of publication, study design, study period, and setting (number and location of research center); 2) study population: inclusion/exclusion criteria, number, mean age (range), breast density, and sampling method; 3) methods: index test, reference test (biopsy or follow-up imaging), comparator, and threshold of Breast Imaging Reporting and Data System (BI-RADS) final assessment category; and 4) analyzed diagnostic accuracy: 2-by-2 contingency table for the presence of breast cancer by index test and comparator (number of true-positive, false-positive, false-negative, and true-negative results), by directly extracting the existing data or estimating the values using the existing data. Any disagreement between the two reviewers was resolved by rechecking the data and discussing the case further with the clinical advisory committee (Supplemental Materials).

The risk of bias and the applicability of the included

 https://doi.org/10.3348/kjr.2020.1227

studies were assessed by the two reviewers using the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool, and all discrepancies were resolved by discussion with the clinical advisory committee.

Statistical Analysis

The pooled summary of sensitivity and specificity were estimated based on the bivariate random effects meta-analysis using STATA version 14.2 (StataCorp.). We generated a hierarchical summary receiver operating characteristic (HSROC) type curve, a summary ROCs curve, and the respective area under the curve. To evaluate study heterogeneity, which refers to the variability in diagnostic accuracy across the primary studies, we used the Higgins I^2 statistic, with $I^2 > 50\%$ indicating the presence of heterogeneity [15]. When there was a substantial heterogeneity in diagnostic accuracy across studies, we investigated a threshold effect by 1) visual assessment of coupled forest plots of sensitivity and specificity, and 2) a Spearman correlation coefficient between the sensitivity and false-positive rate (correlation coefficient > 0.6 indicated a threshold effect) [16]. We also visually assessed the differences between the 95% confidence region and the 95% prediction region in the HSROC curve for examining the presence of heterogeneity between studies [17]. We performed meta-regression analysis to further explore the causes of study heterogeneity by including predefined covariates in a bivariate model: modality of intervention (DBT alone vs. DBT with DM), country where the study was performed (Asian vs. non-Asian), existence of calcifications (lesion with calcification vs. lesion without calcifications), breast density of women included in the study (≥ 2 or b vs. 1–4 or a–d), BI-RADS category threshold (≥ 4 vs. ≥ 3), study design (prospective vs. retrospective), protocol for participant sampling (consecutive vs. non-consecutive), sample size (< 200 vs. ≥ 200), reason for diagnostic examination (symptoms only vs. symptoms or screening recall), and number of readers who interpreted the studies (≥ 5 vs. < 5).

We compared the diagnostic accuracies of DBT and DM alone by adding a covariate for test type to the HSROC model and performed the likelihood ratio tests to compare models with and without covariate terms. We further conducted subgroup comparative analyses using meta-regression between DBT and DM alone according to covariate terms. In addition, we conducted a sensitivity analysis to assess the robustness of our results. By excluding influential research one-by-one, the overall effects of sensitivity or specificity change were estimated. For the non-selected predefined covariates for the subgroup analysis, the following exclusion criteria were chosen considering the characteristics of the included studies: studies with number of participants > 1000 or those comparing DBT with DM alone using unpaired patient cohort. We added 0.5 to all cells of a study's 2-by-2 contingency table if any cell was zero. Deeks' funnel plot was generated to test for publication bias, with statistical significance being assessed based on Deeks' asymmetry test. We used two-tailed tests of significance, and $p < 0.05$ was considered statistically significant.

RESULTS

Study Characteristics

A total of 2501 articles were identified, and 139 full texts were retrieved. Of these, 119 were excluded according to the exclusion criteria (Fig. 1). Finally, 20 articles with 44513 patients and 58388 lesions fulfilled the eligibility criteria [18-37]. Table 1 presents the characteristics of the included studies. They were published between 2013 and 2020. The mean age of the participants was 52 years (range, 45–62 years).

Of these studies, 50% were conducted in Asian countries [18-20,23,25,27,29,31-33], and there were four prospective studies [18,19,32,33]. Five studies had patients with breast density ≥ 2 or b according to American College of Radiology (ACR) [28-31,34], nine studies consecutively sampled participants [18,21,22,24-26,32,34,35], and 13 studies performed a per-lesion analysis [18-23,25,26,28,29,31-34]. Five studies included lesions without calcifications [26,28-30,34], and three studies symptomatic patients only [19,29,34]. A total of 11 studies compared DBT alone and DM alone [18,20,22-24,26,28,29,31,33,36], while nine studies [19,21,25,27,30,32,34,35,37] compared the combination of DBT with DM and DM alone.

Quality Assessment

The results of quality assessments using QUADAS-2 tool are summarized in Figure 2. With regard to patient selection, nine studies had a low risk for bias [18,21,22,24-26,32,34,35], while one study had a high risk for bias [23], mainly due to limited reporting of participant sampling. Only one study had a high risk of bias with regard to index tests [28] as it did not explicitly present a threshold. In

https://doi.org/10.3348/kjr.2020.1227

Korean Journal of Radiology
KJR

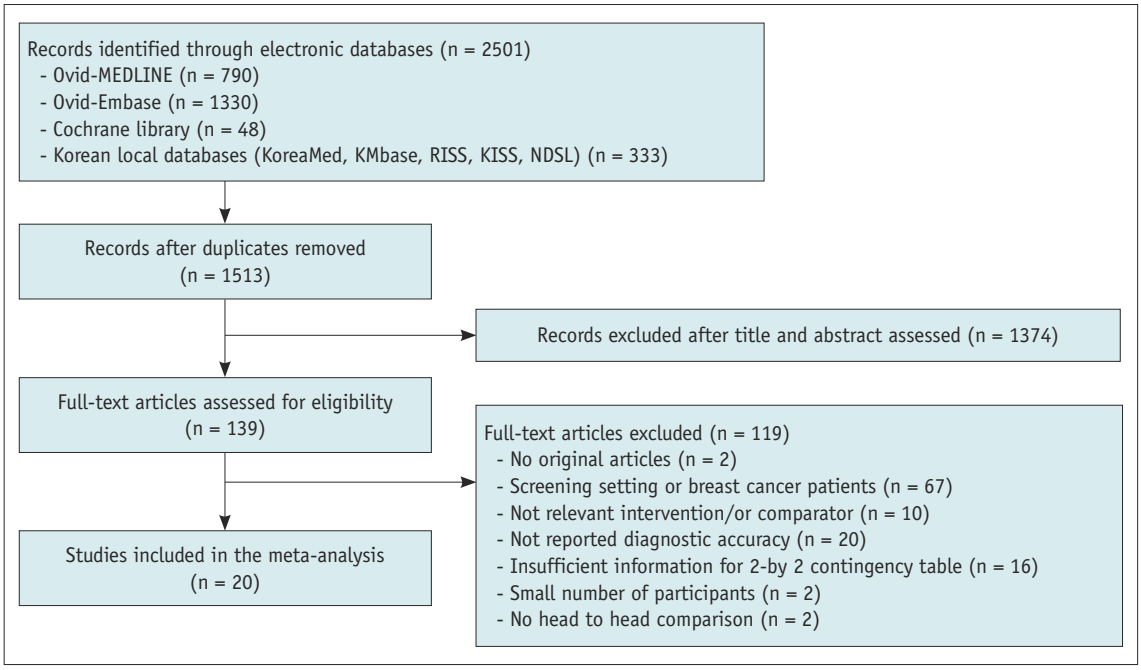

Fig. 1. Flow diagram of the study selection process.

the domain of bias in the reference standard, 12 studies were scored as 'unclear risk' [20-25,27-30,34,37] because of unclear explanation of the blinding results of index test. All studies were scored 'unclear' in the domains of bias in the patient flow and timing, as they did not clearly define the appropriate interval between index test and reference standard. With regard to applicability, all included studies were scored 'low' in the three domains.

Diagnostic Accuracy of DBT and DM

The bivariate random-effects meta-analysis of the 20 studies showed a pooled sensitivity of 0.90 (95% confidence interval [CI] 0.86–0.93) and a pooled specificity of 0.90 (95% CI 0.84–0.94) for DBT (Fig. 3). With regard to DM alone, the pooled estimates of sensitivity and specificity were 0.76 (95% CI 0.68–0.83) and 0.83 (95% CI 0.73–0.89), respectively. In comparison with DM alone, DBT showed a higher sensitivity and specificity ($p < 0.001$). The area under summary ROC curve for DBT and DM alone was 0.95 (95% CI 0.93–0.97) and 0.86 (95% CI 0.82–0.88), respectively (Supplementary Figs. 1, 2). The Higgins I^2 statistic indicated the presence of heterogeneity in both sensitivity (I^2 = 91.4%) and specificity (I^2 = 99.3%) among the studies reporting DBT. In addition, there was a large difference between the 95% confidence and prediction regions in the HSROC curve, which indicated the presence of heterogeneity among studies (Fig. 4A). However, the estimation of the Spearman correlation coefficient (ρ = -0.16, p = 0.49) as well as the visual inspection of coupled forest plot (Fig. 3A) did not reveal a diagnostic threshold effect. Substantial heterogeneity in both pooled sensitivities and pooled specificities were found for DBT and DM alone for the non-threshold effect.

To further explore the causes of study heterogeneity, a meta-regression analysis was conducted (Supplementary Table 1). The results showed that the BI-RADS category threshold was the only significant factor that influenced study heterogeneity with regard to sensitivity and specificity. However, despite extensive meta-regression, no other significant variable was found to influence study heterogeneity. Deeks' funnel plot and the results of regression test for asymmetry of the included studies indicated no direct evidence for publication bias (p = 0.24) (Fig. 5). In our sensitivity analyses, after excluding studies with the number of patients exceeding 1000 or those comparing DBT with DM alone using unpaired patient cohorts (Supplementary Table 2), the pooled sensitivity and specificity of DBT were significantly higher than those of DM alone.

Subgroup Comparative Analyses

We divided the participants into subgroups according to the modality of intervention, country, existence of calcification, breast density, BI-RADS category threshold,

 https://doi.org/10.3348/kjr.2020.1227

Table 1. Characteristics of the Included Studies

Study (Year)	Country	Study Design	Participants					Index Test, Comparator	BI-RADS Threshold	Reference Standard (Biopsy or Follow-Up Imaging)
			No. (Lesions)*	Mean Age	Breast Density†	Con-Secutive Sampling	Per Analysis			
Basha et al. [18] (2020)	Egypt	Prospective	296 (355)	46	All	Yes	Lesion	DBT, DM	4	Mixed
You et al. [19] (2020)	China	Prospective	212 (222)	NR	All	No	Lesion	DBT + DM, DM	4	Biopsy
Asbeutah et al. [20] (2019)	Kuwait	Retrospective	58 (65)	48	All	No	Lesion	DBT, DM	4	Biopsy
Bahl et al. [21]‡ (2019)	USA	Retrospective	32704 (45707)	53.2	All	Yes	Lesion	DBT + DM, DM	4	Mixed
Georgian-Smith et al. [22] (2019)	USA	Retrospective	330 (548)	56	All	Yes	Lesion	DBT, DM	4	Biopsy
Li et al. [23] (2019)	China	Retrospective	305 (312)	49	All	No	Lesion	DBT, DM	4	Biopsy
Mall et al. [24] (2018)	Australia	Retrospective	144	NR	All	Yes	Patient	DBT, DM	4	Mixed
Ohashi et al. [25] (2018)	Japan	Retrospective	628 (1164)	50	All	Yes	Lesion	DBT + DM, DM	4	Mixed
Chan et al. [26] (2017)	USA	Retrospective	134 (142)	46	All	Yes	Lesion	DBT, DM	4	Mixed
Kim et al. [27] (2017)	Korea	Retrospective	116	52	All	No	Patient	DBT + DM, DM	4	Mixed
Krammer et al. [28] (2017)	Germany	Retrospective	66 (69)	62	≥ 3	No	Lesion	DBT, DM	NR	Mixed
Bian et al. [29] (2016)	China	Retrospective	631	45	≥ 3	No	Patient	DBT, DM	4	Biopsy
Elizalde et al. [30] (2016)	Spain	Retrospective	1042	52	≥ 2	No	Patient	DBT + DM, DM	3	Mixed
Kamal et al. [31] (2016)	Egypt	Retrospective	98 (103)	NR	≥ c	No	Lesion	DBT, DM	4	Mixed
Seo et al. [32] (2016)	Korea	Prospective	203 (206)	50	All	Yes	Lesion	DBT + DM, DM	4	Mixed
Taha Ali et al. [33] (2016)	Egypt	Prospective	132 (145)	59	All	No	Lesion	DBT, DM	3	Mixed
Bansal et al. [34] (2015)	UK	Retrospective	103 (106)	53	≥ b	Yes	Lesion	DBT + DM, DM	4	Mixed
Gilbert et al. [35] (2015)	UK	Retrospective	7060	56	All	Yes	Patient	DBT + DM, DM	NR	Mixed
Tagliafico et al. [36] (2015)	Italy	Retrospective	107	52	All	No	Patient	DBT, DM	3	Biopsy
Waldherr et al. [37] (2013)	Switzerland	Retrospective	144	NR	All	No	Patient	DBT + DM, DM	4	Biopsy

*Parenthesis means number of lesions when per lesion analysis was conducted, †Breast density patterns were described in either way of using composition 1, 2, 3, 4 or a, b, c, d, ‡Number of participants (lesions) were 16881 (22824) for combined DBT with DM and 15823 (22883) for DM, respectively. BI-RADS = Breast Imaging Reporting and Data System, DBT = digital breast tomosynthesis, DM = digital mammography, NR = not reported

https://doi.org/10.3348/kjr.2020.1227

Korean Journal of Radiology
KJR

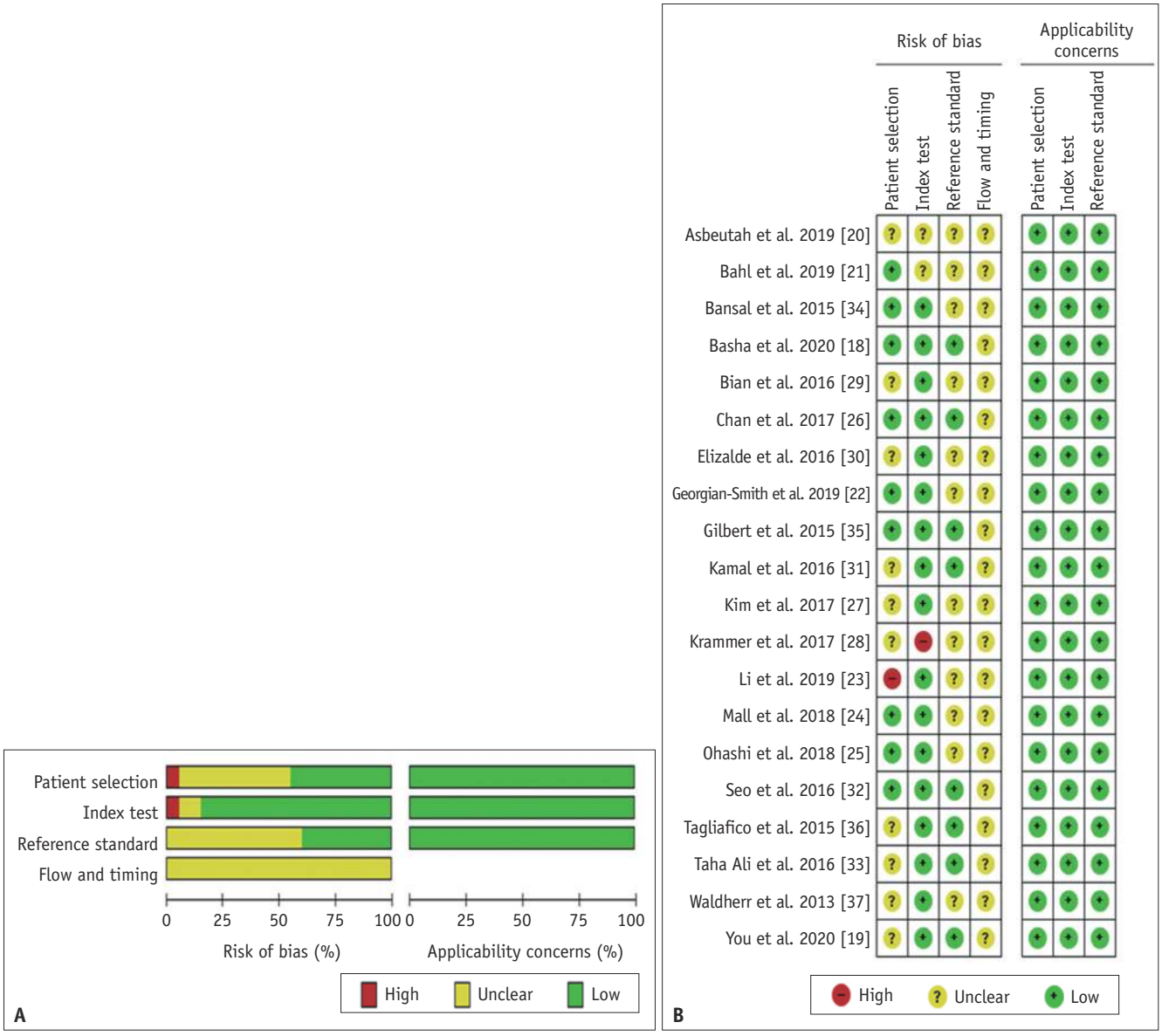

Fig. 2. Risk of bias graph by the Quality Assessment of Diagnostic Accuracy Studies version 2, (A) risk of bias graph, and (B) risk of bias summary for each study.

study design, protocol for participant sampling, number of participants., reason for diagnostic examination (symptoms only vs. symptoms or screening recall), and number of readers who interpreted the studies (≥ 5 vs. < 5). Results of the subgroup analysis revealed that pooled sensitivity and/or specificity for DBT was consistently higher than those of DM alone (Table 2); it was statistically significant in all analyses (all $p < 0.05$), except for the subgroups of non-Asian countries, no calcifications, a BI-RADS category threshold ≥ 3, breast density ≥ 2 or b, and number of readers who interpreted the studies ≥ 5. The higher sensitivities of DBT compared with those of DM alone were preserved in most analyses, except those for non-Asian, BI-RADS category threshold ≥ 3, breast density ≥ 2 or b, no calcification, diagnostic exam for symptomatic reason, and number of readers who interpreted the studies ≥ 5. Compared with DM alone, we found higher specificity of DBT in the subgroups DBT alone, Asian, and non-consecutive sampling ($p < 0.05$).

DISCUSSION

In our systematic review and meta-analysis of 20 comparative studies that investigated 44513 patients, the

https://doi.org/10.3348/kjr.2020.1227

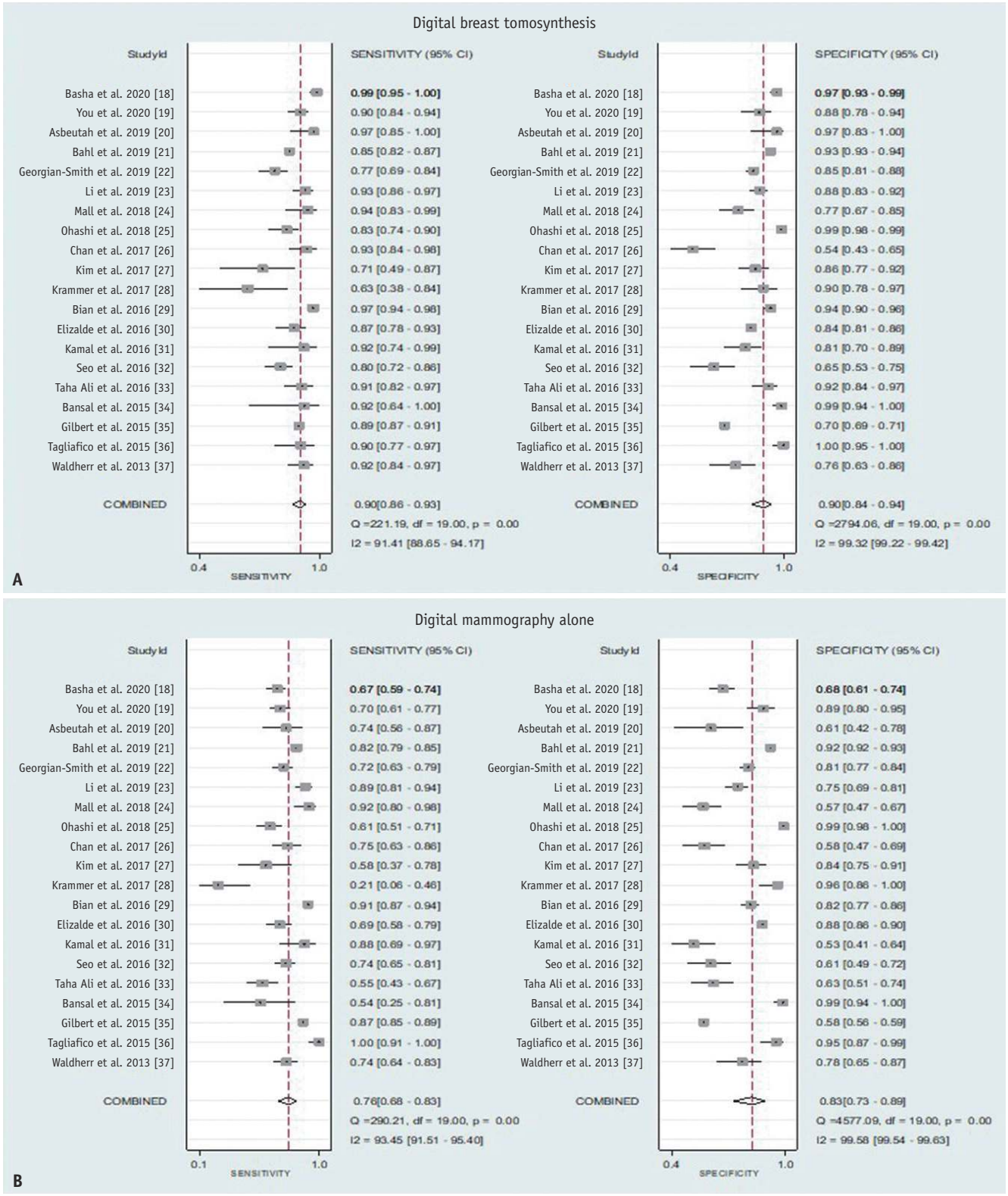

Fig. 3. Coupled forest plots of the pooled sensitivity and specificity of (A) digital breast tomosynthesis and (B) digital mammography alone for breast cancer diagnosis. The black square boxes denote either sensitivity (left panel) or specificity (right panel), and horizontal lines represent 95% CI for each study. The vertical dotted line indicates pooled summary estimates of sensitivity or specificity, and the diamond at the bottom indicates the 95% CIs. Heterogeneity statistics (I^2 value, Q value) for sensitivity and specificity are displayed.
CI = confidence interval

https://doi.org/10.3348/kjr.2020.1227

Korean Journal of Radiology
KJR

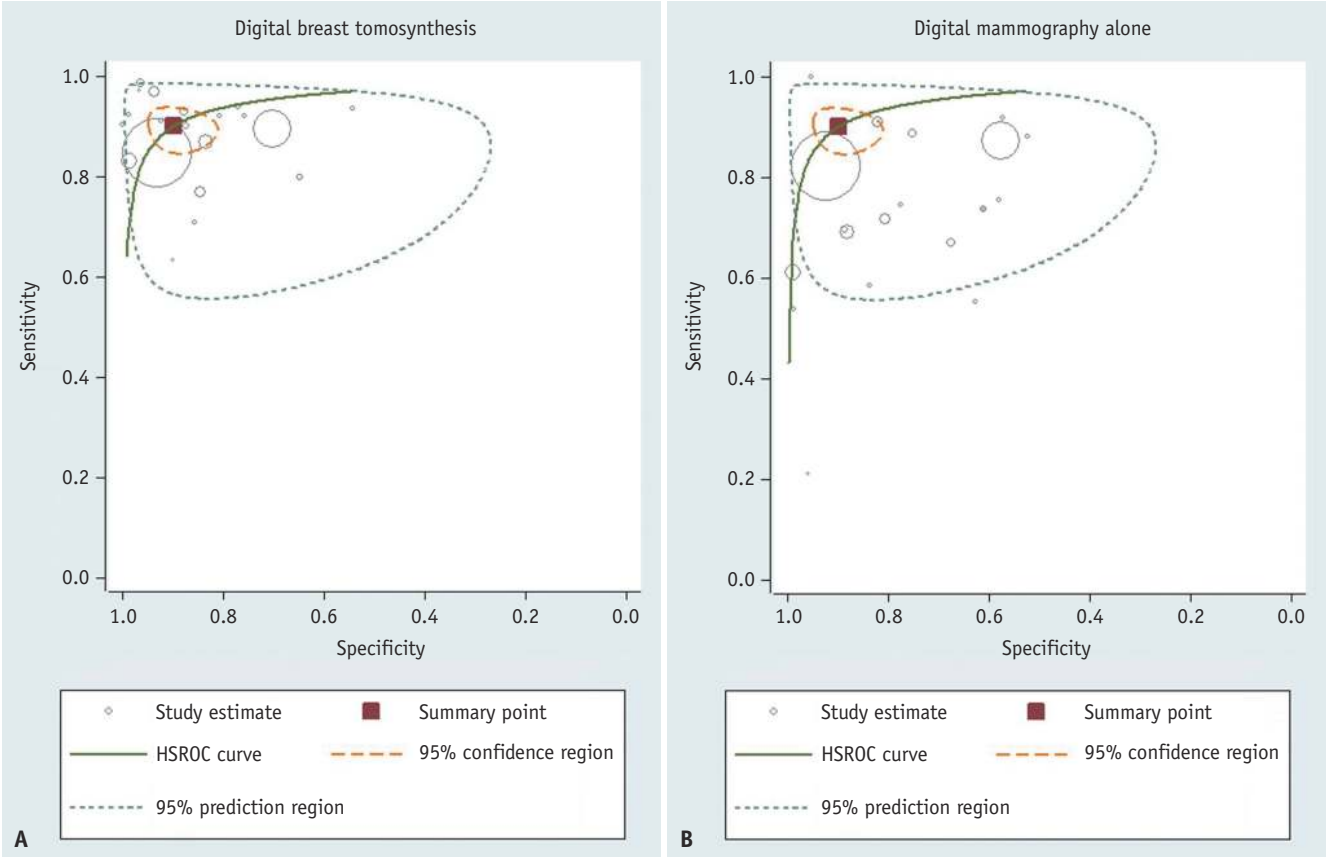

Fig. 4. HSROC curves of (A) digital breast tomosynthesis and (B) digital mammography alone for breast cancer diagnosis. HSROC curves show the individual (circles) and pooled (red square) sensitivity and specificity, and the dimension of each circle indicates the weight from the study sample size. There is a large difference between the 95% prediction (larger oval) and confidence region (small oval) in **(B)** digital mammography alone as well as **(A)** digital breast tomosynthesis, which suggests considerable heterogeneity among studies. HSROC = hierarchical summary receiver operating characteristic

pooled sensitivity was 90% (95% CI 86–93, I^2 = 91.4%) for DBT and 76% (95% CI 68–83, I^2 = 93.5%) for DM alone in diagnosing breast cancer. The pooled specificity to diagnose breast cancer was 90% (95% CI 84–94, I^2 = 99.3%) for DBT and 83% (95% CI 73–89, I^2 = 99.6%) for DM alone. In this study, we only included primary studies with a comparative design; two index tests of interest and meta-regression modeling were performed to evaluate the impact of important potential confounding variables on accuracy. Higher sensitivity and specificity of DBT as compared to DM alone was consistent in subgroup analysis for the modality of intervention, study design, existence of calcification, breast density, BI-RADS category threshold, protocol for participant sampling, number of participants, and country where the study was performed.

A previous meta-analysis by Lei et al. [38] assessed the diagnostic accuracy of DBT alone vs. DM alone in seven studies (2014 patients) and reported sensitivities of 90% and 89%, respectively, and specificities of 79% and 72% for DBT alone and DM alone, respectively. Another meta-analysis that included 38 comparative studies (488099 patients) reported sensitivities of 88%, 88%, and 79%, and specificities of 84%, 81%, and 79% for DBT alone, combined DBT and DM, and DM alone, respectively [39]. Although these previous meta-analyses reported similar results as those of the current study, they only analyzed studies with small sample sizes [38] and evaluated both screening and diagnostic populations [39]. In addition, we performed subgroup analysis providing diagnostic accuracies in each subgroup, and sensitivity analysis, whereas the previous study [39] provided beta coefficients of imaging modality and other covariates based on the multivariate meta-regression model.

Although DBT has been evaluated predominantly in the screening setting, it has been shown to be useful in the diagnostic setting, as well as to improve lesion characterization in noncalcified lesions when compared to conventional mammography [7,40,41]. The diagnostic

https://doi.org/10.3348/kjr.2020.1227

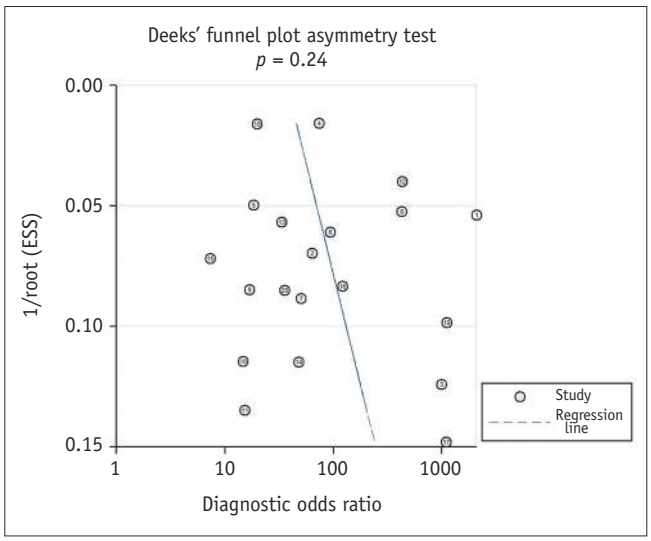

Fig. 5. Deeks' funnel plot for digital breast tomosynthesis. The *p* value of 0.24 for the slope coefficient indicates symmetry in the data and a low likelihood of publication bias. ESS = effective sample size

accuracy of breast DBT in the diagnostic workup of women with clinical signs and symptoms and in women recalled from screening has also been demonstrated to be equivalent to or better than supplemental diagnostic mammographic views in several studies [7,12,37,41]. For these reasons, the ACR Appropriateness Criteria® highly recommended the use of DBT to evaluate symptomatic women with a palpable mass or nipple discharge, especially those ≥ 40 years of age [42,43]. The higher accuracy of DBT than DM enabled the omission of unnecessary recall for additional workup and biopsies and increases workflow efficiency [34].

In most of the studies included in our meta-analysis, a relatively high area under the curve values of DBT compared to DM alone was noted. DBT alone even showed higher values than DM alone and DBT combined with DM. Although the reason for the superior value of DBT alone compared to DBT combined with DM is inevident, the use of synthetic mammography obtained from DBT could explain the high diagnostic value. Moreover, differences in readers' experience who interpreted the DBT also could affect the diagnostic performance. This observation was inconsistent in individual studies, and superiorities in sensitivity or specificity of DBT relative to those of DM were inconsistently noted. In the retrospective cohort study with 22824 DBT diagnostic examinations and 22883 DM conducted by Bahl et al. [21], the cancer detection rate and sensitivity were similar; however, specificity was higher in the DBT combined with DM group than DM alone group. The study that only included cases with calcifications showed equivalent diagnostic performance between DBT alone and DM [36]. It is often considered that calcifications cannot be well assessed by DBT and should therefore be excluded in DBT studies [44]. However, in our subgroup analysis, a statistically significant superior performance of DBT was noted, even in studies that did not excluded calcifications [18-25,27,31-33,35-37]. In studies that included symptomatic patients only, or either symptomatic patients or screening recalls, higher sensitivity and specificity values were noted although no strong evidence for an effect was noted in the subgroup of symptomatic patients only as the number of studies was small. These results indicate that diagnostic performance can be affected by the variabilities in patient characteristics. Furthermore, differing diagnostic thresholds that define a positive test need to be confirmed by more prospective studies.

Our study had a few limitations. First, most studies included in our meta-analysis were retrospective studies, and some "prospective" studies could be classified as retrospective reader studies as the images were collected prospectively; however, the images were evaluated later with the reader study. Second, the reasons for performing the diagnostic DBT and their clinical workflow could be variable accordingly; however, separate analyses for these different indications, including imaging abnormality and various symptoms, were not possible because of lack of separate data by different samplings in primary studies. Third, substantial study heterogeneity was observed. To identify the factors causing heterogeneity, we examined the threshold effect between sensitivity and specificity using coupled forest plot and Spearman correlation coefficient and performed sensitivity analyses as well as extensive meta-regression. Including more prospective studies with a larger study population might help to validate the present conclusions with relatively less heterogeneity. Fourth the obscurity of blinding and time interval between the index test and reference standard were notable areas for quality assessment. To ensure comparability and minimize the bias resulting from confounding factors, we only included primary studies with a comparative design and with the same reference standard. However, lack of detailed description regarding study protocol could influence the results. Further studies should provide a clear description according to each item requested QUADAS-2 tool.

In summary, this systematic review and meta-analysis of comparative studies showed that both DBT alone and

https://doi.org/10.3348/kjr.2020.1227

Table 2. Comparative Analysis of DBT with DM

Covariate	No. of Studies	Sensitivity (95% CI)			Specificity (95% CI)			$P^{‡}$
		DBT	DM Alone	P^{*}	DBT	DM Alone	$P^{†}$	
Modality								
DBT alone	11	0.93 (0.89–0.96)	0.80 (0.69–0.91)	0.030	0.90 (0.84–0.97)	0.75 (0.64–0.86)	0.020	0.010
DBT combined with DM	9	0.86 (0.81–0.92)	0.74 (0.68–0.80)	< 0.001	0.89 (0.82–0.97)	0.89 (0.81–0.98)	0.990	0.005
Country								
Asian	10	0.92 (0.88–0.96)	0.75 (0.65–0.84)	< 0.001	0.92 (0.86–0.97)	0.79 (0.67–0.91)	0.050	< 0.001
Non-Asian	10	0.88 (0.83–0.93)	0.77 (0.67–0.87)	0.070	0.88 (0.79–0.96)	0.86 (0.76–0.95)	0.810	0.130
Existence of calcification								
Yes	15	0.90 (0.86–0.93)	0.78 (0.72–0.85)	0.020	0.90 (0.85–0.96)	0.80 (0.70–0.89)	0.360	< 0.001
No	5	0.90 (0.81–0.99)	0.67 (0.45–0.88)	0.210	0.90 (0.79–1.00)	0.90 (0.80–1.00)	0.680	0.110
Breast density								
≥ 2 or b	5	0.90 (0.80–1.00)	0.70 (0.48–0.92)	0.290	0.92 (0.84–1.00)	0.90 (0.80–1.00)	0.700	0.140
All (1–4 or a–d)	15	0.90 (0.86–0.93)	0.77 (0.71–0.84)	0.020	0.89 (0.83–0.95)	0.80 (0.70–0.90)	0.360	< 0.001
BI-RADS threshold								
≥ 4	15	0.91 (0.87–0.94)	0.77 (0.70–0.84)	0.040	0.90 (0.84–0.96)	0.82 (0.72–0.91)	0.260	< 0.001
≥ 3	3	0.90 (0.82–0.94)	0.77 (0.57–0.98)	0.420	0.95 (0.87–1.00)	0.89 (0.75–1.00)	0.630	0.540
Study design								
Prospective	4	0.92 (0.86–0.98)	0.67 (0.54–0.80)	< 0.001	0.89 (0.78–1.00)	0.71 (0.54–0.88)	0.070	0.010
Retrospective	16	0.89 (0.85–0.93)	0.78 (0.71–0.86)	0.010	0.90 (0.85–0.96)	0.85 (0.77–0.93)	0.310	0.010
Sampling								
Consecutive	9	0.89 (0.84–0.94)	0.76 (0.68–0.84)	< 0.001	0.89 (0.82–0.97)	0.84 (0.70–0.98)	0.520	0.020
Non-consecutive	11	0.90 (0.86–0.95)	0.77 (0.66–0.87)	0.030	0.89 (0.82–0.97)	0.81 (0.74–0.88)	0.050	0.020
Sample sizes								
≥ 200	10	0.90 (0.85–0.94)	0.78 (0.71–0.85)	0.019	0.90 (0.83–0.97)	0.84 (0.75–0.94)	0.320	0.020
< 200	10	0.90 (0.86–0.95)	0.74 (0.61–0.87)	0.030	0.90 (0.83–0.97)	0.81 (0.68–0.94)	0.220	0.030
Indications for diagnostic examination								
Symptom only	3	0.93 (0.87–0.99)	0.76 (0.59–0.93)	0.330	0.95 (0.89–1.00)	0.93 (0.85–1.00)	0.850	0.050
Symptom or screening recall	17	0.89 (0.85–0.93)	0.76 (0.67–0.83)	< 0.001	0.89 (0.83–0.95)	0.80 (0.71–0.89)	0.200	< 0.001
No of readers								
≥ 5	7	0.87 (0.82–0.92)	0.84 (0.77–0.90)	0.400	0.83 (0.71–0.95)	0.76 (0.61–0.92)	0.750	0.610
< 5	13	0.91 (0.87–0.95)	0.70 (0.60–0.80)	< 0.001	0.93 (0.88–0.97)	0.86 (0.78–0.94)	0.210	< 0.001

*p value from likelihood ratio test in sensitivity between DBT and DM alone, †p value from likelihood ratio test in specificity between DBT and DM alone, ‡p value from likelihood ratio test in sensitivity and/or specificity between DBT and DM alone. CI = confidence interval, DBT = digital breast tomosynthesis, DM = digital mammography, NA = not applicable

https://doi.org/10.3348/kjr.2020.1227

Korean Journal of Radiology KJR

DBT combined with DM were more accurate than DM alone for the diagnosis of breast cancer. Women with clinical symptoms or abnormal screening findings could be more effectively evaluated for breast cancer using DBT, which has a superior diagnostic performance than DM alone.

Supplement

The Supplement is available with this article at https://doi.org/10.3348/kjr.2020.1227.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Min Jung Ko, Dong A Park, Sung Hyun Kim, Jung Min Chang. Data curation: Min Jung Ko, Sung Hyun Kim, Jung Min Chang. Formal analysis: Min Jung Ko, Dong A Park. Funding acquisition: Min Jung Ko. Methodology: Min Jung Ko, Dong A Park, Jung Min Chang. Project administration: Min Jung Ko, Sung Hyun Kim. Software: Min Jung Ko, Dong A Park. Validation: Min Jung Ko, Dong A Park, Sung Hyun Kim, Jung Min Chang. Visualization: Min Jung Ko, Jung Min Chang. Writing—original draft: Min Jung Ko, Dong A Park, Jung Min Chang. Writing—review & editing: all authors.

ORCID iDs

Min Jung Ko
https://orcid.org/0000-0003-3599-7173
Dong A Park
https://orcid.org/0000-0001-7225-3152
Sung Hyun Kim
https://orcid.org/0000-0001-9587-674X
Eun Sook Ko
https://orcid.org/0000-0002-0399-7956
Kyung Hwan Shin
https://orcid.org/0000-0002-5852-7644
Woosung Lim
https://orcid.org/0000-0002-2863-9346
Beom Seok Kwak
https://orcid.org/0000-0003-0180-5102
Jung Min Chang
https://orcid.org/0000-0001-5726-9797

REFERENCES

1. American College of Radiology. Practice guideline for the performance of screening and diagnostic mammography. Acr.org Web site. https://www.acr.org/-/media/ACR/Files/Practice-Parameters/Screen-Diag-Mammo.pdf. Published 2013. Accessed June 17, 2020
2. Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, et al. Mammographic density and the risk and detection of breast cancer. *N Engl J Med* 2007;356:227-236
3. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Ann Intern Med* 2003;138:168-175
4. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. *ACR BI-RADS atlas, breast imaging reporting and data system*, 5th ed. Reston, VA: American College of Radiology, 2013
5. Harvey JA, Nicholson BT, Cohen MA. Finding early invasive breast cancers: a practical approach. *Radiology* 2008;248:61-76
6. Zuley ML, Bandos AI, Ganott MA, Sumkin JH, Kelly AE, Catullo VJ, et al. Digital breast tomosynthesis versus supplemental diagnostic mammographic views for evaluation of noncalcified breast lesions. *Radiology* 2013;266:89-95
7. Noroozian M, Hadjiiski L, Rahnama-Moghadam S, Klein KA, Jeffries DO, Pinsky RW, et al. Digital breast tomosynthesis is comparable to mammographic spot views for mass characterization. *Radiology* 2012;262:61-68
8. Marinovich ML, Hunter KE, Macaskill P, Houssami N. Breast cancer screening using tomosynthesis or mammography: a meta-analysis of cancer detection and recall. *J Natl Cancer Inst* 2018;110:942-949
9. Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast cancer screening using tomosynthesis in combination with digital mammography. *JAMA* 2014;311:2499-2507
10. Bernardi D, Macaskill P, Pellegrini M, Valentini M, Fantò C, Ostillio L, et al. Breast cancer screening with tomosynthesis (3D mammography) with acquired or synthetic 2D mammography compared with 2D mammography alone (STORM-2): a population-based prospective study. *Lancet Oncol* 2016;17:1105-1113
11. Poplack SP, Tosteson TD, Kogel CA, Nagy HM. Digital breast tomosynthesis: initial experience in 98 women with abnormal digital screening mammography. *AJR Am J Roentgenol* 2007;189:616-623
12. Brandt KR, Craig DA, Hoskins TL, Henrichsen TL, Bendel EC, Brandt SR, et al. Can digital breast tomosynthesis replace conventional diagnostic mammography views for screening recalls without calcifications? A comparison study in a simulated clinical setting. *AJR Am J Roentgenol* 2013;200:291-298
13. Hakim CM, Chough DM, Ganott MA, Sumkin JH, Zuley ML,

Gur D. Digital breast tomosynthesis in the diagnostic environment: a subjective side-by-side review. *AJR Am J Roentgenol* 2010;195:W172-W176
14. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;319:388-396
15. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-560
16. Devillé WL, Buntinx F, Bouter LM, Montori VM, de Vet HC, van der Windt DA, et al. Conducting systematic reviews of diagnostic studies: didactic guidelines. *BMC Med Res Methodol* 2002;2:9
17. Lee J, Kim KW, Choi SH, Huh J, Park SH. Systematic review and meta-analysis of studies evaluating diagnostic test accuracy: a practical review for clinical researchers-part II. Statistical methods of meta-analysis. *Korean J Radiol* 2015;16:1188-1196
18. Basha MAA, Safwat HK, Alaa Eldin AM, Dawoud HA, Hassanin AM. The added value of digital breast tomosynthesis in improving diagnostic performance of BI-RADS categorization of mammographically indeterminate breast lesions. *Insights Imaging* 2020;11:26
19. You C, Zhang Y, Gu Y, Xiao Q, Liu G, Shen X, et al. Comparison of the diagnostic performance of synthesized two-dimensional mammography and full-field digital mammography alone or in combination with digital breast tomosynthesis. *Breast Cancer* 2020;27:47-53
20. Asbeutah AM, Karmani N, Asbeutah AA, Echreshzadeh YA, AlMajran AA, Al-Khalifah KH. Comparison of digital breast tomosynthesis and digital mammography for detection of breast cancer in Kuwaiti women. *Med Princ Pract* 2019;28:10-15
21. Bahl M, Mercaldo S, Vijapura CA, McCarthy AM, Lehman CD. Comparison of performance metrics with digital 2D versus tomosynthesis mammography in the diagnostic setting. *Eur Radiol* 2019;29:477-484
22. Georgian-Smith D, Obuchowski NA, Lo JY, Brem RF, Baker JA, Fisher PR, et al. Can digital breast tomosynthesis replace full-field digital mammography? A multireader, multicase study of wide-angle tomosynthesis. *AJR Am J Roentgenol* 2019;212:1393-1399
23. Li J, Zhang H, Jiang H, Guo X, Zhang Y, Qi D, et al. Diagnostic performance of digital breast tomosynthesis for breast suspicious calcifications from various populations: a comparison with full-field digital mammography. *Comput Struct Biotechnol J* 2019;17:82-89
24. Mall S, Noakes J, Kossoff M, Lee W, McKessar M, Goy A, et al. Can digital breast tomosynthesis perform better than standard digital mammography work-up in breast cancer assessment clinic? *Eur Radiol* 2018;28:5182-5194
25. Ohashi R, Nagao M, Nakamura I, Okamoto T, Sakai S. Improvement in diagnostic performance of breast cancer: comparison between conventional digital mammography alone and conventional mammography plus digital breast tomosynthesis. *Breast Cancer* 2018;25:590-596
26. Chan HP, Helvie MA, Hadjiiski L, Jeffries DO, Klein KA, Neal CH, et al. Characterization of breast masses in digital breast tomosynthesis and digital mammograms: an observer performance study. *Acad Radiol* 2017;24:1372-1379
27. Kim WH, Chang JM, Koo HR, Seo M, Bae MS, Lee J, et al. Impact of prior mammograms on combined reading of digital mammography and digital breast tomosynthesis. *Acta Radiol* 2017;58:148-155
28. Krammer J, Stepniewski K, Kaiser CG, Brade J, Riffel P, Schoenberg SO, et al. Value of additional digital breast tomosynthesis for preoperative staging of breast cancer in dense breasts. *Anticancer Res* 2017;37:5255-5261
29. Bian T, Lin Q, Cui C, Li L, Qi C, Fei J, et al. Digital breast tomosynthesis: a new diagnostic method for mass-like lesions in dense breasts. *Breast J* 2016;22:535-540
30. Elizalde A, Pina L, Etxano J, Slon P, Zalazar R, Caballeros M. Additional US or DBT after digital mammography: which one is the best combination? *Acta Radiol* 2016;57:13-18
31. Kamal R, Mansour S, ElMesidy D, Moussa K, Hussien A. Detection and diagnosis of breast lesions: performance evaluation of digital breast tomosynthesis and magnetic resonance mammography. *Egypt J Radiol Nucl Med* 2016;47:1159-1172
32. Seo M, Chang JM, Kim SA, Kim WH, Lim JH, Lee SH, et al. Addition of digital breast tomosynthesis to full-field digital mammography in the diagnostic setting: additional value and cancer detectability. *J Breast Cancer* 2016;19:438-446
33. Taha Ali TF, Magid AM, Tawab MA, El-Hariri MA, Abdel-Fattah ES. Potential impact of tomosynthesis on the detection and diagnosis of breast lesions. *Egypt J Radiol Nucl Med* 2016;47:351-361
34. Bansal GJ, Young P. Digital breast tomosynthesis within a symptomatic "one-stop breast clinic" for characterization of subtle findings. *Br J Radiol* 2015;88:20140855
35. Gilbert FJ, Tucker L, Gillan MG, Willsher P, Cooke J, Duncan KA, et al. Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a UK retrospective reading study (TOMMY Trial). *Radiology* 2015;277:697-706
36. Tagliafico A, Mariscotti G, Durando M, Stevanin C, Tagliafico G, Martino L, et al. Characterisation of microcalcification clusters on 2D digital mammography (FFDM) and digital breast tomosynthesis (DBT): does DBT underestimate microcalcification clusters? Results of a multicentre study. *Eur Radiol* 2015;25:9-14
37. Waldherr C, Cerny P, Altermatt HJ, Berclaz G, Ciriolo M, Buser K, et al. Value of one-view breast tomosynthesis versus two-view mammography in diagnostic workup of women with clinical signs and symptoms and in women recalled from screening. *AJR Am J Roentgenol* 2013;200:226-231
38. Lei J, Yang P, Zhang L, Wang Y, Yang K. Diagnostic accuracy of digital breast tomosynthesis versus digital mammography for benign and malignant lesions in breasts: a meta-analysis.

Eur Radiol 2014;24:595-602
39. Alabousi M, Zha N, Salameh JP, Samoilov L, Sharifabadi AD, Pozdnyakov A, et al. Digital breast tomosynthesis for breast cancer detection: a diagnostic test accuracy systematic review and meta-analysis. *Eur Radiol* 2020;30:2058-2071
40. Skaane P, Gullien R, Bjørndal H, Eben EB, Ekseth U, Haakenaasen U, et al. Digital breast tomosynthesis (DBT): initial experience in a clinical setting. *Acta Radiol* 2012;53:524-529
41. Zuley ML, Guo B, Catullo VJ, Chough DM, Kelly AE, Lu AH, et al. Comparison of two-dimensional synthesized mammograms versus original digital mammograms alone and in combination with tomosynthesis images. *Radiology* 2014;271:664-671
42. American College of Radiology. ACR Appropriateness Criteria® Palpable Breast Masses. Acsearch.acr.org Web site. http://acsearch.acr.org/docs/69495/Narrative/. Accessed June 17, 2020
43. American College of Radiology. ACR Appropriateness Criteria® Evaluation of Nipple Discharge. Acsearch.acr.org Web site. http://acsearch.acr.org/docs/3099312/Narrative/. Accessed June 17, 2020
44. Peppard HR, Nicholson BE, Rochman CM, Merchant JK, Mayo RC 3rd, Harvey JA. Digital breast tomosynthesis in the diagnostic setting: indications and clinical applications. *Radiographics* 2015;35:975-990

https://doi.org/10.3348/kjr.2020.1227