

Comparison of New and Established Full-Field Digital Mammography Systems in Diagnostic Performance

Eun Sook Ko, MD¹, Boo-Kyung Han, MD¹, Sun Mi Kim, MD², Eun Young Ko, MD¹, Mijung Jang, MD², Chae Yeon Lyou, MD², Jung Min Chang, MD³, Woo Kyung Moon, MD³, Rock Bum Kim, MD⁴

¹Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul 135-710, Korea; ²Department of Radiology, Seoul National University Bundang Hospital, Seongnam 463-707, Korea; ³Department of Radiology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul 110-744, Korea; ⁴Department of Preventive Medicine, Dong-A University School of Medicine, Busan 602-715, Korea

Objective: To compare the diagnostic performance of new and established full-field digital mammography (FFDM) systems.

Materials and Methods: During a 15-month period, 1038 asymptomatic women who visited for mammography were prospectively included from two institutions. For women with routine two-view mammograms from established FFDM systems, bilateral mediolateral oblique (MLO) mammograms were repeated using the new FFDM system. One of the four reviewers evaluated two-sets of bilateral MLO mammograms at 4-week intervals by using a five-point score for the probability of malignancy according to a Breast Imaging Reporting and Data System. The lesion type and breast density were determined by the consensus of two readers at each institution. The dichotomized mammographic results correlated with a final pathologic outcome and follow-up data. Receiver operating characteristic (ROC) curves, sensitivity, and specificity were compared in general and according to the lesion type and breast density.

Results: Of the 1038 cases, 193 (18.6%) had cancer. The areas under the ROC curve (AUC), sensitivity, and specificity of the established system were 0.815, 65.3%, and 90.2%, respectively. Those of the new system were 0.839, 68.4%, and 91.7%, respectively. There were no significant differences in the AUCs, sensitivities or the specificities in general between new and established systems ($P_s = 0.194, 0.590, 0.322$, respectively). We found no significant difference in these parameters according to lesion type or breast density.

Conclusion: The new FFDM system has a comparable diagnostic performance with established systems.

Index terms: Breast; Digital mammography; FFDM; Analysis; Comparison

Received July 31, 2012; accepted after revision December 6, 2012. This study was supported by a grant of the Korea Health technology R&D Project, Ministry of Health & Welfare, Republic of Korea (A101891).

Corresponding author: Boo-Kyung Han, MD, Department of Radiology and Center for Imaging Science, Samsung Medical Center, Sungkyunkwan University School of Medicine, 50 Irwon-dong, Gangnam-gu, Seoul 135-710, Korea.

• Tel: (822) 3410-0517 • Fax: (822) 3410-0084
• E-mail: bkhan@skku.edu

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Full-field digital mammography (FFDM) can improve the accuracy of mammography over screen-film mammography for pre- and perimenopausal women, women younger than 50 years-of-age, and women with dense breasts (1). A number of comparative studies of screen film mammography and FFDM have been published as screening programs and practices that made the transition to digital imaging (1-5). In addition, there have been a number of physical evaluation studies which investigated the technical performance of individual FFDM systems (6-9). These studies

have proven the benefit of digital mammography in respect to the objectivity of the improved image quality as well as the clinically improved detection of cancer. There are, however, many kinds of digital mammography systems and technologies on the market, and these systems utilize a variety of different X-ray spectra and significantly different approaches for the optimization of image quality and dose.

The incidence of breast cancer in Korea has been increasing due to the westernized lifestyle and eating habits and nowadays, breast cancer is the most common cancer that develops in Korean women next to thyroid cancer (10). Also, Asian women frequently show heterogeneously or homogeneously dense breasts as compared to Western women, so it is hard to detect breast cancer earlier. Therefore, a high image quality digital mammography system is absolutely indispensable (11). Although the number of digital mammograms performed has been increasing in Korea, most of the mammography systems used in Korea are import-dependent requiring a high cost. Recently, the new FFDM system made in Korea was developed with the approval of the Korea Food and Drug Administration. If this Korean digital mammography system shows a similar performance as compared to other established systems, it could enable a wider use of FFDM with high quality. Therefore, the purpose of this study was to compare the clinical performance of this new digital mammography system with established systems.

MATERIALS AND METHODS

Description of Cases

This study received appropriate Institutional Review Board Approval from each participating institution. Informed consents were obtained from all patients. From November 2010 to February 2012, 1038 women (mean age, 59 years; range, 41-69 years) were prospectively enrolled in this study at two institutions. In order to be included in this study, women should have met the following criteria: 1) asymptomatic screening; 2) women aged over 40 years; 3) no history of breast surgery or vacuum-assisted biopsy within 6 months; 4) no history of previous breast cancer. To enrich malignant cases, we included asymptomatic women who wanted further evaluation of imaging abnormalities, regardless of a history of recent core needle biopsy at an outside hospital. In addition, to reach a statistical power of 90%, the sample size was determined as 193 in the case group and 772 in the control group on the assumption that

the cancer rate is approximately 20% and the sensitivity of digital mammography is about 70% (1, 12).

Mammography and Ultrasound (US)

The new FFDM system used in this study was Brestige (Medifuture, Seoul, Korea). The comparable images were obtained from two established FFDM systems; Senographe 2000D or 2000DS (GE Healthcare, Milwaukee, WI, USA) and Selenia (Hologic, Bedford, CT, USA). A Senographe system was used on 917 (88.3%) women, and a Selenia system was used on 121 (11.7%) women. If women had routine two-view digital mammograms obtained from one of the two established systems within 6 weeks, they repeated the mediolateral oblique (MLO) view mammograms for bilateral breasts with Brestige. If not, they underwent routine two-view mammograms with an established system and additional MLO views for bilateral breasts with Brestige. Patients were informed about the additional radiation by taking additional MLO views which were included in our informed consent.

To exclude the mammographically occult but US-visible cancers, US was performed in all cases and if there was a suspicious finding, an US-guided biopsy was immediately performed.

Image Interpretation

All images were stored in a centralized picture archiving and communication system. Digital examinations consisted of digital images acquired with the manufacturer-recommended image processing techniques and were interpreted on 5-megapixel monitors at full resolution with one breast on each monitor. The interpretation of mammograms was performed in a dedicated reading room under controlled lighting conditions.

One of four study radiologists was assigned to evaluate the images of two FFDM systems, MLO mammograms obtained by Brestige first and then those by the other systems with at least a 4 week interval to minimize case recall. No time constraints were placed on readers for either system. All studies were masked in terms of patient identification and any clinical information including age, examination date, other imaging findings, pathologic findings and institution source identification. The study radiologists had a range of experience with breast image interpretation between 3 to 17 years. The cases were presented in a randomized order. The radiologists were able to adjust the viewing window and level and to magnify each

image during the reading. Mammograms were classified into 5 categories according to the Breast Imaging Reporting and Data System (BI-RADS) lexicon (13). The classification system consists of five categories: 1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suspicious for malignancy. They individually described the lesion type and its location of the most significant lesion which contributed to the final assessment category and the breast density. The lesion type was divided into: negative, mass, asymmetry, architectural distortion, and calcification or mass with calcification. Breast density was scored by 4 levels according to the BI-RADS system: 1, fatty; 2, scattered fibroglandular tissue; 3, heterogeneously dense; 4, extremely dense.

After all reading sessions were finished, the consensus meeting was made between two radiologists to confirm the lesion-to-lesion concordancy between two readers and to determine the representative lesion type and breast density for discordant cases. When a decision was difficult, they referred to all two-view mammograms and US imaging. To determine the density, only MLO views were allowed for review.

Outcome Analysis

The clinical outcome data were retrospectively collected using a medical chart review for the histologic results of the breast lesions and follow-up mammography results.

As a result of time constraint by funding institution, this study was initiated shortly after enrollment was complete. Therefore, mammographic follow-up data was obtained in less than one-half of the subjects. To compensate for this limitation, research assistants performed telephone interviews at the time of the closing period of this study. In cases where women were reported as cancer-free at the interview, we also regarded them as having a negative regarding diagnosis for cancer.

To establish a reference routine, the cases were classified as positive for cancer if a malignancy was pathologically verified within a year after the mammogram. The cases were classified as being negative for cancer 1) if there were no suspicious findings in the mammograms and US or if any suspicious findings on mammograms or US were histologically revealed as benign lesions; and 2) when follow-up mammograms or telephone interviews documented a cancer-free state. The follow-up mammograms were only available in 276 (33.8%) of 845 negative women within 6-15 months (mean, 8.3 months; median, 7.0 months) after

enrollment in the study.

The primary aim was to compare the area under the receiver operating characteristics (ROC) curve (AUC), sensitivity and specificity for new and other established systems in general. The ROC curves were constructed using the BI-RADS assessment category. The sensitivity and specificity were calculated from a dichotomized malignancy score, with a BI-RADS category of 1-3 considered as negative and a BI-RADS category of 4-5 considered as positive. The secondary aims were to compare those values on the basis of the lesion type and breast density. Statistical analysis was performed using a Fisher's exact test.

We also assessed the agreement of radiologic assessment between new and established FFDM systems by using a kappa statistic for each case. The rating scale for kappa-values suggests the following correlations: < 0.20 = poor; 0.21-0.40 = fair; 0.41-0.60 = moderate; 0.61-0.80 = good; > 0.81 = very good (14). A McNemar test was used to compare the proportion of benign and malignancy between new and established systems.

All analyses were performed using Statistical Package for the Social Sciences (SPSS) version 18.0 (SPSS Inc., Chicago, IL, USA), and MedCalc for Windows version 12.2.1.0 (MedCalc Software, Mariakerke, Belgium), with a *p* value of < 0.05 indicating a significant difference.

RESULTS

Table 1 shows the final outcomes according to the radiologic assessment by two FFDM systems. In 1038 cases, 193 (18.6%) were cancers. Invasive cancers accounted for 157 (81.3% of the cancers detected) and the remaining 36 (18.7%) cancers were ductal carcinoma *in situ*. The mean size of invasive cancers was 19.1 mm (range, 2-50 mm).

The findings of the study mammograms in terms of the lesion type and the breast density are detailed in Table 2. In terms of lesion type, 605 (58.3%) cases were determined as negative, and 190 (18.3%) were determined as lesions containing calcifications. Of the 193 cases with cancer, 119 (61.7%) had dense breasts. Twenty-nine (15.0%) of the total breast cancers were mammographically occult. These cancers were seen on neither of the mammography systems and were detected by supplementary US screening or were seen only on craniocaudal mammograms. Of the total women, 734 (70.7%) had mammographically dense breasts (BI-RADS score of 3 and 4) and 119 (61.7%) of 193 women

Table 1. Final Outcomes and Radiologic Assessments of 1038 Women with Two Sets of MLO Views from Two Mammographic Systems

	Final Outcome				Total (%)
	Negative	Positive			
		Subtotal	DCIS	Invasive Cancer	
Assessment by established system					
Negative (BI-RADS category 1-3)	762	67	14	53	829 (79.9)
Positive (BI-RADS category 4-5)	83	126	22	104	209 (20.1)
Assessment by new system					
Negative (BI-RADS category 1-3)	775	61	12	49	836 (80.5)
Positive (BI-RADS category 4-5)	70	132	24	108	202 (19.5)
Total (%)	845 (81.4)	193 (18.6)	36 (3.5)	157 (15.1)	1038 (100)

Note.— DCIS = ductal carcinoma *in situ*, BI-RADS = Breast Imaging Reporting and Data System

Table 2. Mammographic Findings of 1038 Women: Lesion Type and Breast Density according to Consensus Reports by Two Mammographic Systems

	Final Outcome				Total (%)
	Negative	Positive			
		Subtotal	DCIS	Invasive Cancer	
Lesion type					
Negative	576	29	7	22	605 (58.3)
Mass or asymmetry	148	90	8	82	238 (22.9)
Calcification only	98	44	17	27	142 (13.7)
Architectural distortion	4	1	0	1	5 (0.5)
Mass with calcification	19	29	4	25	48 (4.6)
Breast density					
Not dense (BI-RADS score 1 and 2)	228 (27.0)	74 (38.3)	7	67	304 (29.3)
Dense (BI-RADS score 3 and 4)	615 (73.0)	119 (61.7)	29	90	734 (70.7)
Total (%)	845	193	36	157	1038 (100)

Note.— DCIS = ductal carcinoma *in situ*, BI-RADS = Breast Imaging Reporting and Data System

with cancers had mammographically dense breasts ($p = 0.025$).

Table 3 presents a comparison of the diagnostic performance for two mammographic systems. In terms of AUC, sensitivity, and specificity, there were no significant differences between new and established systems, although the new system demonstrated better performance compared to established systems ($p = 0.194, 0.590, 0.322$, respectively). We found no significant difference of AUC, sensitivity, or specificity between a new and established system according to the breast density; fatty or dense breasts. We found no significant difference of diagnostic performance between new and established systems according to mammographic lesion types, masses (including asymmetry and architectural distortion) or calcification (including mass with calcification).

Table 4 shows the agreement of radiological assessment between new and established FFDM systems. There was no statistically significant difference in the proportion of negative and positive assessments between the two systems

($p = 0.608$). The kappa-value for agreement between the two systems was 0.584.

DISCUSSION

In this study, a new digital mammography system made in Korea showed a diagnostic performance similar to the established western digital mammography systems to detect breast cancers in women over 40 years of age. Two systems demonstrated 65.3-68.4% sensitivity and 90.2-91.7% specificity. The proportion of dense breasts was 70.7% and this performance was only achieved by one-view mammograms not by routine two-view mammograms. The slightly lower value of sensitivity and specificity might be due to these factors. Surely, it is better to perform routine two views (i.e., CC and MLO view) from each system to compare the diagnostic performance between two systems more accurately. However, as we were trying to minimize the radiation hazard to patients, only MLO views were obtained.

Table 3. Comparison of Diagnostic Performance between New and Established Systems

	AUC		Sensitivity (%)		Specificity (%)		P
	New (95% CI)	Established (95% CI)	New (95% CI)	Established (95% CI)	New (95% CI)	Established (95% CI)	
All lesion (n = 1038)	0.839 (0.815-0.861)	0.815 (0.790-0.838)	68.4 (61.3-74.9)	65.3 (58.1-72.0)	91.7 (89.7-93.5)	90.2 (88.0-92.1)	0.322
Mass (n = 243)*	0.790 (0.734-0.840)	0.747 (0.688-0.801)	79.1 (69.3-87.0)	75.8 (65.7-84.2)	75.7 (68.0-82.3)	71.1 (63.2-78.1)	0.437
Calcification (n = 190) [†]	0.858 (0.800-0.904)	0.824 (0.762-0.875)	78.1 (66.9-87.0)	78.1 (66.9-87.0)	83.8 (75.8-90.0)	76.1 (67.3-83.5)	0.191
Breast density							
Not dense (n = 304)	0.866 (0.822-0.902)	0.862 (0.818-0.898)	74.3 (62.9-83.8)	78.4 (67.3-87.1)	90.0 (85.4-93.6)	87.0 (82.0-91.0)	0.389
Dense (n = 734)	0.821 (0.792-0.848)	0.784 (0.753-0.814)	64.7 (55.4-73.2)	57.1 (47.8-66.2)	92.4 (90.0-94.3)	91.4 (88.9-93.5)	0.590

Note.— *Including asymmetry or architectural distortion, †Including mass with calcification. AUC = areas under receiver operating characteristic curve, CI = confidence interval

Table 4. Agreement of Radiologic Assessment between Two Mammographic Systems

	New System		Total
	Benign	Malignant	
Established systems			
Benign	764 (92.2%)	65 (7.8%)	829
Malignant	72 (34.4%)	137 (65.6%)	209
Total	836	202	1038

Traditional film mammography has been the single best breast cancer screening test to date and it has been shown to reduce mortality from breast cancer in large randomized trials (15, 16). The reason why mammography using an X-ray has been used as the primary method to detect breast cancer is that mammography is the most sensitive method to detect microcalcifications which are one of the main findings of breast cancer and even the recent technical development of other modalities (i.e., breast US and MRI) cannot reach the power of mammography to detect suspicious calcifications (17). A mammography, however, is far from perfect. Approximately 10% of cancers are mammographically occult even after they are palpable (18). Moreover, if the lesion is masked by surrounding fibroglandular tissue, it is difficult to differentiate breast cancers from superimposed normal parenchyma. The false-negative rate of mammography is related to mammographic breast density, which is known for about 25% of breast cancers (19).

Ever since the earliest comparative trials, known as the Colorado-Massachusetts study published in 2002 (20) and the Oslo I trial (4) published in 2003 showed only minimal inferiority that was not statistically significant of digital mammography compared with screen film, several studies showing at least the equivalence of the two technologies were followed (1, 3, 5). Consequently, as digital mammography has been widely accepted, much of all newly-adopted mammography systems in Korea are now changing into digital systems. It is well known that digital mammography has many benefits (21); Digital mammography, like other digital modalities, allows for the digital storage and transmission of each study, eliminating lost films and eventually eliminating the need for a film library. Also, images can be sent electronically to several treating physicians simultaneously, or given to the patient, without any loss of quality. In addition, it could eliminate film artifacts, such as dust and structural noise caused by film processing. A digital system also reduces the variability in contrast, density, dose, and exposure

time associated with film emulsion and processing. From a patient's perspective, the biggest advantage of a digital system is speed. By not having to wait for films to develop, the time for a diagnostic mammography examination is shortened from the exposure to the radiologist's final assessment. Also, wire localization becomes much faster and a patient does not have to stay in compression while the last image taken is processed and then reviewed by the radiologist. Furthermore, because of the large dynamic range, digital mammography is ideal for imaging women with breast implants. Similarly, a large dynamic range of digital mammography makes it ideal to evaluate the skin and tissues that are just beneath the skin. This tissue is typically blackened on a well-exposed film mammogram, requiring a special hot light partially to recover the information.

The results reported here suggest that the diagnostic performance of the new FFDM system, in terms of AUC, sensitivity and specificity is equivalent although the agreement to the established FFDM system is not as high as expected (κ value = 0.584). We guess the reason for relatively low agreement between the two systems is because we reviewed only MLO views and assessment in each system could be different if a lesion was better seen in one system. As demonstrated in Table 4, 65 mammograms were assessed as negative (BI-RADS assessment category 1-3) on the established system but as positive (BI-RADS assessment category 4, 5) on the new system. Similarly, 72 mammograms were assessed differently between the two systems, positive on the established system but negative on the new system. Among 65 discordant mammograms, 24 (36.9%) were finally identified as positive for malignancy. Similarly, 18 (25%) of the 72 mammograms were finally proven to be malignant. These discordant cases explain the low agreement between the two systems. The most common cause of discordance was the different delineation of the lesion in different systems probably caused by the technician or physical factors.

Our study has several limitations. First, we could not review routine two-view mammograms in each patient. Thus, calculated AUC, sensitivity, and specificity might not reflect actual clinical practice. However, to reduce the radiation dose exposure to patients, performing only MLO views was inevitable. Second, the number of patients having a follow-up mammogram was relatively small due to the short study period. To prevent a false-negative diagnosis, we performed breast US or a telephone interview for patients who did not

have a follow-up mammogram. Therefore, we believe not having follow-up mammograms in all patients did not affect the results of diagnostic performance.

In conclusion, we found no significant difference in diagnostic performance to detect breast cancers between new and established FFDM systems for all lesions, for both mass-type lesions and calcification-type lesions, and for both fatty and dense breasts. Based on these results, we suggest that the new Korean FFDM system, Brestige, is not inferior to the established systems.

REFERENCES

1. Pisano ED, Gatsonis C, Hendrick E, Yaffe M, Baum JK, Acharyya S, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med* 2005;353:1773-1783
2. Karssemeijer N, Bluekens AM, Beijerinck D, Deurenberg JJ, Beekman M, Visser R, et al. Breast cancer screening results 5 years after introduction of digital mammography in a population-based screening program. *Radiology* 2009;253:353-358
3. Vinnicombe S, Pinto Pereira SM, McCormack VA, Shiel S, Perry N, Dos Santos Silva IM. Full-field digital versus screen-film mammography: comparison within the UK breast screening program and systematic review of published data. *Radiology* 2009;251:347-358
4. Skaane P, Young K, Skjennald A. Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading--Oslo I study. *Radiology* 2003;229:877-884
5. Skaane P, Skjennald A. Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program--the Oslo II Study. *Radiology* 2004;232:197-204
6. Lazzari B, Belli G, Gori C, Rosselli Del Turco M. Physical characteristics of five clinical systems for digital mammography. *Med Phys* 2007;34:2730-2743
7. Baldelli P, Phelan N, Egan G. A novel method for contrast-to-noise ratio (CNR) evaluation of digital mammography detectors. *Eur Radiol* 2009;19:2275-2285
8. Baldelli P, Phelan N, Egan G. Investigation of the effect of anode/filter materials on the dose and image quality of a digital mammography system based on an amorphous selenium flat panel detector. *Br J Radiol* 2010;83:290-295
9. Hendrick RE, Pisano ED, Averbukh A, Moran C, Berns EA, Yaffe MJ, et al. Comparison of acquisition parameters and breast dose in digital mammography and screen-film mammography in the American College of Radiology Imaging Network digital mammographic imaging screening trial. *AJR Am J Roentgenol* 2010;194:362-369
10. Ministry for Health, Welfare and Family Affairs. Annual Report of cancer incidence (2007), cancer prevalence (2007) and

- survival (1993-2007) in Korea. Seoul: Ministry for Health Welfare and Family Affairs, 2009.
11. del Carmen MG, Halpern EF, Kopans DB, Moy B, Moore RH, Goss PE, et al. Mammographic breast density and race. *AJR Am J Roentgenol* 2007;188:1147-1150
 12. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. *J Clin Epidemiol* 2005;58:859-862
 13. American College of Radiology. Breast imaging reporting and data system (BI-RADS), ultrasound, 4th ed. Reston, VA: American College of Radiology, 2003
 14. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-174
 15. Nyström L, Rutqvist LE, Wall S, Lindgren A, Lindqvist M, Rydén S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. *Lancet* 1993;341:973-978
 16. Tabàr L, Fagerberg G, Duffy SW, Day NE, Gad A, Gröntoft O. Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am* 1992;30:187-210
 17. Kim HS, Han BK, Choo KS, Jeon YH, Kim JH, Choe YH. Screen-film mammography and soft-copy full-field digital mammography: comparison in the patients with microcalcifications. *Korean J Radiol* 2005;6:214-220
 18. Bird RE. Low-cost screening mammography: report on finances and review of 21,716 consecutive cases. *Radiology* 1989;171:87-90
 19. Mandelson MT, Oestreich N, Porter PL, White D, Finder CA, Taplin SH, et al. Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. *J Natl Cancer Inst* 2000;92:1081-1087
 20. Lewin JM, D'Orsi CJ, Hendrick RE, Moss LJ, Isaacs PK, Karellas A, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. *AJR Am J Roentgenol* 2002;179:671-677
 21. Lewin JM, D'Orsi CJ, Hendrick RE. Digital mammography. *Radiol Clin North Am* 2004;42:871-884, vi