Radiology

ORIGINAL RESEARCH • BREAST IMAGING

Effect of Digital Mammography for Breast Cancer Screening: A Comparative Study of More than 8 Million Korean Women

Seri Hong, MD, PhD • Soo Yeon Song, MPH • Boyoung Park, MD, PhD • Mina Suh, MD, PhD • Kui Son Choi, PhD • Seung Eun Jung, MD, PhD • Min Jung Kim, MD, PhD • Eun Hye Lee, MD, PhD • Chan Wha Lee, MD, PhD • Jae Kwan Jun, MD, PhD

From the National Cancer Control Institute, National Cancer Center, 323 Ilsan-ro, Ilsandong-gu, Goyang 10408, Republic of Korea (S.H., S.Y.S., B.P., M.S., K.S.C., J.K.J.); Graduate School of Cancer Science and Policy, National Cancer Center, Goyang, Republic of Korea (K.S.C., J.K.J.); Department of Radiology, Eunpyeong St Mary's Hospital, the Catholic University of Korea College of Medicine, Seoul, Republic of Korea (S.E.J.); Department of Radiology, Severance Hospital, Yonsei University College of Medicine, Seoul, Republic of Korea (M.J.K.); Department of Radiology, Bucheon Hospital, Soonchunhyang University College of Medicine, Bucheon, Republic of Korea (E.H.L.); and Department of Radiology, National Cancer Center Hospital, National Cancer Center, Goyang, Republic of Korea (C.W.L.). Received April 25, 2019; revision requested June 7; final revision received October 4; accepted October 8. **Address correspondence to** J.K.J. (e-mail: *jkjun@ncc.re.kr*).

Supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute funded by the Ministry of Health and Welfare, Republic of Korea (grant HI15C1257). The sponsors had no role in the study design, data collection, data analysis, data interpretation, or manuscript writing.

Conflicts of interest are listed at the end of this article.

Radiology 2019; 00:1–9 • https://doi.org/10.1148/radiol.2019190951 • Content codes: BR OI

Background: Full-field digital mammography (FFDM) has been accepted as a superior modality for breast cancer screening compared with conventional screen-film mammography (SFM), especially in women younger than 50 years or with dense breasts.

Purpose: To evaluate the accuracy of FFDM for breast cancer screening.

Materials and Methods: Data from January 1, 2011 to December 31, 2013 in the database from a nationwide breast cancer screening program linked with the national cancer registry were retrospectively analyzed. The study included Korean women aged 40–79 years who had undergone screening mammography with SFM, computed radiography (CR), or FFDM. The sensitivity, specificity, positive predictive value (PPV), and absolute and relative differences among these modalities were calculated, followed by pairwise comparison tests with multiple testing corrections. The areas under the receiver operating characteristic curve (AUCs) were also estimated and compared by using the DeLong method with Bonferroni correction.

Results: Among the 8 482 803 women included (mean age, 55 years ± 10), 34.4% (2 920 279 of 8 482 803), 51.7% (4 385 807 of 8 482 803), and 13.9% (1 176 717 of 8 482 803) underwent SFM, CR, and FFDM, respectively. The sensitivity and PPV were higher for FFDM than for SFM (adjusted odds ratio, 1.77 [95% confidence interval {CI}: 1.62, 1.95] for sensitivity and 1.36 [95% CI: 1.29, 1.43] for PPV) and CR (adjusted odds ratio, 1.70 [95% CI: 1.56, 1.85] for sensitivity and 1.26 [95% CI: 1.20, 1.32] for PPV), whereas specificity was lower with FFDM. The overall AUC for FFDM was 0.80 (95% CI: 0.80, 0.81), which was higher than that for SFM (0.75 [95% CI: 0.75, 0.76]) and CR (0.76 [95% CI: 0.75, 0.76]). $P < .05$ was found for differences in sensitivity, PPV, and AUC among modalities overall and in most of the subgroups of age, breast density, and screening round.

Conclusion: Full-field digital mammography allows better discrimination or prediction of breast cancer in the general female population than screen-film mammography or computed radiography, regardless of age, breast density, or screening round.

Online supplemental material is available for this article.

Many studies have demonstrated that the performance of digital mammography is at least equal to that of film-based mammography (1–13), which has demonstrated a reduction in breast cancer mortality in multiple randomized trials. A few comparative studies have reported that digital mammography is more accurate than screen-film mammography (SFM) (1,2). However, those studies had various limitations, and the results were valid only in certain conditions. First, they regarded two different systems—offline computed radiography (CR) and full-field digital mammography (FFDM)—as a single digital mammography system and performed combined estimations; second, the results were confined to women with dense breasts and to specific age groups. Although a 2013 meta-analysis (14) compared the summary receiver operating characteristic (ROC) curves and concluded that FFDM had better accuracy than SFM, those results were limited to women younger than 50 years; the study showed comparable accuracy in the overall female population.

The 2016 U.S. Preventive Services Task Force recommendation statement (15) also suggests digital mammography as a preferred screening modality because it has a higher sensitivity than SFM, and, like a conclusion of the previous study, the suggestion is that it is indicated for women younger than 50 years. It should be noted, however, that most of the studies based on U.S. cohorts did not treat CR and FFDM systems separately, which could have revealed obvious differences in their screening performances (3,16). Therefore, it is necessary to reevaluate CR and FFDM separately to achieve more precise test results among the different systems (17). Moreover, most studies involving digital mammography systems were performed

Abbreviations

ACR = American College of Radiology, AUC = area under the ROC curve, BI-RADS = Breast Imaging Reporting and Data System, CI = confidence interval, CR = computed radiography, FFDM = full-field digital mammography, NCSP = National Cancer Screening Program, PPV = positive predictive value, ROC = receiver operating characteristic, SFM = screen-film mammography

Summary

Full-field digital mammography had higher sensitivity and superior screening accuracy than screen-film mammography and computed radiography, despite its slightly lower specificity.

Key Results

- In this study of 8 482 803 mammograms, full-field digital mammography (FFDM) showed an overall sensitivity of 73% and an area under the receiver operating characteristic curve (AUC) of 0.80; in contrast, screen-film mammography (SFM) had a sensitivity of 61% and an AUC of 0.75 ($P < .001$ for comparison), and values for computed radiography (CR) were similar (sensitivity, 62%; AUC, 0.76).
- FFDM was superior to SFM (adjusted odds ratio for sensitivity was 1.77; $P < .001$ was found for differences in both sensitivity and AUC) and CR (adjusted odds ratio for sensitivity was 1.70; $P < .001$ was found for differences in both sensitivity and AUC) in the general female population, regardless of age, breast density, or screening round.

with data from the early 2000s. Because the technology had recently been introduced, the overall performance and accuracy indicators may have been affected by nonstandardization of the mammography equipment or the radiologists' lack of experience with the new systems. However, it has been more than 10 years since the implementation of digital mammography systems, and their reliability has increased. The equipment has been standardized, and the reading skills of the radiologists have improved and are more consistent. These changes may produce results significantly different from those of prior studies, but current research is lacking.

We hypothesized that FFDM compared with other modalities would now exhibit superior accuracy in the general population. This study aimed to investigate and compare the screening accuracy of FFDM with that of other mammography systems (CR and SFM) for breast cancer screening in the general Korean female population, regardless of age or breast density.

Materials and Methods

Study Population

Since 2002, all Korean women aged 40 years or older receive biennial invitations for mammographic screening through the National Cancer Screening Program (NCSP). A detailed explanation of the NCSP is available elsewhere (18). We retrospectively reviewed the data of breast cancer screening participants from 2011–2013 collected from nationwide screening units by the National Health Insurance Service. There were 3 380 461 mammograms from 2288 screening units in 2011, 3 346 958 mammograms from 2404 units in 2012, and 3 322 971 mammograms from 2469 units in 2013. We excluded women with a history of breast cancer, those who were older than 80 years at the time of screening, and those with incomplete screening results. We also excluded mammograms from screening units that did not fulfill the quality control standard for each year (165 units in 2011, 178 units in 2012, and 183 units in 2013) and those with inadequate information regarding the mammography equipment used. Of 10 050 390 mammograms (from 2828 screening units) obtained from 2011–2013 by the NCSP, 8 482 803 (from 2379 screening units) were included in the analysis (Fig 1).

Of the 2379 screening units included in the 2011–2013 national cancer screening data, 1511 institutions participated in all 3 years, 452 institutions participated in 2 years, and 416 institutions were involved in only 1 year. For when the annual average reading volume was calculated from the total number of examinations performed at each screening unit, the distribution is shown in Figure 2. Approximately two-thirds of the institutions (screening units) had an annual average number of mammographic screenings of less than 1000, and the remaining one-third of the screening units (ie, units with 1000 or more annual average mammograms) accounted for about 20% of the total number of annual mammographic screening examinations. The average number of mammograms or reading volume per year was 1303 for all the screening units included.

The participants consented to use of their data for public purposes; the requirement for written informed consent was waived. These NCSP data included information regarding age; screening round; screening results; screening units where mammography was performed; and socioeconomic status, classified into Medical Aid Program recipients and National Health Insurance Service beneficiaries (divided into two groups according to income bracket). Approval was obtained from the institutional review board of the National Cancer Center.

Mammography Equipment and Cancer Identification

All participants underwent breast cancer screening with one of three mammography systems (SFM, CR, or FFDM), depending on the screening unit and time of the examination. Information regarding the mammography equipment used at each screening unit was obtained from the Korean Institute for Accreditation of Medical Images 2011−2015 database of mammography. The data were collected through the medical imaging quality control program in Korea (19), which gathers equipment information from each screening unit every year. The mammography equipment information was linked to patient information by screening-unit registration numbers, allowing us to identify whether the examinations were performed with SFM, CR, or FFDM.

Screening results were reported as “negative” (category 1), “benign disease” (category 2), “suspicious finding” (category 4 or category 5), or “incomplete evaluation” (category 0), based on the assessment categories of the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS), 4th edition (20). In our analysis, categories 1 and 2 were considered negative results, whereas categories 4, 5, and 0 were considered positive results. Breast tissue density was classified into “dense” and “fatty” subgroups, according to the categories defined in ACR BI-RADS; “dense” includes “heterogeneously

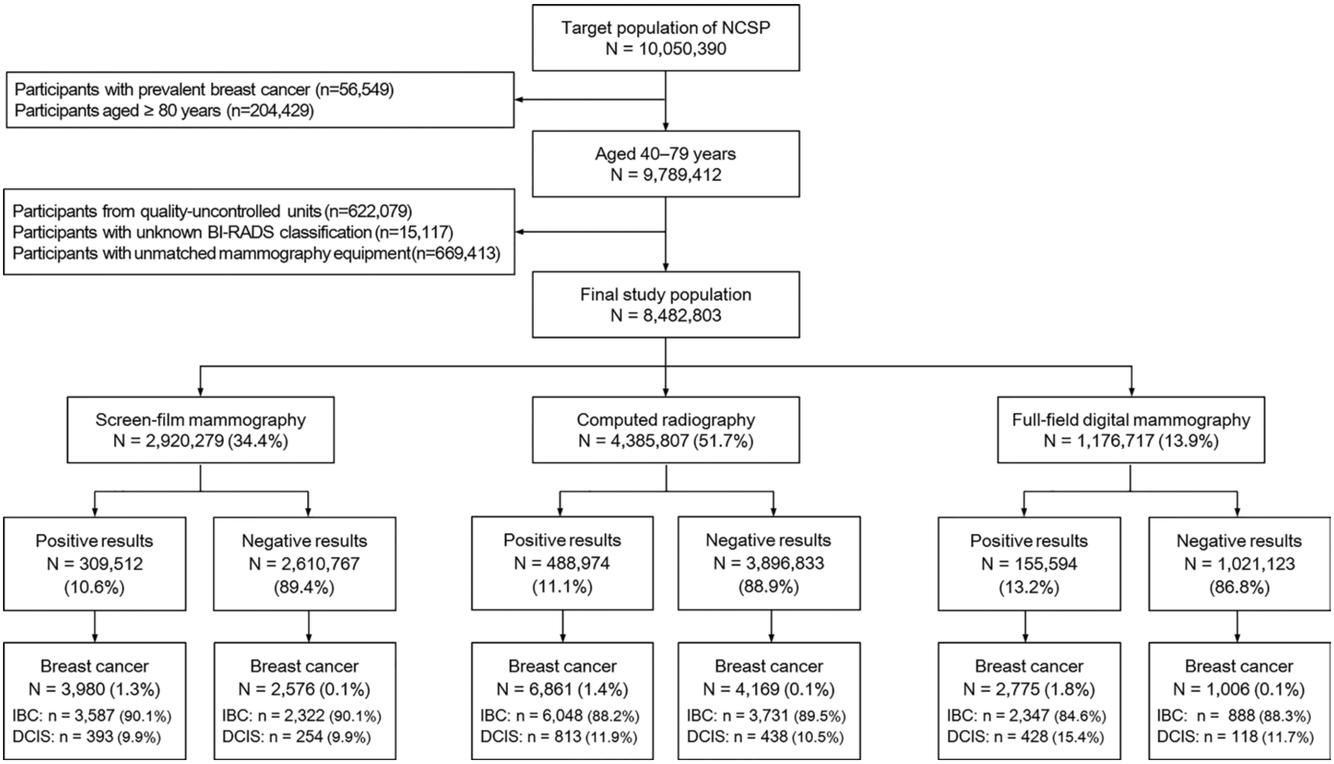


Figure 1: Flowchart of screening mammography examinations during study period. BI-RADS = Breast Imaging Reporting and Data System, DCIS = ductal carcinoma in situ, IBC = invasive breast cancer, NCSP = National Cancer Screening Program.

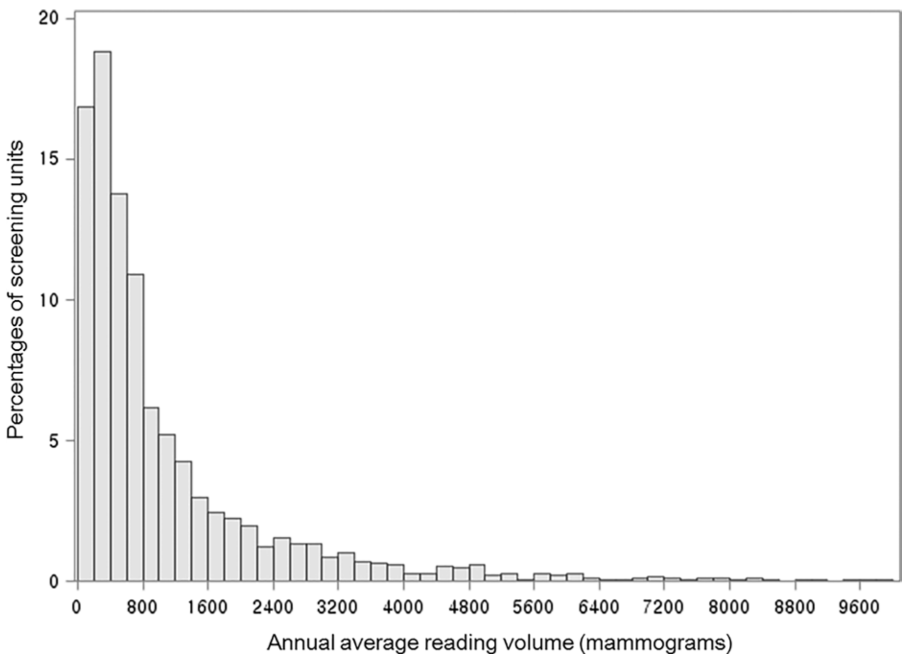


Figure 2: Bar graph shows distribution of average mammography reading volume per year among 2379 included screening units.

dense” and “extremely dense” breast tissue, and “fatty” includes the categories “almost entirely fat” and “scattered fibroglandular densities.” Breast cancer was confirmed by reviewing the records in the Korea Central Cancer Registry database. The records were estimated to be 98.2% complete (21) up to December 2014; patients diagnosed within 365 days after the baseline screening mammogram of the year were identified and linked to the NCSP database by using an individual registration number. Breast cancer included invasive cancer (International Classification of Diseases, 10th revision, code C50) and intraductal carcinoma in situ (International Classification of Diseases, 10th revision, code D05.1).

Statistical Analysis

The distribution of numbers and proportions among mammograms and the corresponding screening results and mammography equipment were assessed. Differences between the proportions were evaluated by using the χ^2 test. Sensitivity, specificity, and positive predictive value (PPV or PPV_1, based on positivity at screening examination) were estimated according to the patients’ characteristics, with exact 95% confidence intervals (CIs) determined by using binomial distribution. Absolute differences between SFM and FFDM and between CR and FFDM were calculated, and pairwise comparisons between pairs of proportions with correction for multiple testing (pairwise.prop.test—a pairwise two-sample test to determine the equality of proportions based on the χ^2 test, adjusted for multiple testing by means of the Holm

Table 1: Basic Characteristics of All 8 482 803 Korean Women in the Study according to Mammography Type

Characteristic	Screen-Film Mammography (*n* = 2 920 279)	Computed Radiography (*n* = 4 385 807)	Full-Field Digital Mammography (*n* = 1 176 717)	*P* Value
Year				<.001
2011	1 221 557 (41.8)	1 303 008 (29.7)	273 866 (23.3)	
2012	931 737 (31.9)	1 514 986 (34.5)	392 404 (33.3)	
2013	766 985 (26.3)	1 567 813 (35.7)	510 447 (43.4)	
Age (y)				<.001
Mean ± standard deviation	56.2 ± 10.7	55.0 ± 10.0	54.7 ± 9.7	
40–49	878 689 (30.1)	1 411 996 (32.2)	373 747 (31.8)	
50–59	934 395 (32.0)	1 525 148 (34.8)	426 607 (36.3)	
60–69	640 427 (21.9)	951 289 (21.7)	266 040 (22.6)	
70–79	466 768 (16.0)	497 374 (11.3)	110 323 (9.4)	
Breast density				<.001
Almost entirely fat	795 792 (27.3)	1 090 862 (24.9)	178 942 (15.2)	
Scattered fibroglandular densities	785 210 (26.9)	1 270 148 (28.9)	362 575 (30.8)	
Heterogeneously dense	830 836 (28.5)	1 380 023 (31.5)	471 379 (40.1)	
Extremely dense	508 441 (17.4)	644 774 (14.7)	163 821 (13.9)	
Screening round				<.001
Baseline screening	597 247 (20.5)	899 920 (20.5)	244 646 (20.8)	
Subsequent screening	2 323 032 (79.5)	3 485 887 (79.5)	932 071 (79.2)	
Health insurance type				<.001
National Health Insurance, upper 50%	1 249 757 (42.8)	1 968 417 (44.9)	604 610 (51.4)	
National Health Insurance, lower 50%	1 549 746 (53.1)	2 278 405 (51.9)	544 753 (46.3)	
Medical Aid Program	120 776 (4.1)	138 985 (3.2)	27 354 (2.3)	

Note.—Except where indicated, data are numbers of patients, with percentages in parentheses.

method) were performed to evaluate their significance. Logistic regression analysis was performed to evaluate relative differences in diagnostic accuracy among the types of mammography equipment after adjustment for screening year, patient age, breast density, screening round, and health insurance type. The areas under the ROC curve (AUCs) with 95% CIs for each mammography modality were calculated by using ROC analysis and were compared by using the DeLong method with Bonferroni adjustment for a multiple-comparisons correction with the pROC package (22). Sensitivity analyses were performed for screening units with an average of 1000 or more annual screening examinations (812 screening units, from which 6 900 164 mammograms [about 81.3% of the final study mammograms] were included) according to the radiologists' yearly interpretive volume number recommended in the quality guidelines for breast cancer screening in Korea (23). *P* values were two sided, and $P < .05$ was considered to indicate a statistically significant difference. All statistical analyses were performed with software (SAS version 9.3; SAS Institute, Cary, NC; and R version 3.4.2; R Foundation for Statistical Computing, Vienna, Austria].

Results

Basic Characteristics

The final data set (n = 8 482 803) consisted of 2 920 279 (34.4%) SFM mammograms, 4 385 807 (51.7%) CR mammograms, and 1 176 717 (13.9%) FFDM mammograms. Mammograms obtained with FFDM were more recent on an average, compared with SFM or CR, and a higher proportion of participants who underwent FFDM belonged to the higher socioeconomic status group compared with those who underwent SFM or CR. The proportions of women in their 40s and those with dense breasts were greater than one-third and approximately half of the total study population, respectively. The distribution of the participants' characteristics based on the type of mammography is presented in Table 1.

The recall rates and cancer detection rates of SFM, CR, and FFDM were 10.6% and 1.4 cancer cases per 1000 screens (n = 309 512 and n = 3980 of 2 920 279 screens); 11.1% and 1.6 per 1000 screens (n = 488 974 and n = 6861 of 4 385 807 screens); and 13.2% and 2.4 per 1000 screens (n = 155 594 and n = 2775 of 1 176 717 screens), respectively. In total, 13 616 detected breast cancer cases were included (Fig 1). Among the cancer cases detected with the three types of mammography, those diagnosed with FFDM had the largest proportion of ductal carcinoma in situ and invasive breast cancer (Table E1 [online]). The recall rates and cancer detection rates were similar in a subanalysis of screening units with an average number of screening mammography examinations of 1000 or more per year.

Performance Indicators in Three Mammography Modalities

The screening accuracy indicators in FFDM were distinct from those in SFM and CR, with higher sensitivity (adjusted odds ratios for sensitivity were 1.77 compared with SFM and 1.70

Table 2: Screening Accuracy of Screen-Film Mammography, Computed Radiography, and Full-Field Digital Mammography in Korean Women

Accuracy Indicator	Screen-Film Mammography	Computed Radiography	Full-Field Digital Mammography	Difference 1*			Difference 2†		
				Result	*P* Value	aOR‡	Result	*P* Value	aOR‡
Total									
Sensitivity	60.7 (59.5, 61.9)	62.2 (61.3, 63.1)	73.4 (72.0, 74.8)	12.7 (10.8, 14.6)	<.001	1.77 (1.62, 1.95)	11.2 (9.5, 12.9)	<.001	1.70 (1.56, 1.85)
Specificity	89.5 (89.5, 89.6)	89.0 (89.0, 89.0)	87.0 (86.9, 87.0)	−2.5 (−2.6, −2.4)	<.001	0.85 (0.84, 0.85)	−2.0 (−2.1, −1.9)	<.001	0.90 (0.90, 0.91)
PPV	1.3 (1.2, 1.3)	1.4 (1.4, 1.4)	1.8 (1.7, 1.9)	0.5 (0.4, 0.6)	<.001	1.36 (1.29, 1.43)	0.4 (0.3, 0.5)	<.001	1.26 (1.20, 1.32)
Age 40–44 y									
Sensitivity	49.0 (46.2, 51.9)	54.6 (52.4, 56.7)	65.8 (62.1, 69.3)	16.8 (12.2, 21.4)	<.001	1.96 (1.61, 2.39)	11.2 (7.0, 15.4)	<.001	1.57 (1.31, 1.88)
Specificity	86.0 (85.9, 86.0)	85.3 (85.2, 85.4)	83.2 (83.0, 83.3)	−2.8 (−3.0, −2.6)	<.001	0.80 (0.79, 0.81)	−2.1 (−2.3, −1.9)	<.001	0.87 (0.86, 0.88)
PPV	0.8 (0.8, 0.9)	0.9 (0.9, 1.0)	1.2 (1.1, 1.3)	0.4 (0.3, 0.5)	<.001	1.44 (1.27, 1.64)	0.3 (0.2, 0.4)	<.001	1.28 (1.14, 1.42)
Age 45–49 y									
Sensitivity	53.7 (50.8, 56.6)	54.5 (52.3, 56.7)	67.8 (64.0, 71.4)	14.1 (9.4, 18.8)	<.001	1.85 (1.50, 2.29)	13.3 (8.9, 17.7)	<.001	1.77 (1.46, 2.15)
Specificity	86.2 (86.1, 86.3)	86.0 (85.8, 86.0)	83.8 (83.6, 84.0)	−2.4 (−2.6, −2.2)	<.001	0.81 (0.80, 0.83)	−2.2 (−2.3, −1.9)	<.001	0.88 (0.87, 0.89)
PPV	1.3 (1.2, 1.4)	1.3 (1.3, 1.4)	1.8 (1.6, 2.0)	0.5 (0.3, 0.7)	<.001	1.26 (1.11, 1.44)	0.5 (0.3, 0.7)	<.001	1.29 (1.15, 1.44)
Age 50–59 y									
Sensitivity	61.4 (59.4, 63.4)	62.0 (60.5, 63.5)	75.0 (72.7, 77.3)	13.6 (10.6, 16.6)	<.001	1.80 (1.55, 2.10)	13.0 (10.2, 15.8)	<.001	1.88 (1.64, 2.17)
Specificity	88.8 (88.8, 88.9)	88.9 (88.9, 89.0)	86.7 (86.6, 86.8)	−2.1 (−2.2, −2.0)	<.001	0.84 (0.83, 0.85)	−2.2 (−2.3, −2.1)	<.001	0.88 (0.88, 0.89)
PPV	1.4 (1.3, 1.5)	1.5 (1.4, 1.5)	1.8 (1.7, 1.9)	0.4 (0.3, 0.5)	<.001	1.21 (1.11, 1.31)	0.3 (0.2, 0.4)	<.001	1.22 (1.13, 1.31)
Age 60–69 y									
Sensitivity	72.1 (69.4, 74.7)	72.0 (70.0, 74.0)	77.3 (74.2, 80.3)	5.2 (1.1, 9.3)	.03	1.38 (1.10, 1.72)	5.3 (1.6, 9.0)	.02	1.36 (1.12, 1.66)
Specificity	91.9 (91.8, 91.9)	91.7 (91.7, 91.8)	90.2 (90.1, 90.3)	−1.7 (−1.8, −1.6)	<.001	0.93 (0.91, 0.94)	−1.5 (−1.6, −1.4)	<.001	0.97 (0.95, 0.98)
PPV	1.6 (1.5, 1.7)	1.9 (1.8, 2.0)	2.2 (2.0, 2.4)	0.6 (0.4, 0.8)	<.001	1.41 (1.26, 1.58)	0.3 (0.1, 0.5)	.003	1.19 (1.08, 1.32)
Age 70–79 y									
Sensitivity	74.4 (70.6, 77.9)	76.8 (73.8, 79.6)	86.2 (81.7, 90.0)	11.6 (6.2, 17.4)	<.001	2.01 (1.35, 3.00)	9.4 (4.3, 14.5)	.002	1.97 (1.35, 2.87)
Specificity	94.1 (94.0, 94.2)	93.6 (93.5, 93.7)	92.2 (92.1, 92.4)	−1.9 (−2.1, −1.7)	<.001	0.94 (0.91, 0.96)	−1.4 (−1.6 to −1.2)	<.001	1.00 (0.98, 1.03)
PPV	1.5 (1.4, 1.7)	2.0 (1.8, 2.1)	2.8 (2.5, 3.2)	1.3 (0.9, 1.7)	<.001	1.82 (1.54, 2.15)	0.8 (0.5, 1.3)	<.001	1.50 (1.29, 1.75)
Almost entirely fatty breasts									
Sensitivity	68.3 (65.0, 71.5)	70.0 (67.5, 72.4)	82.4 (77.7, 86.4)	14.1 (8.6, 19.6)	<.001	1.89 (1.35, 2.66)	12.4 (7.3, 17.5)	<.001	1.80 (1.31, 2.48)
Specificity	96.7 (96.6, 96.7)	96.5 (96.5, 96.6)	94.9 (94.8, 95.0)	−1.8 (−1.9, −1.7)	<.001	0.66 (0.64, 0.68)	−1.6 (−1.7, −1.5)	<.001	0.67 (0.65, 0.68)
PPV	2.1 (1.9, 2.2)	2.4 (2.3, 2.6)	2.7 (2.4, 3.1)	0.6 (0.2, 1.0)	.003	1.36 (1.16, 1.58)	0.3 (−0.1, 0.7)	.10	1.14 (0.99, 1.31)
Scattered fibroglandular breast densities									
Sensitivity	69.4 (67.1, 71.7)	70.1 (68.3, 71.8)	77.0 (74.0, 79.8)	7.6 (3.9, 11.3)	<.001	1.39 (1.14, 1.69)	6.9 (3.5, 10.3)	<.001	1.38 (1.15, 1.66)
Specificity	91.0 (91.0, 91.1)	91.8 (91.8, 91.9)	91.0 (90.9, 91.1)	0.0 (−0.1, 0.1)	.999	1.00 (0.98, 1.01)	−0.8 (−0.9, −0.7)	<.001	0.90 (0.89, 0.92)
PPV	1.6 (1.5, 1.7)	1.9 (1.8, 2.0)	2.0 (1.8, 2.1)	0.4 (0.2, 0.6)	<.001	1.22 (1.10, 1.35)	0.1 (−0.1, 0.3)	.26	1.06 (0.97, 1.16)
Heterogeneously dense breasts									
Sensitivity	58.2 (56.1, 60.2)	59.6 (58.1, 61.0)	74.8 (72.7, 76.7)	16.6 (13.7, 19.5)	<.001	2.15 (1.88, 2.47)	15.2 (12.7, 17.7)	<.001	2.00 (1.78, 2.26)
Specificity	86.6 (86.5, 86.7)	84.4 (84.4, 84.5)	82.5 (82.4, 82.6)	−4.1 (−4.2, −4.0)	<.001	0.72 (0.71, 0.73)	−1.9 (−2.0, −1.8)	<.001	0.85 (0.85, 0.86)
PPV	1.2 (1.1, 1.3)	1.2 (1.2, 1.3)	1.7 (1.6, 1.8)	0.5 (0.4, 0.6)	<.001	1.37 (1.26, 1.48)	0.5 (0.4, 0.6)	<.001	1.34 (1.25, 1.43)
Extremely dense breasts									
Sensitivity	52.5 (50.1, 54.8)	53.6 (51.6, 55.6)	62.0 (58.4, 65.5)	9.5 (5.2, 13.8)	<.001	1.50 (1.26, 1.80)	8.4 (4.3, 12.5)	<.001	1.41 (1.19, 1.67)
Specificity	80.7 (80.6, 80.8)	80.3 (80.2, 80.4)	82.4 (82.2, 82.6)	1.7 (1.5, 1.9)	<.001	1.08 (1.07, 1.10)	2.1 (1.9, 2.3)	<.001	1.13 (1.11, 1.14)
PPV	1.0 (0.9, 1.0)	1.0 (1.0, 1.1)	1.6 (1.4, 1.7)	0.6 (0.5, 0.9)	<.001	1.57 (1.40, 1.77)	0.6 (0.4, 0.8)	<.001	1.47 (1.32, 1.64)
Baseline screening									
Sensitivity	62.2 (59.9, 64.4)	64.1 (62.3, 65.8)	76.4 (73.7, 79.1)	14.2 (10.6, 17.8)	<.001	1.99 (1.66, 2.40)	12.3 (9.0, 15.6)	<.001	1.84 (1.56, 2.19)
Specificity	87.3 (87.2, 87.4)	86.3 (86.2, 86.4)	83.9 (83.8, 84.1)	−3.4 (−3.6, −3.2)	<.001	0.78 (0.77, 0.79)	−2.4 (−2.6, −2.2)	<.001	0.87 (0.86, 0.88)
PPV	1.5 (1.4, 1.6)	1.5 (1.4, 1.5)	1.9 (1.7, 2.0)	0.4 (0.2, 0.6)	<.001	1.21 (1.10, 1.34)	0.4 (0.2, 0.6)	<.001	1.20 (1.10, 1.31)
Subsequent screening									
Sensitivity	60.2 (58.8, 61.6)	61.5 (60.5, 62.6)	72.4 (70.7, 74.0)	12.2 (10.0, 14.4)	<.001	1.71 (1.54, 1.90)	10.9 (8.9, 12.9)	<.001	1.65 (1.50, 1.82)
Specificity	90.1 (90.0, 90.1)	89.7 (89.6, 89.7)	87.8 (87.7, 87.8)	−2.3 (−2.4, −2.2)	<.001	0.87 (0.86, 0.88)	−1.9 (−2.0, −1.8)	<.001	0.91 (0.91, 0.92)
PPV	1.2 (1.2, 1.3)	1.4 (1.3, 1.4)	1.8 (1.7, 1.8)	0.6 (0.5, 0.7)	<.001	1.41 (1.33, 1.50)	0.4 (0.3, 0.5)	<.001	1.28 (1.21, 1.35)

Note.—Except where indicated, data are percentages, with 95% confidence intervals in parentheses. aOR = adjusted odds ratio, PPV = positive predictive value.

* Differences between full-field digital mammography and screen-film mammography, based on screen-film mammography; *P* values obtained with a two-proportions test.

† Differences between full-field digital mammography and computed radiography, based on computed radiography; *P* values obtained with a two-proportions test.

‡ Adjusted for screening year, age, breast density, screening round, and health insurance type.

Table 3: AUCs for Screen-Film Mammography, Computed Radiography, and Full-Field Digital Mammography in Korean Women

Characteristic	Screen-Film Mammography			Computed Mammography			Full-Field Digital Mammography	
	AUC	95% CI	*P* Value*	AUC	95% CI	*P* Value*	AUC	95% CI
Total	0.75	0.75, 0.76	<.001	0.76	0.75, 0.76	<.001	0.80	0.80, 0.81
Age (y)								
40–44	0.67	0.66, 0.69	<.001	0.70	0.69, 0.71	<.001	0.74	0.73, 0.76
45–49	0.70	0.69, 0.71	<.001	0.70	0.69, 0.71	<.001	0.76	0.74, 0.78
50–54	0.74	0.73, 0.75	<.001	0.74	0.73, 0.75	<.001	0.79	0.78, 0.81
55–59	0.77	0.75, 0.78	<.001	0.77	0.76, 0.78	<.001	0.84	0.82, 0.86
60–64	0.81	0.79, 0.83	.20	0.81	0.79, 0.82	.11	0.83	0.81, 0.84
65–69	0.84	0.82, 0.86	.17	0.84	0.83, 0.86	.21	0.86	0.84, 0.89
70–74	0.84	0.82, 0.86	<.001	0.84	0.83, 0.86	<.001	0.90	0.88, 0.92
75–79	0.85	0.82, 0.89	.67	0.88	0.86, 0.91	.62	0.87	0.82, 0.92
Breast density								
Almost entirely fat	0.82	0.81, 0.84	<.001	0.83	0.82, 0.84	<.001	0.89	0.87, 0.91
Scattered fibroglandular densities	0.80	0.79, 0.81	<.001	0.81	0.80, 0.82	<.001	0.84	0.83, 0.85
Heterogeneously dense	0.72	0.71, 0.73	<.001	0.72	0.71, 0.73	<.001	0.79	0.78, 0.80
Extremely dense	0.67	0.65, 0.68	<.001	0.67	0.66, 0.68	<.001	0.72	0.70, 0.74
Screening round								
Baseline screening	0.75	0.74, 0.76	<.001	0.75	0.74, 0.76	<.001	0.80	0.79, 0.81
Subsequent screening	0.75	0.74, 0.76	<.001	0.76	0.75, 0.76	<.001	0.80	0.79, 0.81

Note.—AUC = area under the receiver operating characteristic curve, CI = confidence interval.

* *P* values compared with full-field digital mammography.

compared with CR [$P < .001$]) and PPV (adjusted odds ratios for PPV were 1.36 compared with SFM and 1.26 compared with CR [$P < .001$]) but lower specificity (adjusted odds ratios for specificity were 0.85 compared with SFM and 0.90 compared with CR [$P < .001$]) (Table 2). Similar results were seen when the women were dichotomized into age younger than 50 years versus age 50 years or older and with dense versus nondense mammographic breast density (Table E2 [online]). Both absolute and relative differences between these estimates of SFM and FFDM were larger than the differences between CR and FFDM, and $P < .05$ was found for differences among the modalities for the total study population and in most of the subgroups based on age, breast density, and screening round, for all indicators.

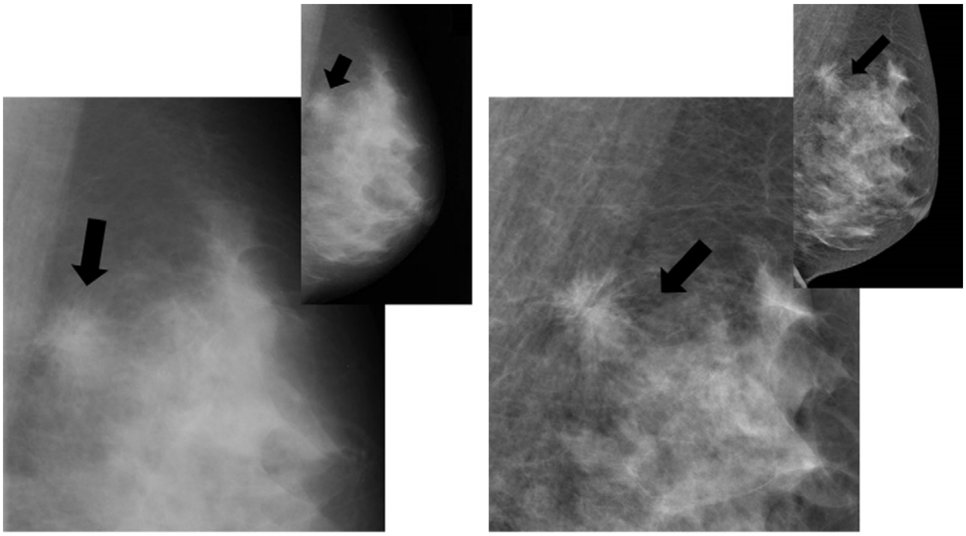
The discriminative performance of each type of mammography equipment, presented as the AUC, also revealed that FFDM had higher diagnostic accuracy than SFM and CR at the population level. Furthermore, FFDM was found to have better performance when compared with SFM and CR, with more remarkable differences in younger women with dense breasts (Table 3, Figs 3, 4). The overall AUCs of SFM, CR, and FFDM were 0.751 (95% CI: 0.745, 0.757), 0.755 (95% CI: 0.751, 0.760), and 0.802 (95% CI: 0.795, 0.809), respectively, with differences between SFM and FFDM ($P < .001$) and CR and FFDM ($P < .001$). Similar results were seen when the women were dichotomized into age younger than 50 years versus age 50 years or older and with dense versus nondense mammographic breast density (Table E3 [online]). The AUCs were similar in a subanalysis of screening units with an average number of screening mammography examinations of 1000 or more per year.

Sensitivity Analysis

In sensitivity analysis (in which 6 900 164 mammograms from 812 screening units with an average of 1000 or more annual screening examinations were analyzed), the overall performance indicators showed better or similar estimates compared with those of original data (marked improvement in SFM; minor differences in CR and FFDM); the overall AUCs of SFM, CR, and FFDM were 0.762 (95% CI: 0.755, 0.769), 0.756 (95% CI: 0.751, 0.761), and 0.803 (95% CI: 0.796, 0.810), respectively. This resulted in a small reduction in the differences between SFM and FFDM, but in general the reduction was trivial (adjusted odds ratios for sensitivity and PPV in FFDM were 1.67 and 1.36, respectively, compared with SFM; for differences in sensitivity, PPV, and AUC, respectively, $P < .001$).

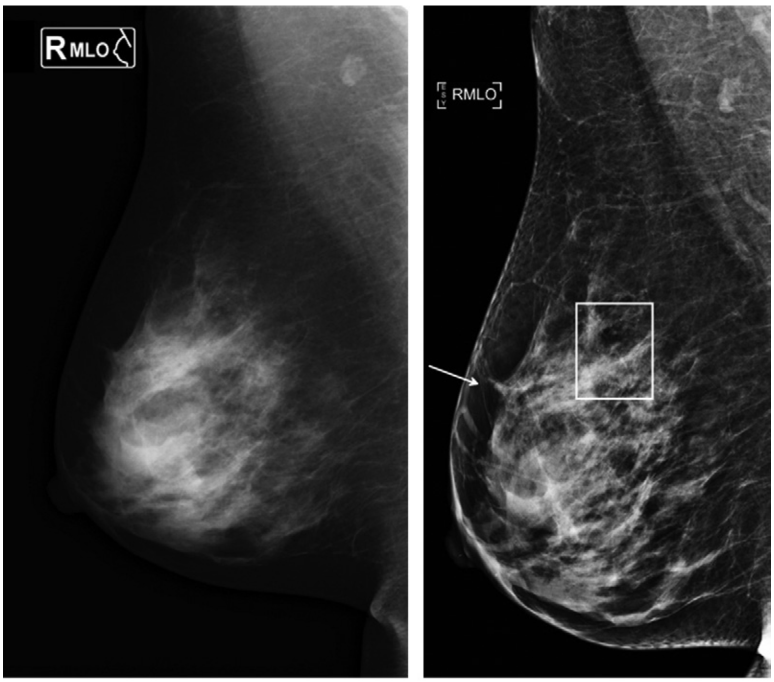
Discussion

In this large-scale retrospective study, 3 years of recent National Cancer Screening Program (NCSP) data on breast cancer screening were used to perform an up-to-date evaluation of the accuracy of full-field digital mammography (FFDM) for breast cancer screening. The study revealed the superiority of FFDM over screen-film mammography (SFM) and computed radiography (CR) in screening accuracy (higher sensitivity, positive predictive value [PPV], and area under the receiver operating characteristic [ROC] curve [AUC]) for the general female population, regardless of age, breast density, or screening round. The results were maintained in the sensitivity analysis, which restricted the included screening units according to the number of screening mammograms performed annually.



a. b.

Figure 3: Mammographic images in 43-year-old woman with hormone receptor–negative and human epidermal growth factor receptor 2–positive invasive ductal carcinoma. Screening mammograms (left mediolateral oblique views) were obtained with **(a)** screen-film mammography (SFM) and **(b)** full-field digital mammography (FFDM) within 1 month. (**a**) In SFM image and a spot compression film, an asymmetry (arrow) in the upper posterior portion of the left breast is inconclusive, whereas (**b**) in FFDM image and a spot compression film, an irregular spiculated mass (arrow) at the same location is clearly visible and highly suggestive of malignancy.



a. b.

Figure 4: Mammographic images in 47-year-old woman with microcalcifications suggestive of malignancy. Screening mammograms (right mediolateral oblique [RMLO] views) were obtained with **(a)** computed radiography (CR) and **(b)** full-field digital mammography (FFDM) within 1 month. (**a**) In CR image, faint microcalcifications in right breast with extremely dense breast tissue (Breast Imaging Reporting and Data System breast composition category D) are inconclusive, whereas (**b**) in FFDM image, grouped calcifications are obviously visible and are suggestive of malignancy.

Our results differ from those of earlier studies (1–13); this could be attributed to three factors. First, because more than a decade has elapsed since the introduction of the newest equipment, the number of screening units and examinations in which FFDM is used has increased significantly, allowing radiologists to advance and stabilize their image-reading skills. Second, FFDM equipment and image quality have improved, which may have positively affected FFDM performance. Third, until recently, FFDM performance could not be precisely evaluated because the distinction between CR and FFDM had not been made. This may have led to underestimated performances of digital mammography systems in previous studies.

The Digital Mammographic Imaging Screening Trial (1), which compared the diagnostic accuracy of SFM and digital mammography, reported better performance of digital mammography compared with SFM only in limited subgroups (young age, dense breasts, and premenopause or perimenopause); it reported no significant difference in the population as a whole. Our study showed higher performance of FFDM across all age groups and breast densities. However, the Digital Mammographic Imaging Screening Trial analysis combined two mammography systems (CR and FFDM) into a single “digital mammography” category; the inferior performance of CR, first reported in a later study (3), may have affected the diagnostic accuracy of that category in the randomized trial. The meta-analysis by Souza et al (14) reported results similar to those of the Digital Mammographic Imaging Screening Trial study. However, it included studies performed in the early 2000s, when image quality was relatively low and radiologists lacked experience in reading FFDM images; this may have offset a difference in SFM and FFDM diagnostic accuracy compared with what has been seen since 2010. In this study, we used more recent mammography data and separately assessed the performance indicators of the three modalities. This allowed us to expand the acceptable range and generalize (ie, without limitations on patient age or breast tissue density) the superiority of the screening accuracy of FFDM when compared with the other two modalities, even though it had a slightly lower specificity (probably because of a higher recall rate). However, when we focused on the subgroups based on patient characteristics, the differences between modalities were much larger in the dense breast and young age groups. Of note, our sensitivity analysis showed similar AUCs for women who were aged 40–44 years and those who were aged 45–49 years. These results are important given the differences in the guidelines for screening mammography that have been published by the ACR and the American Cancer Society (24,25).

Moreover, FFDM revealed more invasive cancer and ductal carcinoma in situ lesions than SFM and CR, even after the original distribution or prevalence of patients with cancer with those modalities was taken into consideration. In addition, a higher proportion of ductal carcinoma in situ lesions was detected with FFDM than with SFM and CR, even after the original relative distribution or ratio between two cancer types (ie, invasive vs ductal carcinoma in situ) for those modalities was taken into consideration. These results were similar in context to those of earlier studies (26–28). This superior accuracy or performance of FFDM can be attributed to technical reasons: The method is simpler, with a digitized process for archiving and transmitting images, which improves logistics and provides better mammographic images with more consistent quality and fewer artifacts at similar dose levels than with other mammography modalities (29–31). In addition, it enables radiologists to alter the window and contrast of the image and to optimize viewing conditions on a digital monitor (30,32).

Our study had some limitations. First, it was an observational study involving records of screening examinations from 3 previous years rather than a prospective randomized trial. Hence, a potential selection bias could exist with respect to the screening modality assigned to each participant. Confounding factors between each mammography modality and the overall screening results reported—for example, the participants' residence, hospital visited, or socioeconomic status—could be associated with the results. Second, manufacturer-related differences in the characteristics and diagnostic performances of FFDM were not reflected in our analysis. In addition, we did not consider the use of computer-assisted detection or diagnosis software to aid in interpretation of mammograms. Third, individual variability among the radiologists who interpreted the mammograms may also have affected our results, and we were unable to account for this factor. Fourth, the number of mammograms in which FFDM was used was small in our study, compared with the number in which SFM and CR were used.

In conclusion, our results indicate a higher screening accuracy of full-field digital mammography (FFDM); we recommend it for the general screening participant, regardless of age or breast tissue density. Today, considerable use of computed radiography and screen-film mammography (SFM) is found in many countries, although FFDM has substantially replaced SFM in some developed countries such as the United States. Therefore, universal application requires consideration of the economic aspects of FFDM, within the context of each country's public health environment.

Acknowledgments: The authors thank the staff of the Korean National Health Insurance System and the Korean Institute for Accreditation of Medical Images.

Author contributions: Guarantors of integrity of entire study, M.S., J.K.J.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, S.H., S.Y.S., B.P., M.J.K., J.K.J.; clinical studies, S.E.J., E.H.L., C.W.L.; statistical analysis, S.H., B.P., M.S., J.K.J.; and manuscript editing, S.H., S.Y.S., K.S.C., S.E.J., M.J.K., E.H.L., J.K.J.

Disclosures of Conflicts of Interest: **S.H.** disclosed no relevant relationships. **S.Y.S.** disclosed no relevant relationships. **B.P.** disclosed no relevant relationships. **M.S.** disclosed no relevant relationships. **K.S.C.** disclosed no relevant relationships. **S.E.J.** disclosed no relevant relationships. **M.J.K.** disclosed no relevant relationships. **E.H.L.** disclosed no relevant relationships. **C.W.L.** disclosed no relevant relationships. **J.K.J.** disclosed no relevant relationships.

References

1. Pisano ED, Gatsonis C, Hendrick E, et al. Diagnostic performance of digital versus film mammography for breast-cancer screening. N Engl J Med 2005;353(17):1773–1783.
2. Kerlikowske K, Hubbard RA, Miglioretti DL, et al. Comparative effectiveness of digital versus film-screen mammography in community practice in the United States: a cohort study. Ann Intern Med 2011;155(8):493–502.
3. Prummel MV, Muradali D, Shumak R, et al. Digital compared with screen-film mammography: measures of diagnostic accuracy among women screened in the Ontario Breast Screening Program. Radiology 2016;278(2):365–373.
4. de Munck L, de Bock GH, Otter R, et al. Digital vs screen-film mammography in population-based breast cancer screening: performance indicators and tumour characteristics of screen-detected and interval cancers. Br J Cancer 2016;115(5):517–524.
5. Sala M, Domingo L, Macià F, Comas M, Burón A, Castells X. Does digital mammography suppose an advance in early diagnosis? Trends in performance indicators 6 years after digitalization. Eur Radiol 2015;25(3):850–859.
6. Nederend J, Duijm LE, Louwman MW, et al. Impact of the transition from screen-film to digital screening mammography on interval cancer characteristics and treatment - a population based study from the Netherlands. Eur J Cancer 2014;50(1):31–39.
7. Chiarelli AM, Edwards SA, Prummel MV, et al. Digital compared with screen-film mammography: performance measures in concurrent cohorts within an organized breast screening program. Radiology 2013;268(3):684–693.
8. Juel IM, Skaane P, Hoff SR, Johannessen G, Hofvind S. Screen-film mammography versus full-field digital mammography in a population-based screening program: The Sogn and Fjordane study. Acta Radiol 2010;51(9):962–968.
9. 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(2):347–358.
10. Skaane P, Hofvind S, Skjennald A. Randomized trial of screen-film versus full-field digital mammography with soft-copy reading in population-based screening program: follow-up and final results of Oslo II study. Radiology 2007;244(3):708–717.
11. Skaane P, Skjennald A, Young K, et al. Follow-up and final results of the Oslo I Study comparing screen-film mammography and full-field digital mammography with soft-copy reading. Acta Radiol 2005;46(7):679–689.
12. Lewin JM, D'Orsi CJ, Hendrick RE, et al. Clinical comparison of full-field digital mammography and screen-film mammography for detection of breast cancer. AJR Am J Roentgenol 2002;179(3):671–677.
13. Heddson B, Rönnow K, Olsson M, Miller D. Digital versus screen-film mammography: a retrospective comparison in a population-based screening program. Eur J Radiol 2007;64(3):419–425.
14. Souza FH, Wendland EM, Rosa MI, Polanczyk CA. Is full-field digital mammography more accurate than screen-film mammography in overall population screening? A systematic review and meta-analysis. Breast 2013;22(3):217–224.
15. Siu AL; U.S. Preventive Services Task Force. Screening for Breast Cancer: U.S. Preventive Services Task Force Recommendation Statement. Ann Intern Med 2016;164(4):279–296.
16. Yaffe MJ, Bloomquist AK, Hunter DM, et al. Comparative performance of modern digital mammography systems in a large breast screening program. Med Phys 2013;40(12):121915.
17. Pisano ED. Digital Compared with Screen-Film Mammography: Measures of Diagnostic Accuracy among Women Screened in the Ontario Breast Screening Program-Evidence that Direct Radiography Is Superior to Computed Radiography for Cancer Detection. Radiology 2016;278(2):311–312.
18. Suh M, Song S, Cho HN, et al. Trends in Participation Rates for the National Cancer Screening Program in Korea, 2002-2012. Cancer Res Treat 2017;49(3):798–806.
19. Kim YS, Jung SE, Choi BG, et al. Image quality improvement after implementation of a CT accreditation program. Korean J Radiol 2010;11(5):553–559.
20. D'Orsi CJ, Bassett LW, Berg WA, et al. ACR BI-RADS Mammography. In: ACR BI-RADS Atlas, Breast Imaging Reporting and Data System. 4th ed. Reston, Va: American College of Radiology, 2003.
21. Jung KW, Won YJ, Kong HJ, Lee ES; Community of Population-Based Regional Cancer Registries. Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2015. Cancer Res Treat 2018;50(2):303–316.
22. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12(1):77.

23. Ministry of Health and Welfare, National Cancer Center. Quality Guidelines of Breast cancer Screening. 2nd rev. Seoul, Korea: Ministry of Health and Welfare, National Cancer Center, 2018.
24. Monticciolo DL, Newell MS, Hendrick RE, et al. Breast cancer screening for average-risk women: recommendations from the ACR Commission on Breast Imaging. J Am Coll Radiol 2017;14(9):1137–1143.
25. Oeffinger KC, Fontham ET, Etzioni R, et al. Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. JAMA 2015;314(15):1599–1614.
26. Séradour B, Heid P, Estève J. Comparison of direct digital mammography, computed radiography, and film-screen in the French national breast cancer screening program. AJR Am J Roentgenol 2014;202(1):229–236.
27. van Luijt PA, Fracheboud J, Heijnsdijk EA, den Heeten GJ, de Koning HJ; National Evaluation Team for Breast Cancer Screening in Netherlands Study Group (NETB). Nation-wide data on screening performance during the transition to digital mammography: observations in 6 million screens. Eur J Cancer 2013;49(16):3517–3525.
28. Bluekens AM, Holland R, Karssemeijer N, Broeders MJ, den Heeten GJ. Comparison of digital screening mammography and screen-film mammography in the early detection of clinically relevant cancers: a multicenter study. Radiology 2012;265(3):707–714.
29. Hofvind S, Skaane P, Elmore JG, Sebuødegård S, Hoff SR, Lee CI. Mammographic performance in a population-based screening program: before, during, and after the transition from screen-film to full-field digital mammography. Radiology 2014;272(1):52–62.
30. Sala M, Comas M, Macià F, Martinez J, Casamitjana M, Castells X. Implementation of digital mammography in a population-based breast cancer screening program: effect of screening round on recall rate and cancer detection. Radiology 2009;252(1):31–39.
31. Berns EA, Hendrick RE, Cutter GR. Performance comparison of full-field digital mammography to screen-film mammography in clinical practice. Med Phys 2002;29(5):830–834.
32. Iared W, Shigueoka DC, Torloni MR, et al. Comparative evaluation of digital mammography and film mammography: systematic review and meta-analysis. Sao Paulo Med J 2011;129(4):250–260.