The Positive Predictive Value for Diagnosis of Breast Cancer: Full-Field Digital Mammography Versus Film-Screen Mammography in the Diagnostic Mammographic Population

Bo Kyoung Seo, MD, PhD, Etta D. Pisano, MD, FACR, Cherie M. Kuzmiak, DO, Marcia Koomen, MD, Dag Pavic, MD, Robert McLelland, MD, FACR, Yeonhee Lee, MD, PhD, Elodia B. Cole, MS, Dianne Mattingly, RN, Juneyoung Lee, PhD

Rationale and Objectives. Diagnostic mammography is performed on women with clinical symptoms that suggest breast cancer or women for whom further mammographic evaluation has been requested because of an abnormal screening mammography. We assessed whether the use of full-field digital mammography would improve the positive predictive value (PPV) for the diagnosis of breast cancer in a diagnostic population compared with film-screen mammography.

Materials and Methods. From January 2002 to December 2003, 11,621 patients underwent diagnostic mammography at the University of North Carolina Hospital, Chapel Hill. Among these 11,621 patients, 1400 lesions in 1121 patients underwent biopsy. We included the biopsy-performed lesions, so PPV3 was used for comparison of PPVs between film-screen mammography and full-field digital mammography. Six breast radiologists interpreted the images using the Breast Imaging Reporting and Data System of the American College of Radiology. PPV3s were compared between film-screen and full-field digital mammography in the entire study cohort and in specified subgroups according to different radiologists, breast density, and lesion type on mammography. The χ2 and Fisher’s exact tests were used for comparison of PPV3s between two modalities of mammography with the Bonferroni procedure for subgroup analysis.

Results. In the entire study cohort, PPV3s of full-field digital mammography and film-screen mammography were similar (difference in PPV3, −0.007; 95% confidence interval, −0.081 to 0.068; P = .8602). In predefined subgroups, there was no difference in PPV3 by the radiologist, breast density, or lesion type between two modalities of mammography (P > .005).

Conclusion. There is no improvement in PPV for the diagnosis of breast cancer with full-field digital mammography compared with film-screen mammography in a large diagnostic population.

Key Words. Mammography; breast neoplasms; predictive value of tests; comparative study.

© AUR, 2006

Mammography has been used successfully as a screening test for breast cancer over the past 20 to 30 years. Randomized controlled trials of mammographic screening for breast cancer demonstrate that screening reduces mortality from the disease by 20% to 30% (1). However, film-screen mammography (FSM) has substantial limitations. Approximately 10% to 20% of palpable breast cancers are not visible on images obtained with this modality, mainly due to insufficient contrast between normal and abnormal breast tissue (2–4). In FSM, the film serves simultaneously as the image receptor, display medium, and long-term stor-
age system. This can lead to loss of image contrast, especially when exposure or film-processing conditions cause lower optical densities in lesion-containing tissues (5, 6). In addition, because image acquisition factors are often interdependent in FSM, the optimization of one factor often comes at the expense of another.

On the other hand, full-field digital mammography (FFDM) systems separate the processes of image acquisition, processing, and display (7). Because image acquisition and display are separated, each can be optimized. Digital detectors have a linear response to x-ray intensity in contrast to the sigmoidal response of FSM. Thus, digital detectors provide a broader dynamic range of densities and higher contrast resolution (7). Some reports suggested that FFDM may be superior to FSM in visualization of small masses and microcalcifications caused by improving contrast between lesions and normal background tissue (8–12).

Comparative study in the diagnostic accuracy for breast cancer between FSM and FFDM has been reported in large screening populations (13–16). Although FFDM has many physical advantages, the positive predictive values (PPVs) for the diagnosis of breast cancer in screening population have not shown a significant difference between FSM and FFDM (13–16). Screening mammography is performed on asymptomatic women to detect early clinically unsuspected breast cancer (17). On the other hand, diagnostic mammography examinations are performed on women with clinical signs or symptoms that suggest breast cancer (17). Our study was designed to compare the PPV for the diagnosis of breast cancer between FSM and FFDM in patients who underwent diagnostic mammography in a high-volume breast cancer referral clinic.

The PPV for the diagnosis of breast cancer is defined as the percentage of mammographic examinations with positive findings that lead to a diagnosis of breast cancer at pathology within 1 year (17). There are three separate definitions of PPV at a mammography audit using the Breast Imaging Reporting and Data System (BI-RADS) of the American College of Radiology (17). PPV1 is the percentage of all positive screening examinations (BI-RADS categories 0, 4, and 5) that result in a tissue diagnosis of cancer within 1 year. PPV2 is the percentage of all screening or diagnostic examinations recommended for biopsy or surgical consultation (BI-RADS categories 4 and 5) that result in a tissue diagnosis of cancer within 1 year. PPV3 is the percentage of all known biopsies performed as a result of positive screening or diagnostic examinations or additional imaging evaluations of positive examination (BI-RADS categories 4 and 5) that result in a tissue diagnosis of cancer within 1 year. Thus, PPV1 is based on positive findings at screening examination, PPV2 is based on recommendation for biopsy or surgical consultation, and PPV3 is based on results of biopsy. PPV3 is also known as the biopsy yield of malignancy or the positive biopsy rate. In this study, we included the biopsy-performed lesions, thus, PPV3 was used for comparison of PPV between FSM and FFDM in a diagnostic population. We compared the PPV3s according to different radiologists, breast densities, and lesion types on FSM and FFDM over a 12-month period.

**MATERIALS AND METHODS**

**Image Production**

The institutional review board at University of North Carolina, Chapel Hill, approved the study. Diagnostic mammography examination is performed on a woman with clinical signs or symptoms that suggest breast cancer and a woman for whom further mammographic evaluation has been requested because of an abnormal screening mammographic examination (17). Diagnostic mammography is also performed for special screening examination in a woman with a personal history of breast cancer treated with breast conservations and in a woman with breast augmentation (17). From January 2002 to December 2003, 11,621 patients underwent diagnostic mammography at the University of North Carolina Hospital, Chapel Hill. Among these 11,621 patients, the consecutive 1121 patients who underwent diagnostic mammography in a high-volume breast cancer referral clinic.

The PPV for the diagnosis of breast cancer is defined as the percentage of mammographic examinations with positive findings that lead to a diagnosis of breast cancer at pathology within 1 year (17). There are three separate definitions of PPV at a mammography audit using the Breast Imaging Reporting and Data System (BI-RADS) of the American College of Radiology (17). PPV1 is the percentage of all positive screening examinations (BI-RADS categories 0, 4, and 5) that result in a tissue diagnosis of cancer within 1 year. PPV2 is the percentage of all screening or diagnostic examinations recommended for biopsy or surgical consultation (BI-RADS categories 4 and 5) that result in a tissue diagnosis of cancer within 1 year. PPV3 is the percentage of all known biopsies performed as a result of positive screening or diagnostic ex-
Data Collection

FFDM was performed with a General Electric Seno- graph 2000D (GE Medical Systems, Milwaukee, WI). The digital detector of this machine consists of an amorphous silicon transistor/photodiode array deposited as a thin film over a large-area glass substrate. A cesium iodide phosphor layer above the silicon layer is used to convert X-ray photons to light photons, which are then used to create a signal in the amorphous silicon. The parameters of the machine are a detector element size of 100 μm, matrix size of 1900 × 2304, field of view of 19 × 23 cm, and 14-bit gray resolution. The digital images were evaluated in soft-copy and printed-film display. The soft copy was displayed on a prototype two-monitor workstation supplied with the unit. The monitor resolution was 2000 × 2500 pixels. The digital images were also printed on film (Kodak EktaScan HN; Eastman Kodak, Rochester, NY) with a laser film printer (Kodak 2180 EktaScan; Eastman Kodak). Ongoing quality control for the FFDM systems included weekly phantom imaging and flat-field calibration of the detectors.

Data Analysis

The statistical analysis of the data was supervised by a statistician (J.L.). The PPV3 was defined as the percentage of diagnostic examinations with positive findings that led to a diagnosis of breast cancer within 1 year. PPV3s of the two modalities of mammography were computed on the basis of the BI-RADS categories. The BI-RADS categories were dichotomized as negative (cate-
category of 1, 2, or 3) and positive (category of 4 or 5). The \( \chi^2 \) and Fisher’s exact tests were used to compare PPV3s between FSM and FFDM (SAS/STAT software, version 6.12; SAS Institute, Cary, NC).

The PPV3s were compared between FSM and FFDM in the entire study cohort as well as in prespecified subgroups of participants. For analysis of subgroups, breast density was dichotomized as “dense” (extremely dense or heterogeneously dense) and “less dense” (scattered fibroglandular densities, or almost completely fat), and lesion type was dichotomized as “mass” (mass, asymmetric density, or architectural distortion) and “calcification.” The subgroups were defined according to radiologists (six radiologists), breast density (dense versus less dense), and lesion type (mass versus calcification) on mammography. The Bonferroni procedure was used to account for the 10 multiple comparisons in the subgroup analysis, with \( P \leq .005 \) considered to indicate statistical significance.

### RESULTS

Of the 1400 lesions included in this study, 731 were classified as BI-RADS category 4 or 5 on mammography, indicating a “positive” mammographic finding for calculation of the PPV3. Table 1 demonstrates the pathologic diagnoses of the 731 lesions. The pathologic diagnoses of the lesions detected by the two types of mammography were similar. Of the 731 lesions that were recommended for biopsy, 349 (48%) proved to be malignant. Of the 349 malignant lesions, 172 (49%) lesions were invasive carcinoma and DCIS, 109 (31%) lesions were only invasive carcinoma, and the remaining 68 (20%) lesions were only DCIS. There was no difference in tumor subtype for invasive carcinomas and DCIS cases between FFDM and FSM (\( P > .005 \)).

There were 281 lesions classified as BI-RADS 4 or 5 by FFDM, and of these, 133 lesions were malignant. There were 450 lesions classified as BI-RADS 4 or 5 by FSM, and 216 of these were malignant. Therefore, PPV3 of FFDM and FSM was similar, as reflected by a mean (±SE) of 0.47 ± 0.03 for FFDM and of 0.48 ± 0.02 for FSM (difference in PPV3, −0.007; 95% confidence interval, −0.081 to 0.068; \( P = .8602 \)) in the entire study cohort.

Table 2 demonstrates a comparison of the PPV3 between FFDM and FSM by interpreting radiologist, breast density, and lesion type. Three of six radiologists (radiologists 1, 2, and 4) showed higher PPV3 for FFDM, and three radiologists (radiologists 3, 5, and 6) demonstrated higher PPV3 for FSM. These results were not significantly different (\( P > .005 \)).

For comparison of PPV3 between FFDM and FSM according to breast density, PPV3 was evaluated for dense and less dense breasts. For dense breasts, the PPV3 for FFDM (0.51 ± 0.04) was better than that for FSM (0.45 ± 0.03), and the difference was 0.051 (95% confidence interval, −0.046 to 0.148). On the other hand, for less dense breasts, the PPV3 for FSM (0.51 ± 0.03) was superior to PPV3 for FFDM (0.42 ± 0.05), and the difference was −0.091 (95% confidence interval, −0.207 to 0.026). These results were not significantly different (\( P > .005 \)).

Of the 731 lesions recommended for biopsy, 463 (63%) presented as masses on mammography, and the remaining 270 (37%) lesions were calcifications. For masses, the PPV3 was no different between the two modalities of mammography (PPV3 for FFDM, 0.59 ± 0.04; PPV3 for FSM, 0.59 ± 0.03; difference, −0.006; 95% confidence interval, −0.099 to 0.087; \( P > .005 \)). For cal-

### Table 1

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>FFDM (n = 281)</th>
<th>FSM (n = 450)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign lesion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCC</td>
<td>87/148 (59)</td>
<td>135/234 (58)</td>
</tr>
<tr>
<td>ADH</td>
<td>14/148 (9)</td>
<td>16/234 (7)</td>
</tr>
<tr>
<td>FA</td>
<td>15/148 (10)</td>
<td>27/234 (12)</td>
</tr>
<tr>
<td>LCIS</td>
<td>3/148 (2)</td>
<td>2/234 (1)</td>
</tr>
<tr>
<td>Others</td>
<td>29/148 (20)</td>
<td>54/234 (23)</td>
</tr>
<tr>
<td>Malignant lesion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invasive carcinoma ± DCIS</td>
<td>105/133 (79)</td>
<td>176/216 (81)</td>
</tr>
<tr>
<td>DCIS</td>
<td>28/133 (21)</td>
<td>40/216 (19)</td>
</tr>
<tr>
<td>Invasive carcinoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ductal carcinoma</td>
<td>80/105 (76)</td>
<td>137/176 (78)</td>
</tr>
<tr>
<td>Mixed ductal and lobular carcinoma</td>
<td>15/105 (14)</td>
<td>27/176 (15)</td>
</tr>
<tr>
<td>Lobular carcinoma</td>
<td>8/105 (8)</td>
<td>4/176 (2)</td>
</tr>
<tr>
<td>Others</td>
<td>2/105 (2)</td>
<td>8/176 (5)</td>
</tr>
<tr>
<td>DCIS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comedo DCIS</td>
<td>28/85 (33)</td>
<td>52/155 (34)</td>
</tr>
<tr>
<td>Noncomedo DCIS</td>
<td>57/85 (67)</td>
<td>103/155 (66)</td>
</tr>
</tbody>
</table>

Numbers in parentheses are percentages.
cifications, the PPV3 was also not significantly different between FFDM and FSM (PPV3 for FFDM, 0.29 ± 0.04; PPV3 for FSM, 0.28 ± 0.04; difference, 0.008; 95% confidence interval, 0.103 to 0.118; P = .005).

**DISCUSSION**

In this study, we compared PPV for FFDM and FSM in a diagnostic population. The use of screening populations results in a relatively low number of cancers, typically 2 to 10 per 1000 women screened, depending on the proportion of first-time screens (22). This low cancer rate decreases the power to detect a difference between the modalities. Thus, we used diagnostic population for comparison of PPV between to modalities of mammography in the current study. PPV in diagnostic population is meaningful for reducing unnecessary biopsy and surgery, medical costs, and patients’ emotional stress. There are three separate definition of PPV at a mammography audit using BI-RADS (17). We used PPV3, based on the results of biopsy in this study. In the entire patient cohort of this study, PPV3 for FFDM was 0.47 and PPV3 for FSM was 0.48. PPV has been reported in large screening populations (13–16). The study by Pisano et al. (16) demonstrated that PPV1 in the large screening cohort, 42,760 women, was 0.12 for FFDM and 0.13 for FSM. Skaaane et al. (14) demonstrated that PPV2 in a large screening population was 0.39 for FFDM and 0.46 for FSM. Lewin et al. (13) showed that PPV3 in a large screening population was 0.30 for FFDM and 0.19 for FSM. Thus, PPV in the current study was higher than that in previous reports using screening populations (13–16). The higher PPV in the current study might be caused by a higher cancer detection rate for diagnostic mammography than for screening mammography. This is the first published report of the comparative PPV of digital mammography in a diagnostic mammography population.

In this study, we found there was no difference in PPV3 between two modalities of mammography in either the entire study cohort or predefined subgroups. FFDM offers physical advantages over FSM: separate of the processes of image acquisition, processing, and display; a broader dynamic range of densities; higher contrast resolution of digital detector; and image processing algorithms (23–26). Although FFDM has these advantages, FFDM did not improve PPV for the detection of breast cancer in a diagnostic population. In previous reports utilizing a screening population, there was no difference in PPV between FFDM and FSM (13–16). Lewin et al. (13) reported that the PPV3 for FFDM (30%; 21 of 69) was higher than that for FSM (19%; 22 of 114), but this difference was not statistically significant.

PPV3 was compared according to different six radiologists in the current study. Although the six radiologists were experts in breast imaging, PPV3s for FSM and FFDM varied. The variability in PPV3 was higher for

**Table 2**

<table>
<thead>
<tr>
<th>Radiologist</th>
<th>Full-Field Digital Mammography* (n = 281)</th>
<th>Film-Screen Mammography* (n = 450)</th>
<th>Difference† (95% CI)</th>
<th>P Value‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast density</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dense</td>
<td>21/45 (0.47 ± 0.07)</td>
<td>24/58 (0.41 ± 0.06)</td>
<td>0.053 (−0.140, 0.246)</td>
<td>0.5915</td>
</tr>
<tr>
<td>Less dense</td>
<td>8/14 (0.57 ± 0.13)</td>
<td>12/24 (0.50 ± 0.10)</td>
<td>0.071 (−0.256, 0.399)</td>
<td>0.6706</td>
</tr>
<tr>
<td>Mass</td>
<td>26/63 (0.41 ± 0.06)</td>
<td>61/128 (0.48 ± 0.04)</td>
<td>−0.064 (−0.213, 0.085)</td>
<td>0.4047</td>
</tr>
<tr>
<td>Calcification</td>
<td>56/97 (0.58 ± 0.05)</td>
<td>60/134 (0.45 ± 0.04)</td>
<td>0.130 (0.0001, 0.2590)</td>
<td>0.0519</td>
</tr>
<tr>
<td>Lesion type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dense</td>
<td>20/57 (0.35 ± 0.06)</td>
<td>54/97 (0.56 ± 0.05)</td>
<td>−0.206 (−0.364, 0.047)</td>
<td>0.0136</td>
</tr>
<tr>
<td>Less dense</td>
<td>2/5 (0.40 ± 0.22)</td>
<td>5/9 (0.56 ± 0.17)</td>
<td>−0.156 (−0.694, 0.383)</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

*The values are mean ± SE.
†The difference was obtained by subtracting the value for film-screen mammography from the value for full-field digital mammography.
‡P value was calculated by χ² or Fisher’s exact test.

The use of screening populations results in a relatively low number of cancers, typically 2 to 10 per 1000 women screened, depending on the proportion of first-time screens. This low cancer rate decreases the power to detect a difference between the modalities. Thus, we used diagnostic population for comparison of PPV between two modalities of mammography in the current study. PPV in diagnostic population is meaningful for reducing unnecessary biopsy and surgery, medical costs, and patients’ emotional stress. There are three separate definition of PPV at a mammography audit using BI-RADS. We used PPV3, based on the results of biopsy in this study. In the entire patient cohort of this study, PPV3 for FFDM was 0.47 and PPV3 for FSM was 0.48. PPV has been reported in large screening populations. The study by Pisano et al. demonstrated that PPV1 in the large screening cohort, 42,760 women, was 0.12 for FFDM and 0.13 for FSM. Skaaane et al. demonstrated that PPV2 in a large screening population was 0.39 for FFDM and 0.46 for FSM. Lewin et al. showed that PPV3 in a large screening population was 0.30 for FFDM and 0.19 for FSM. Thus, PPV in the current study was higher than that in previous reports using screening populations. The higher PPV in the current study might be caused by a higher cancer detection rate for diagnostic mammography than for screening mammography. This is the first published report of the comparative PPV of digital mammography in a diagnostic mammography population.
FFDM than for FSM. The PPV3 for FFDM ranged from 0.35 to 0.58, and the PPV3 for FSM ranged from 0.41 to 0.56. Variability of PPV3 according to the radiologists may be related to the limitations of this study. In this study, we included consecutive patients who underwent diagnostic mammography and breast biopsy in a single facility; thus, each patient underwent only one type of mammography, either FSM or FFDM, and each lesion was evaluated by only one radiologist. Therefore, the PPV3 of each radiologist can be influenced by variation of breast density and lesion type for each case. In addition, the radiologists had far more advanced experience with FSM than with FFDM.

Breast density on mammography is influenced by many factors, such as obesity, ethnicity, age, stage of menstrual cycle, and parity. Breast density is a risk factor for missed cancers, and both false-positive and false-negative mammographic interpretations are more likely with dense breasts (5, 6). Digital detectors can improve lesion detection in dense breasts by higher contrast compared with film-screen system (7). In this study, PPV3 for FFDM was superior to FSM in dense breasts, but this was not significantly different. However, there are some reports suggesting that FFDM may improve lesion detection in dense breasts using variable image processing algorithms and advanced techniques such as tomosynthesis and digital subtraction mammography (23–26). Therefore, further study is warranted to evaluate the potential advantage of FFDM in dense breasts in a diagnostic population.

In the current study, we also compared PPV3 according to lesion type, mass versus calcification. Some reports have suggested that FFDM is superior to FSM for the detection and characterization of small masses and microcalcifications due to improved contrast between lesions and normal background tissue (8–12). These studies were performed in relatively small patient cohorts. We compared PPV3 between the two types of mammography in a relatively large patient cohort. Even so, the PPV3s we measured were not different for either masses or calcifications between the two types of mammography. In this study, the PPV3s were higher for masses (0.59 for FFDM and FSM) than for calcifications (0.29 for FFDM and 0.28 for FSM). Liberman et al. (27) calculated PPV in nonpalpable lesions with FSM according to lesion type and demonstrated higher PPV for masses (0.55) than for calcifications (0.42). Our results are concordant with theirs.

There are some limitations to this study. First, each patient was imaged using only one type of mammography, FFDM or FSM. The two populations using FFDM and FSM were different. A randomized prospective design would have yielded more credible results. However, there was no difference in pathology for malignant and benign lesions between FFDM and FSM in our study, suggesting the populations of FFDM and FSM were similar. Another limitation is each woman was evaluated by only one radiologist. Even though the radiologists included in this study were experts in breast imaging, interreader variability in the determination of breast density and BI-RADS categories was present in our study. Again, randomization of readers would have been a way to avoid this problem.

In conclusion, the PPV for the diagnosis of breast cancer for FFDM is similar to that for FSM in the diagnostic population studied here. Additionally, PPV did not vary between FFDM and FSM according to breast density or lesion type.

REFERENCES


