ORIGINAL ARTICLE

Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening

Etta D. Pisano, M.D., Constantine Gatsonis, Ph.D., Edward Hendrick, Ph.D., Martin Yaffe, Ph.D., Janet K. Baum, M.D., Suddhasatta Acharyya, Ph.D., Emily F. Conant, M.D., Laurie L. Fajardo, M.D., Lawrence Bassett, M.D., Carl D'Orsi, M.D., Roberta Jong, M.D., and Murray Rebner, M.D., for the Digital Mammographic Imaging Screening Trial (DMIST) Investigators Group*

ABSTRACT

BACKGROUND

Film mammography has limited sensitivity for the detection of breast cancer in women with radiographically dense breasts. We assessed whether the use of digital mammography would avoid some of these limitations.

METHODS

A total of 49,528 asymptomatic women presenting for screening mammography at 33 sites in the United States and Canada underwent both digital and film mammography. All relevant information was available for 42,760 of these women (86.3 percent). Mammograms were interpreted independently by two radiologists. Breast-cancer status was ascertained on the basis of a breast biopsy done within 15 months after study entry or a follow-up mammogram obtained at least 10 months after study entry. Receiver-operating-characteristic (ROC) analysis was used to evaluate the results.

RESULTS

In the entire population, the diagnostic accuracy of digital and film mammography was similar (difference between methods in the area under the ROC curve, 0.03; 95 percent confidence interval, −0.02 to 0.08; P=0.18). However, the accuracy of digital mammography was significantly higher than that of film mammography among women under the age of 50 years (difference in the area under the curve, 0.15; 95 percent confidence interval, 0.05 to 0.25; P=0.002), women with heterogeneously dense or extremely dense breasts on mammography (difference, 0.11; 95 percent confidence interval, 0.04 to 0.18; P=0.003), and premenopausal or perimenopausal women (difference, 0.15; 95 percent confidence interval, 0.05 to 0.24; P=0.002).

CONCLUSIONS

The overall diagnostic accuracy of digital and film mammography as a means of screening for breast cancer is similar, but digital mammography is more accurate in women under the age of 50 years, women with radiographically dense breasts, and premenopausal or perimenopausal women. (clinicaltrials.gov number, NCT00008346.)

From the Departments of Radiology and Biomedical Engineering, the Biomedical Research Imaging Center, and the Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill (E.D.P.); the Center for Statistical Sciences, Brown University, Providence, R.I. (C.G., S.A.); the Department of Radiology, Feinberg School of Medicine, Northwestern University, Chicago (E.H.); the Departments of Medical Imaging (M.Y., R.J.) and Medical Biophysics (M.Y.), University of Toronto, Toronto; the Department of Radiology, Beth Israel Deaconess Medical Center, Boston (J.K.B.); the Department of Radiology, University of Pennsylvania Medical School, Philadelphia (E.F.C.); the Department of Radiology, University of Iowa, Iowa City (L.L.F.); the Department of Radiology, University of California at Los Angeles, Los Angeles (L.B.); the Department of Radiology, Emory University, Atlanta (C.D.); and the Department of Radiology, William Beaumont Hospital, Royal Oak, Mich. (M.R.). Address reprint requests to Dr. Pisano at etta_pisano@med.unc.edu.

*Members of the DMIST Investigators Group are listed in the Appendix.

N Engl J Med 2005;353.

There is now general agreement that screening mammography reduces the rate of death from breast cancer among women who are 40 years of age or older.[1,2] Meta-analyses of eight large, randomized trials found a reduction in the mortality rate of 16 to 35 percent among women 50 to 69 years of age who were assigned to screening mammography,[1] whereas women who were 40 to 49 years of age at entry had a smaller but significant reduction of 15 to 20 percent.[1-3]

The smaller benefit of screening in younger women is probably due to a lower incidence of breast cancer, more rapidly growing tumors, and greater radiographic density of breast tissue in women less than 50 years of age.[4] Greater density reduces the sensitivity of mammography[5,6] and increases the risk of breast cancer.[7-9] Digital mammography, which was developed in part to address some of the limitations of film mammography,[10] separates image acquisition and display, allowing the optimization of both. Image processing of digital data allows the degree of contrast in the image to be manipulated, so that contrast can be increased in the dense areas of the breast with the lowest contrast.[11,12]

Despite these apparent differences between the two approaches, previous trials have not found digital mammography to be significantly more accurate than film mammography in the diagnosis of breast cancer.[13-17] These studies were limited, however, in that they included only one type of digital detector and had insufficient statistical power to identify relatively small differences in diagnostic accuracy. The Digital Mammographic Imaging Screening Trial (DMIST) was designed to measure relatively small but potentially clinically important differences in diagnostic accuracy between digital and film mammography.

METHODS

A detailed account of the design of DMIST has been published previously.[18] This trial was conducted by the American College of Radiology Imaging Network. During a two-year period, 49,528 women were recruited to the study at 33 sites. The protocol was approved by the institutional review boards at all sites. All women gave written informed consent. The study was monitored by a data and safety monitoring board. Women who presented for screening mammography at the study sites were eligible to participate unless they reported symptoms, had breast implants, believed they might be pregnant, had undergone mammography for any purpose within the preceding 11 months, or had a history of breast cancer treated with both lumpectomy and radiation.

All participants underwent both digital and film mammography in random order. Five digital-mammography systems were used: the SenoScan (Fischer Medical), the Computed Radiography System for Mammography (Fuji Medical), the Senographe 2000D (General Electric Medical Systems), the Digital Mammography System (Hologic), and the Selenia Full Field Digital Mammography System (Hologic).[18]

The digital and film examinations for each woman were independently interpreted by two radiologists, one reader for each examination. Readers rated the mammograms using a seven-point malignancy scale suitable for receiver-operating-characteristic (ROC) analysis and the classification of the Breast Imaging Reporting and Data System (BIRADS)[19] and recorded whether they recommended that additional tests be performed. A score of 1 on the seven-point malignancy scale indicates a result that is definitely not malignant, a score of 2 findings that are almost definitely not malignant, a score of 3 findings that are probably not malignant, a score of 4 findings that may be malignant, a score of 5 findings that are probably malignant, a score of 6 findings that are almost definitely malignant, and a score of 7 findings that are definitely malignant. A BIRADS score of 0 indicates incomplete data, a score of 1 negative results, a score of 2 benign findings, a score of 3 findings that are probably benign, a score of 4 the presence of a suspicious-appearing abnormality, and a score of 5 findings highly suggestive of cancer.

Readers also rated breast density according to the standard BIRADS scale (extremely dense, heterogeneously dense, scattered fibroglandular densities, and almost completely fat). Radiologists in the United States were all qualified interpreters of mammograms under federal law. Canadian readers met equivalent standards. Each site's lead radiologist was trained in the use of the malignancy scale and trained the site's other readers.

A workup, including a biopsy or aspiration of the suspicious-appearing lesion, was performed if either reader recommended it. A single pathologist or the principal investigator of the study coded all pathological diagnoses on the basis of a review of

the cytologic or histologic material or of the local pathology report. All participants were asked to return for a follow-up mammogram at one year.

To establish a reference standard, participants were classified as positive for cancer if breast cancer was pathologically verified within 455 days after the initial study mammogram and negative for cancer if their study records showed negative findings on a pathology report of a biopsy specimen, if the follow-up mammogram at 1 year was normal, or if both criteria were met. The 455-day period gave women more than a year after study entry to undergo follow-up mammography. Some analyses were repeated with the use of an additional reference standard based on information from the first 365 days after initial mammography, an interval used in other publications on screening mammography.[5,6,20-26] The status of participants who were classified as neither positive nor negative for cancer was considered indeterminate if they had a breast biopsy with indeterminate results (owing to insufficient material or an inability to interpret the results); had a follow-up mammogram with a BIRADS score[19] of 3, 4, or 5; or died during the follow-up period without receiving a diagnosis of breast cancer. All women whose cancer status was indeterminate had no additional pathological or imaging information available. The reference standard for all other participants who did not fall into these three categories was classified as unknown. Participants with either positive or negative reference-standard status made up the fully verified group.

ROC curves for digital and film mammography were estimated from the pooled data across the study with the use of the malignancy score assigned to each woman at the time of screening mammography and before further workup was conducted. The full areas under the curve (AUCs) were compared with the use of the bivariate, binormal model, which accounts for the paired test design.[27,28] A corroborating, nonparametric AUC analysis was also performed.[29,30] The AUCs were compared in the entire study cohort (primary study aim) as well as in prespecified subgroups of participants (secondary aims). The latter included subgroups defined according to age (younger than 50 years vs. 50 years or older), breast density (heterogeneously dense or extremely dense vs. less dense), menopausal status (premenopausal or perimenopausal vs. postmenopausal), race (white vs. black vs. other), risk of breast cancer (a lifetime risk of ≥25 percent vs. <25 percent, as determined by the Gail model[31]), and the four digital-machine manufacturers. The Bonferroni procedure was used to account for the 15 multiple comparisons in the subgroup analysis, with a P value of 0.003 or less considered to indicate statistical significance.

For descriptive purposes, estimates of the sensitivity, specificity, and positive and negative predictive values of the two methods of mammography were computed on the basis of the seven-point malignancy scale, the BIRADS scale, and the presence or absence of a workup recommendation by the radiologist. For this purpose, the scores for the seven-point malignancy scale were dichotomized as negative (score of 1, 2, or 3) and positive (score of 4, 5, 6, or 7), and the BIRADS ratings were dichotomized as negative (score of 1, 2, or 3) and positive (score of 0, 4, or 5). Results were evaluated for 365 and 455 days of follow-up. McNemar's test was used to compare estimates.

The analysis was confined to the fully verified group. We assessed the effect of missing information on disease status by deriving and comparing estimates of AUCs and sensitivity and specificity using methods for correcting for verification bias in the ROC analysis[30] and in the comparisons of sensitivity and specificity.[32] Both methods incorporate available information on covariates and assume that the verification status depends only on test outcomes and observed covariates.

RESULTS

STUDY POPULATION

A total of 49,528 women were enrolled in the trial. Of these, 195 (0.4 percent) were subsequently determined to be ineligible and 194 (0.4 percent) withdrew from the study. In addition, 1489 women (3.0 percent) were excluded from the analysis because the study protocol had not been followed at one participating institution, as determined by on-site audits. Thirty-nine additional women were excluded because the same radiologist interpreted both examinations or the radiologist knew the results of the other examination at the time of interpretation, and 12 were excluded because the examinations were technically inadequate (9 with inadequate film examinations and 3 with inadequate digital examinations). Of the 47,599 remaining women, follow-up information was lacking for 4339 (9.1 percent), and 500 (1.1 percent) had an indeterminate cancer status (474 with follow-up mammograms interpreted as having a BIRADS

score of 3, 4, or 5; 20 who had insufficient biopsy specimens or nondiagnostic biopsy findings; 6 who died without receiving a diagnosis of breast cancer; and none of whom had definitive information concerning pathological or imaging results). Thus, we were left with data on 42,760 women (86.7 percent of those eligible) for the primary analysis. All interpreted mammograms other than the listed exclusions were included in the analysis, including those obtained from 203 women who underwent only one type of mammography (188 [0.4 percent] underwent film mammography alone, and 15 [0.04 percent] digital mammography alone, primarily owing to equipment malfunctions). Table 1 lists the characteristics of the eligible women and the women who were included in the analysis.

Table 1. Characteristics of Eligible Women and Women Whose Cancer Status Was Verified.*

Characteristic	Eligible Women (N=49,333)	Women with Verified Cancer Status (N=42,760)
Age at enrollment — yr		
Mean	54.6	54.9
Interquartile range	47–61	47–62
Race or ethnic group — no. (%)†		
White	40,409 (81.9)	36,079 (84.4)
Hispanic or Latina	2,012 (4.1)	1,266 (3.0)
Black or African American	5,439 (11.0)	4,243 (9.9)
Native Hawaiian or other Pacific Islander	64 (0.1)	61 (0.1)
Asian	923 (1.9)	793 (1.9)
American Indian or Alaskan Native	46 (0.1)	37 (0.1)
Other race specified	396 (0.8)	244 (0.6)
Unknown or data missing	44 (0.1)	37 (0.1)
Menopausal status — no. (%)‡		
Premenopausal	14,349 (29.1)	12,024 (28.1)
Perimenopausal	4,294 (8.7)	3,779 (8.8)
Postmenopausal	29,569 (59.9)	26,087 (61.0)
Unknown or data missing	1,121 (2.3)	870 (2.0)
Breast density — no. (%)		
Almost entirely fat	5,184 (10.5)	4,364 (10.2)
Scattered fibroglandular densities	21,171 (42.9)	18,480 (43.2)
Heterogeneously dense	19,089 (38.7)	16,793 (39.3)
Extremely dense	3,690 (7.5)	3,104 (7.3)
Data missing	199 (0.4)	19 (<0.1)

* Because of rounding, percentages may not total 100.
† Race or ethnic group was self-assigned.
‡ Premenopausal women had had their last menstrual period less than one month before mammography. Perimenopausal women had had their last menstrual period at least 1 month but less than 12 months before mammography.

INTERPRETATION OF THE IMAGES

Using the dichotomized seven-point malignancy scale, we found that 223 women (0.5 percent) had both positive digital and positive film mammograms, 947 women (2.2 percent) had only positive digital mammograms, 832 women (1.9 percent) had only positive film mammograms, and 40,553 women (94.8 percent) had neither positive film nor positive digital examinations. For the remaining 205 women (0.5 percent), interpretations for either digital or film mammograms were missing (187 negative and 3 positive film examinations and 15 negative digital mammograms).

Using the dichotomized BIRADS scale, we found that 1249 women (2.9 percent) had both positive digital and positive film mammograms, 2399 women (5.6 percent) had only positive digital mammograms, 2416 women (5.7 percent) had only positive film mammograms, and 36,696 (85.8 percent) had neither positive film nor positive digital examinations.

BREAST CANCERS

A total of 335 breast cancers were diagnosed in the DMIST cohort on the basis of reference-standard information during the 455 days after study entry (Table 2). Of these 335 cancers, 254 (75.8 percent) were diagnosed within 365 days after study mammography and 81 (24.2 percent) were diagnosed between 366 and 455 days after study mammography. The histologic findings and the stage of the breast cancers detected by the two methods were similar.

DIAGNOSTIC PERFORMANCE OF DIGITAL AND FILM MAMMOGRAPHY

The diagnostic accuracy of digital and film mammography was similar in the fully verified group, as reflected by a mean (±SE) AUC of 0.78±0.02 for digital mammography and of 0.74±0.02 for film mammography (difference in AUC, 0.03; 95 percent confidence interval, −0.02 to 0.08; P=0.18) (Fig. 1A). The AUC for digital mammography also did not vary significantly from that for film mammography according to race, the risk of breast cancer, or the type of digital machine used.

The performance of digital mammography was, however, significantly better than that of film mammography among women under the age of 50 years,

Table 2. Pathological Diagnosis and Stage of 335 Cancers among Women Referred for a Workup after Initial Imaging.*

Diagnosis	Both Film and Digital Mammography				Film Mammography Alone				Digital Mammography Alone				Neither Type of Mammography				Total
	All	Women <50 yr Old	Premenopausal and Perimenopausal Women	Women with Heterogeneously Dense or Extremely Dense Breasts	All	Women <50 yr Old	Premenopausal and Perimenopausal Women	Women with Heterogeneously Dense or Extremely Dense Breasts	All	Women <50 yr Old	Premenopausal and Perimenopausal Women	Women with Heterogeneously Dense or Extremely Dense Breasts	All	Women <50 yr Old	Premenopausal and Perimenopausal Women	Women with Heterogeneously Dense or Extremely Dense Breasts	
	number of women (percent)																
Invasive carcinoma†	85 (25.4)	12 (3.6)	16 (4.8)	36 (10.7)	35 (10.4)	3 (0.9)	7 (2.1)	12 (3.6)	38 (11.3)	14 (4.2)	19 (5.7)	26 (7.8)	73 (21.8)	14 (4.2)	18 (5.4)	41 (12.2)	231 (69.0)
Invasive ductal carcinoma +/– DCIS	73 (21.8)	9 (2.7)	13 (3.9)	33 (9.9)	26 (7.8)	2 (0.6)	6 (1.8)	8 (2.4)	30 (9.0)	11 (3.3)	14 (4.2)	19 (5.7)	60 (17.9)	10 (3.0)	13 (3.9)	32 (9.6)	189 (56.4)
Invasive lobular carcinoma +/– DCIS	5 (1.5)	2 (0.6)	3 (0.9)	1 (0.3)	5 (1.5)	1 (0.3)	1 (0.3)	3 (0.9)	6 (1.8)	3 (0.9)	4 (1.2)	5 (1.5)	5 (1.5)	2 (0.6)	3 (0.9)	3 (0.9)	21 (6.3)
Mixed invasive ductal and lobular carcinoma +/– DCIS	7 (2.1)	1 (0.3)	0	2 (0.6)	4 (1.2)	0	0	1 (0.3)	2 (0.6)	0	1 (0.3)	2 (0.6)	8 (2.4)	2 (0.6)	2 (0.6)	6 (1.8)	21 (6.3)
DCIS†	36 (10.7)	14 (4.2)	16 (4.8)	18 (5.4)	17 (5.1)	3 (0.9)	4 (1.2)	7 (2.1)	25 (7.5)	8 (2.4)	14 (4.2)	14 (4.2)	25 (7.5)	4 (1.2)	6 (1.8)	11 (3.3)	103 (30.7)
High grade	15 (4.5)	6 (1.8)	8 (2.4)	9 (2.7)	6 (1.8)	2 (0.6)	1 (0.3)	3 (0.9)	7 (2.1)	3 (0.9)	6 (1.8)	4 (1.2)	12 (3.6)	1 (0.3)	3 (0.9)	4 (1.2)	40 (11.9)
Medium grade	14 (4.2)	5 (1.5)	4 (1.2)	6 (1.8)	4 (1.2)	0	1 (0.3)	1 (0.3)	12 (3.6)	4 (1.2)	6 (1.8)	5 (1.5)	7 (2.1)	1 (0.3)	1 (0.3)	3 (0.9)	37 (11.0)
Low grade	6 (1.8)	3 (0.9)	4 (1.2)	3 (0.9)	6 (1.8)	1 (0.3)	2 (0.6)	3 (0.9)	6 (1.8)	1 (0.3)	2 (0.6)	5 (1.5)	6 (1.8)	2 (0.6)	2 (0.6)	4 (1.2)	24 (7.2)
Unknown grade	1 (0.3)	0	0	0	1 (0.3)	0	0	0	0	0	0	0	0	0	0	0	2 (0.6)
Other malignant cancer†	1 (0.3)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (0.3)
T stage	122 (36.4)	26 (7.8)	32 (9.6)	54 (16.1)	52 (15.5)	6 (1.8)	11 (3.3)	19 (5.7)	63 (18.8)	22 (6.6)	33 (9.9)	40 (11.9)	98 (29.3)	18 (5.4)	24 (7.2)	52 (15.5)	335 (100.0)
No stage‡	1 (0.8)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1 (0.3)
Tx§	12 (9.8)	4 (15.4)	3 (9.4)	7 (13.0)	4 (7.7)	2 (33.3)	0	0	9 (14.3)	4 (18.2)	5 (15.2)	6 (15.0)	13 (13.3)	0	0	4 (7.7)	38 (11.3)
Tis§	36 (29.5)	14 (53.8)	16 (50.0)	18 (33.3)	17 (32.7)	3 (50.0)	4 (36.4)	7 (36.8)	25 (39.7)	8 (36.4)	14 (42.4)	14 (35.0)	25 (25.5)	4 (22.2)	6 (25.0)	11 (21.2)	103 (30.7)
T1mic§	2 (1.6)	1 (3.8)	1 (3.1)	0	1 (1.9)	0	0	0	1 (1.6)	0	0	1 (2.5)	1 (1.0)	1 (5.6)	1 (4.2)	1 (1.9)	5 (1.5)
T1a§	11 (9.0)	3 (11.5)	2 (6.3)	6 (11.1)	3 (5.8)	0	1 (9.1)	1 (5.3)	4 (6.3)	2 (9.1)	1 (3.0)	1 (2.5)	5 (5.1)	0	0	1 (1.9)	23 (6.9)
T1b§	20 (16.4)	0	2 (6.3)	6 (11.1)	8 (15.4)	1 (16.7)	3 (27.3)	2 (10.5)	8 (12.7)	1 (4.5)	1 (3.0)	6 (15.0)	18 (18.4)	1 (5.6)	2 (8.3)	7 (13.5)	54 (16.1)
T1c§	29 (23.8)	1 (3.8)	3 (9.4)	13 (24.1)	15 (28.8)	0	1 (9.1)	6 (31.6)	11 (17.5)	4 (18.2)	8 (24.2)	7 (17.5)	18 (18.4)	5 (27.8)	6 (25.0)	13 (25.0)	73 (21.8)
T2§	10 (8.2)	3 (11.5)	5 (15.6)	3 (5.6)	4 (7.7)	0	2 (18.2)	3 (15.8)	2 (3.2)	1 (4.5)	1 (3.0)	2 (5.0)	17 (17.3)	7 (38.9)	9 (37.5)	14 (26.9)	33 (9.9)
T3§	1 (0.8)	0	0	1 (1.9)	0	0	0	0	3 (4.8)	2 (9.1)	3 (9.1)	3 (7.5)	1 (1.0)	0	0	1 (1.9)	5 (1.5)
N stage (invasive tumors)§	85 (36.8)	12 (5.2)	16 (6.9)	36 (15.6)	35 (15.2)	3 (1.3)	7 (3.0)	12 (5.2)	38 (16.5)	14 (6.1)	19 (8.2)	26 (11.3)	73 (31.6)	14 (6.1)	18 (7.8)	41 (17.7)	231 (100.0)
Nx	25 (29.4)	5 (41.7)	4 (25.0)	14 (38.9)	9 (25.7)	3 (100.0)	2 (28.6)	3 (25.0)	12 (31.6)	4 (28.6)	5 (26.3)	9 (34.6)	18 (24.7)	1 (7.1)	1 (5.6)	7 (17.1)	64 (27.7)
N0	44 (51.8)	6 (50.0)	7 (45.8)	16 (44.4)	19 (54.3)	0	2 (28.6)	4 (33.3)	20 (52.6)	7 (50.0)	9 (47.4)	12 (46.2)	39 (53.4)	8 (57.1)	9 (50.0)	23 (56.1)	122 (52.8)
N1	16 (18.8)	1 (8.3)	5 (31.3)	6 (16.7)	7 (20.0)	0	3 (42.9)	5 (41.7)	6 (15.8)	3 (21.4)	5 (26.3)	5 (19.2)	13 (17.8)	4 (28.6)	6 (33.3)	8 (19.5)	42 (18.2)
N2	0	0	0	0	0	0	0	0	0	0	0	0	3 (4.1)	1 (7.1)	2 (11.1)	3 (7.3)	3 (1.3)

* The number of women in the subgroups does not equal the total because some women were included in more than one subgroup. Premenopausal women had had their last menstrual period less than one month before mammography. Perimenopausal women had had their last menstrual period at least 1 month but less than 12 months before mammography. DCIS denotes ductal carcinoma in situ, Tx inability to assess tumor size, Tis carcinoma in situ alone, T1 tumor no larger than 2 cm in diameter, T1mic tumor 0.1 cm or smaller, T1a tumor larger than 0.1 cm but not larger than 0.5 cm, T1b tumor larger than 0.5 cm but not larger than 1.0 cm, T1c tumor larger than 1.0 cm but not larger than 2.0 cm, T2 tumor larger than 2.0 cm but not larger than 5.0 cm, T3 tumor larger than 5.0 cm, Nx inability to assess nodes, N0 no cancer in axillary nodes, N1 axillary nodes moveable but have cancer, and N2 axillary nodes fixed and have cancer.
† The percentages are the percentages of the 335 cases of cancer detected in the 455 days after initial imaging.
‡ The one case without an assigned stage was a lymphoma.
§ The percentages are the percentages of the total number of cases of cancer with information about tumor (T) or nodal (N) stage.

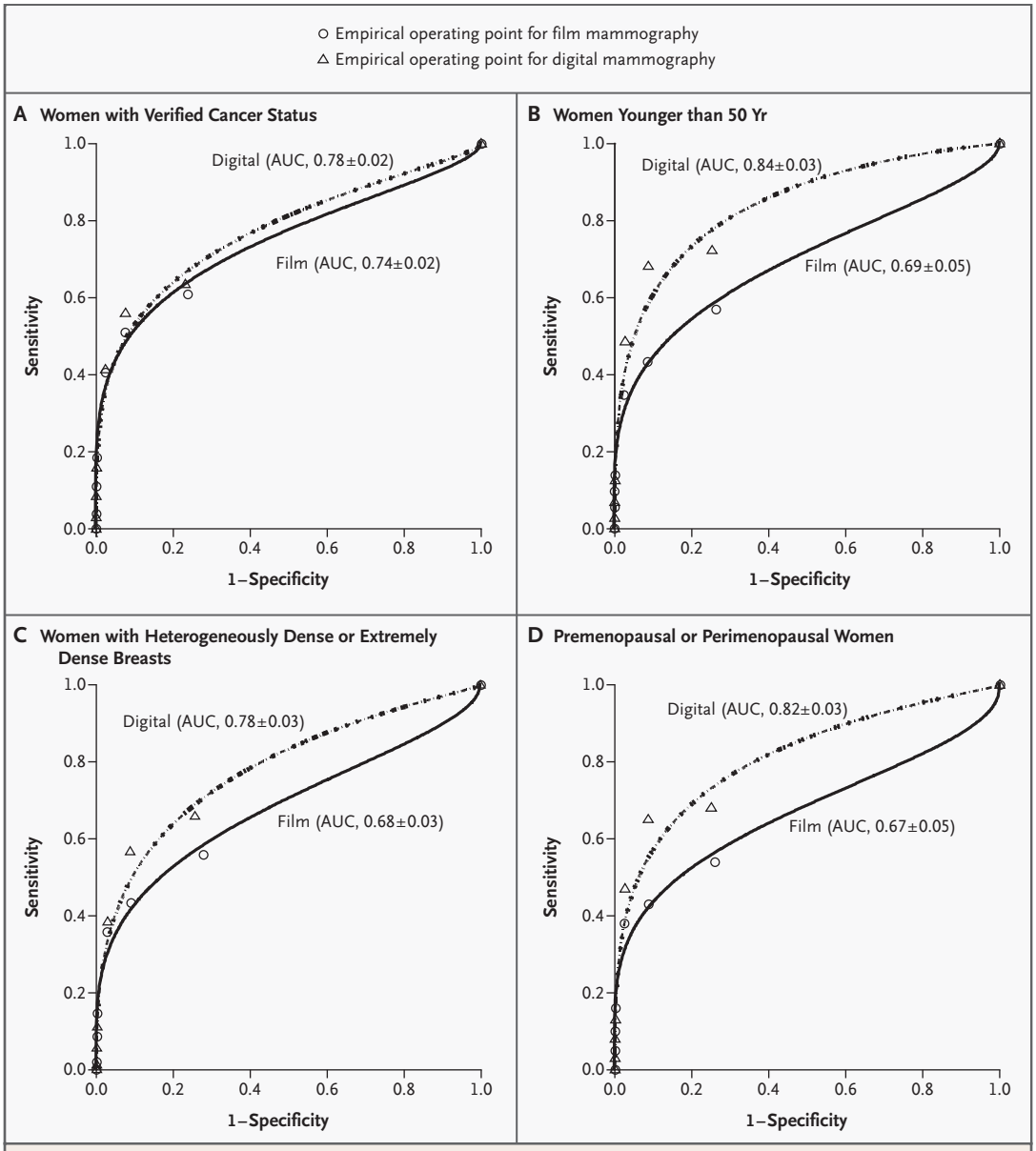

Figure 1. ROC Points and Fitted AUCs for Digital and Film Mammography for the 42,760 Women with Fully Verified Breast-Cancer Status (Panel A), the 14,335 Women under the Age of 50 Years (Panel B), the 19,897 Women with Heterogeneously or Extremely Dense Breasts (Panel C), and the 15,803 Premenopausal or Perimenopausal Women (Panel D).

Diagnostic performance was determined with the use of the seven-point malignancy scale. Premenopausal women were defined as those whose last menstrual period was less than one month before mammography. Perimenopausal women were defined as those whose last menstrual period was at least 1 month but less than 12 months before mammography.

as compared with those who were at least 50 years of age (AUC for digital mammography, 0.84±0.03; AUC for film mammography, 0.69±0.05; difference, 0.15; 95 percent confidence interval, 0.05 to 0.25; P=0.002) (Fig. 1B), women classified by the readers as having heterogeneously dense or extremely dense breasts (AUC for digital mammography, 0.78±0.03; AUC for film mammography, 0.68±0.03; difference, 0.11; 95 percent confidence interval, 0.04 to 0.18; P=0.003) (Fig. 1C), and premenopausal or perimenopausal women (AUC for digital mammography, 0.82±0.03; AUC for film mammography,

0.67±0.05; difference, 0.15; 95 percent confidence interval, 0.05 to 0.24; P=0.002) (Fig. 1D). The results of the AUC comparison in the full cohort and the prespecified subgroups were qualitatively similar to those obtained in the analysis that corrected for potential verification bias. There was no significant difference in the AUC between digital and film mammography among women 50 years of age or older, women with fatty breasts or scattered fibroglandular densities, and postmenopausal women.

Tables 3 and 4 show estimates of the sensitivity, specificity, and positive predictive value of each method on the basis of the seven-point malignancy scale after 455 days of follow-up and the BIRADS scale after 365 days of follow-up, dichotomized at each possible threshold. The tables also show digital and film mammography in terms of their sensitivities and specificities, computed at the main thresholds specified above. Detailed results of statistical analyses for sensitivity and specificity with the use of the seven-point malignancy scale at the 365-day follow-up and the BIRADS scale at the 455-day follow-up are provided in the Supplementary Appendix (available with the full text of this article at www.nejm.org). When the comparisons of sensitivities and specificities were adjusted for verification bias, the results were qualitatively similar.

DISCUSSION

We found that digital mammography was significantly better than conventional film mammography at detecting breast cancer in young women, premenopausal and perimenopausal women, and women with dense breasts. There was no significant difference in diagnostic accuracy between digital and film mammography in the population as a whole or in other predefined subgroups. However, digital mammography offers other advantages over film mammography — namely, easier access to images and computer-assisted diagnosis; improved means of transmission, retrieval, and storage of images; and the use of a lower average dose of radiation without a compromise in diagnostic accuracy.[33] We believe that the significant improvement in accuracy in specific subgroups of women justifies the use of digital mammography in these groups.

Our results are understandable in the light of the technical advantages of digital mammography over film mammography. In a digital image, the x-ray transmission can be manipulated to enhance visualization of subtle structural changes in tissue over the entire breast. For mammograms, the most problematic areas are those in which cancers can be hidden by adjacent dense tissue owing to small differences in contrast between lesions and the fibroglandular background. The visibility of a subtle mass or cluster of calcifications present in the image can be increased if the image contrast is adjusted.[34,35]

DMIST did not measure mortality end points. The assumption inherent in the design of the trial is that screening mammography reduces the rate of death from breast cancer and that if digital mammography detects cancers at a rate that equals or exceeds that of film mammography, its use in screening is likely to reduce the risk of death by as much as or more than that conferred by film mammography. The evidence supporting this view is given in Table 2. The cancers detected by digital mammography and missed by film mammography in women under the age of 50 years, women with heterogeneously dense or extremely dense breasts, and premenopausal and perimenopausal women included many invasive and high-grade in situ cases. These are precisely the lesions that must be detected early to save lives through screening. Neither digital nor film mammography found all the breast cancers in the population. Palpable findings and symptoms that develop after screening should be evaluated even if a woman has negative findings on digital mammography.

Why were the sensitivities of both digital and film mammography measured in this study apparently lower than the sensitivities in other studies?[20-23] Estimates of sensitivity depend on the definition used.[24] We considered any woman presenting with breast cancer within 455 days after study entry to have been positive for breast cancer at the time of her initial screening mammogram. All women with negative findings on mammography at study entry who had breast cancer at the annual follow-up mammography were thus considered to have false negative results for the analysis. The longer follow-up interval was selected to allow study sites to complete the one-year follow-up and subsequent workup. Some of the cancers detected up to 455 days after study entry were probably present at the time of the initial mammogram, but the use of the 455-day follow-up interval for reporting estimates of diagnostic accuracy is unconventional. Table 4 gives estimates of the diagnostic performance of both digital and film mammography at all cutoff points of the BIRADS scale during the 365-day follow-up peri-

Table 3. Diagnostic Accuracy of Digital and Film Mammography with the Use of the Seven-Point Malignancy Scale after 455 Days of Follow-up.*

Variable	Malignancy Score on Digital Mammography							Malignancy Score on Film Mammography							Difference†	
	7	6	5	4	3	2	1	7	6	5	4	3	2	1	Value (95% CI)	P Value
All women																
No. of tests	11	29	69	1061	2224	6588	32,588	17	29	70	942	2291	6,910	32,486		
No. of breast cancers	10	18	25	85	49	25	122	13	24	25	74	35	33	131		
Cumulative no. of tests	11	40	109	1170	3394	9982	42,570‡	17	46	116	1058	3349	10,259	42,745‡		
Cumulative no. of true positive results	10	28	53	138	187	212	334§	13	37	62	136	171	204	335		
Sensitivity for all cancers	0.03	0.08	0.16	0.41	0.56	0.63	1.00	0.04	0.11	0.19	0.41	0.51	0.61	1.00		
Sensitivity for invasive cancers	0.04	0.11	0.19	0.40	0.53	0.62	1.00	0.05	0.13	0.21	0.42	0.52	0.63	1.00		
Specificity for all cancers	1.00	1.00	1.00	0.98	0.92	0.77	0.00	1.00	1.00	1.00	0.98	0.93	0.76	0.00		
Positive predictive value	0.91	0.70	0.49	0.12	0.06	0.02	0.01	0.76	0.80	0.53	0.13	0.05	0.02	0.01		
No. of women who underwent biopsies	11	22	41	296	314	210	460	14	27	50	271	321	206	467		
Sensitivity				0.41±0.03							0.41±0.03				0.01 (−0.06 to 0.07)	0.92
Specificity				0.98±0.001							0.98±0.001				−0.002 (−0.005 to −0.001)	0.006
Positive predictive value				0.12±0.01							0.13±0.01					
Women <50 yr old																
Sensitivity				0.49±0.06							0.35±0.06				0.14 (−0.01 to 0.28)	0.06
Specificity				0.97±0.001							0.98±0.001				−0.003 (−0.007 to 0.0003)	0.07
Positive predictive value				0.08±0.01							0.07±0.01					
Pre- and perimenopausal women																
Sensitivity				0.47±0.05							0.38±0.05				0.09 (−0.04 to 0.22)	0.20
Specificity				0.97±0.001							0.98±0.001				−0.002 (−0.006 to 0.001)	0.20
Positive predictive value				0.10±0.01							0.09±0.01					
Women with heterogeneously dense or extremely dense breasts																
Sensitivity				0.38±0.04							0.36±0.04				0.03 (−0.07 to 0.12)	0.69
Specificity				0.97±0.001							0.97±0.001				−0.002 (−0.005 to 0.002)	0.33
Positive predictive value				0.10±0.01							0.10±0.01					

* Plus–minus values are means ±SE. Scores for the seven-point malignancy scale range from 1 (definitely not malignant) to 7 (definitely malignant). In the sensitivity analyses, scores of 4, 5, 6, and 7 were defined as positive and scores of 1, 2, and 3 were defined as negative. Premenopausal women had had their last menstrual period less than one month before mammography. Perimenopausal women had had their last menstrual period at least 1 month but less than 12 months before mammography.

† The difference was obtained by subtracting the value for film mammography from the value for digital mammography. CI denotes confidence interval.

‡ Of the 42,760 women whose cancer status was fully verified, only 42,555 had complete follow-up information: 190 women did not undergo digital mammography, and 15 did not undergo film mammography.

§ One woman who received a diagnosis of cancer did not undergo digital mammography.

Table 4. Diagnostic Accuracy of Digital and Film Mammography with the Use of the BIRADS Score after 365 Days of Follow-up.*

Variable	BIRADS Score on Digital Mammography						BIRADS Score on Film Mammography						Difference†	
	5	4	0	3	2	1	5	4	0	3	2	1	Value (95% CI)	P Value
All women														
No. of tests	9	23	3623	68	8,564	30,283	8	23	3648	53	8,113	30,900		
No. of breast cancers	7	8	162	0	26	51	7	6	154	0	23	64		
Cumulative no. of tests	9	32	3655	3723	12,287	42,570‡	8	31	3679	3732	11,845	42,745‡		
Cumulative no. of true positive results	7	15	177	177	203	254§	7	13	167	167	190	254§		
Sensitivity for all cancers	0.03	0.06	0.70	0.70	0.80	1.00	0.03	0.05	0.66	0.66	0.75	1.00		
Sensitivity for invasive cancers	0.03	0.06	0.67	0.67	0.77	1.00	0.03	0.06	0.66	0.66	0.77	1.00		
Specificity for all cancers	1.00	1.00	0.92	0.92	0.71	0.00	1.00	1.00	0.92	0.92	0.73	0.00		
Positive predictive value	0.78	0.47	0.05	0.05	0.02	0.01	0.88	0.42	0.05	0.04	0.02	0.01		
No. of women who underwent biopsies	8	18	655	4	127	235	8	19	658	3	105	256		
Sensitivity				0.70±0.03						0.66±0.03			0.04	(−0.04 to 0.12) 0.37
Specificity				0.92±0.001						0.92±0.001			0.001	(−0.003 to 0.004) 0.74
Positive predictive value				0.05±0.004						0.05±0.003				
Women <50 yr old														
Sensitivity				0.78±0.05						0.51±0.07			0.27	(0.11 to 0.44) 0.002
Specificity				0.90±0.003						0.90±0.003			0	(−0.006 to 0.006) 0.89
Positive predictive value				0.03±0.005						0.02±0.004				
Pre- and perimenopausal women														
Sensitivity				0.72±0.05						0.51±0.06			0.21	(0.06 to 0.36) 0.008
Specificity				0.90±0.002						0.90±0.002			0.002	(−0.003 to 0.008) 0.37
Positive predictive value				0.04±0.005						0.03±0.004				
Women with heterogeneously dense or extremely dense breasts														
Sensitivity				0.70±0.04						0.55±0.04			0.14	(0.03 to 0.26) 0.02
Specificity				0.91±0.002						0.90±0.002			0.004	(−0.001 to 0.010) 0.09
Positive predictive value				0.04±0.005						0.03±0.004				

* Plus–minus values are means ±SE. BIRADS scores can range from 0 (incomplete data) to 5 (highly suggestive of cancer). In the sensitivity analyses, scores of 0, 4, and 5 were defined as positive and scores of 1, 2, and 3 were defined as negative. Premenopausal women had had their last menstrual period less than one month before mammography. Perimenopausal women had had their last menstrual period at least 1 month but less than 12 months before mammography. One woman who received a diagnosis of cancer did not undergo digital mammography.

† The difference was obtained by subtracting the value for film mammography from the value for digital mammography. CI denotes confidence interval.

‡ Of the 42,760 women whose cancer status was fully verified, only 42,555 had complete follow-up information: 190 women did not undergo digital mammography, and 15 did not undergo film mammography.

§ A total of 254 cancers were diagnosed in the 365-day follow-up period.

od. This allows our estimates of diagnostic performance to be compared with those of others.[22,23,25]

Although the lead radiologists at each site were trained in the use of the seven-point malignancy scale and they then trained the other radiologists interpreting mammograms, this scale has not been used in other large, published studies. Our results using the BIRADS or follow-up scales can more readily be compared with those published elsewhere.[5,6,25] In addition, the percentage of the total population recalled for further workup (14.0 percent) is relatively high, because women underwent two screening tests (digital and film mammography), not just one. The call-back rate of 8.4 percent for both digital and film mammography is similar to or lower than those reported elsewhere for U.S. screening programs.[21,26,36]

One of the major impediments to the adoption of digital mammography will be its cost: digital systems currently cost approximately 1.5 to 4 times as much as film systems. As part of DMIST, we are performing a formal cost-effectiveness analysis and study of the quality of life of asymptomatic women.

Supported by grants from the National Cancer Institute (U01 CA80098, U01 CA80098-S1, U01 CA79778, and U01 79778-S1).

We are indebted to the many people at the headquarters of the American College of Radiology Imaging Network and at the recruiting sites for their important contributions to the study; to the radiologists, physicists, and research associates at the clinical sites; to Dennis Fryback, Anna Tosteson, Shahla Masood, Bruce Hillman, Mitchell Schnall, Thomas Caldwell, Stephen King, Charles Apgar, Irene Mahon, Sophia Sabina, Bernadine Dunning, Jamie Downs, Tess Thompson, Heather Wallace, Elaine Pakuris, Donna Hartfeil, Jessie Flaim-Spetsas, Boris Ginsburgs, Sharon Jones, Maria Oh, Rex Welsh, Tim Welsh, Fraser Wilton, Anthony Levering, Anita Murray, Brenda Young, Cheryl Crozier, Mary Kelly Truran, Chris Steward, Thomas Iarocci, Crystal Wright, Janet Vogel, Karan Boparai, Rolma Mancinow, Josephine Schloesser, Sharlene Snowdon, Vish Iyer, JoAnn Stetz, Robert Smith, and the other members of the data and safety monitoring board; to Aili Bloomquist, Gordon Mawdsley, Sam Shen, Mary Brown, Elodia Cole, Beverly Currence, Cherie Kuzmiak, Ann Sherman, Jason Hauser, Dag Pavic, Marcia Koomen, Robert McLelland, Richard Clark, and the following members of the American College of Radiology Biostatistics Center: Lucy Hanna, Alicia Toledano, Ben Herman, Minran Li, Jean Cormack, Prashni Paliwal, Shang-Ying Shiu, and Helga Marques; and to the late Jo-Ann D'Amato for her important work on this project.

APPENDIX

The following persons served as principal investigators (PIs) or lead physicists (LPs) at the DMIST clinical sites: Allegheny General Hospital, Pittsburgh — W. Poller (PI), J. Och (LP); Beth Israel Deaconess Medical Center, Boston — J. Baum (PI), R. Zamenhof (LP); Brown University, Providence, R.I. — B. Schepps (PI), D. Shearer (LP); Columbia University, New York — S.J. Smith (PI), E. Nickoloff (LP); Elizabeth Wende Breast Clinic, Rochester, N.Y. — E. Bonaccio (PI), M. Zuley (PI), A. Tibold (LP); Emory University, Atlanta — C. D'Orsi (PI), P. Sprawls (LP); Johns Hopkins University, Baltimore — N. Khouri (PI), M. Mahesh (LP); LaGrange Hospital, LaGrange, Ill. — T. Merrill (PI), C. Vyborny (PI), R. Nishikawa (LP); Lahey Clinic, Burlington, Mass. — R.B. Shah (PI), N. Shaikh (LP); Massachusetts General Hospital, Boston — D. Georgian-Smith (PI), J. Quattrochi (LP); Memorial Sloan-Kettering Cancer Center, New York — M. Cohen (PI), R. Fleischman (LP); H. Lee Moffitt Cancer Center, Tampa, Fla. — A.P. Romilly (PI), K. Coleman (LP); Monmouth County Hospital, Long Branch, N.J. — M. Staiger (PI), T. Piccoli (LP); Mount Sinai University, New York — S. Feig (PI), Jose Burgos (LP); Northwestern University, Chicago — R.E. Hendrick (PI), E. Berns (LP); Shore Memorial Hospital, Somers Point, N.J. — R. Menghetti (PI), J. Law (LP); Thomas Jefferson University, Philadelphia — C. Piccoli (PI), A. Maidment (LP), E. Gingold (LP); University of California at Davis, Davis — K. Lindfors (PI), A. Seibert (LP), J. Boone (LP); University of California at Los Angeles, Los Angeles — L. Bassett (PI), V. Cooper (LP); University of Cincinnati, Cincinnati — M. Mahoney (PI), R. Samaratunga (LP); University of Colorado, Denver — P. Isaacs (PI), J Lewin (PI), F. Larke (LP); University of Iowa, Iowa City — L. Fajardo (PI), K. Berbaum (LP), M. Madsen (LP); University of North Carolina, Chapel Hill — E. Pisano (PI), R.E. Johnston (LP); University of Pennsylvania, Philadelphia — E. Conant (PI),M. O'Shea (LP), A. Maidment (LP); University of Texas Southwestern Medical Center, Dallas — W.P. Evans III (PI), M. Hatab (LP); University of Toronto, Toronto — M. Yaffe (PI), A. Bloomquist (LP), G. Mawdsley (LP); University of Virginia, Charlottesville — J. Harvey (PI), M. Williams (LP); University of Washington, Seattle — A. Freitas (PI), K. Kanal (LP); Washington Radiology Associates, Washington, D.C. — L. Glassman (PI), J. Greenberg (PI), M. Goodwill (LP); Washington University, St. Louis — D. Farria (PI), G. Fletcher (LP); William Beaumont Hospital, Royal Oak, Mich. — M. Rebner (PI), D. Bakalyar (LP).

REFERENCES

1. Humphrey LL, Helfand M, Chan BK, Woolf SH. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002; 137:347-60.
2. Institute of Medicine. Saving women's lives: integration and innovation: a framework for progress in early detection and diagnosis of breast cancer. Washington, D.C.: National Academies Press, 2005.
3. Fletcher SW, Elmore JG. Mammographic screening for breast cancer. N Engl J Med 2003;348:1672-80.
4. Buist DSM, Porter PL, Lehman C, Taplin SH, White E. Factors contributing to mammography failure in women aged 40-49 years. J Natl Cancer Inst 2004;96:1432-40.
5. Carney PA, Miglioretti DL, Yankaskas BC, et al. Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 2003; 138:168-75. [Erratum, Ann Intern Med 2003;138:771.]
6. Kerlikowske K, Grady D, Barclay J, Sickles EA, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 1996; 276:33-8.
7. Wolfe JN. Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer 1976;37:2486-92.
8. Byrne C, Schairer C, Wolfe J, et al. Mammographic features and breast cancer risk: effects with time, age, and menopause status. J Natl Cancer Inst 1995;87:1622-9.
9. Boyd NF, Dite GS, Stone J, et al. Heritability of mammographic density, a risk for breast cancer. N Engl J Med 2002;347:886-94.
10. Shtern F. Digital mammography and related technologies: a perspective from the National Cancer Institute. Radiology 1992; 183:629-30.

11. Pisano ED, Yaffe MJ, Hemminger BM, et al. Current status of full-field digital mammography. Acad Radiol 2000;7:266-80.
12. Pisano ED, Yaffe MJ. Digital mammography. Radiology 2005;234:353-61.
13. Cole E, Pisano ED, Brown M, et al. Diagnostic accuracy of Fischer Senoscan Digital Mammography versus screen-film mammography in a diagnostic mammography population. Acad Radiol 2004;11:879-86.
14. Hendrick RE, Lewin JM, D'Orsi CJ, et al. Non-inferiority study of FFDM in an enriched diagnostic cohort: comparison with screen-film mammography in 625 women. In: Yaffe MJ, ed. IWDM 2000: 5th International Workshop on Digital Mammography. Madison, Wis.: Medical Physics Publishing, 2001:475-81.
15. 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;179:671-7.
16. 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;229:877-84.
17. 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.
18. Pisano ED, Gatsonis CA, Yaffe MJ, et al. The American College of Radiology Imaging Network Digital Mammographic Imaging Screening Trial: objectives and methodology. Radiology 2005;236:404-12.
19. Breast Imaging Reporting and Data System (BI-RADS). 4th ed. Reston, Va.: American College of Radiology, 2003.
20. Duffy SW, Chen HH, Tabar L, Fagerberg G, Paci E. Sojourn time, sensitivity and positive predictive value of mammography screening for breast cancer in women aged 40-49. Int J Epidemiol 1996;25:1139-45.
21. Poplack SP, Tosteson AN, Grove MR, Wells WA Carney PA. Mammography in 53,803 women from the New Hampshire mammography network. Radiology 2000; 217:832-40.
22. Banks E, Reeves G, Beral V, et al. Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: cohort study. BMJ 2004;329:477.
23. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. J Natl Cancer Inst 2005; 97:358-67.
24. Rosenberg RD, Yankaskas BC, Hunt WC, et al. Effect of variations in operational definitions on performance estimates for screening mammography. Acad Radiol 2000;7:1058-68.
25. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol 1997;169:1001-8.
26. Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K. The performance of mammography screening in the USA and the UK. J Med Screen 2005;12:50-4.
27. Metz C, Wang P, Kronman HA. New approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck F, ed. Information processing in medical imaging. The Hague, the Netherlands: Nijihoff, 1984.
28. Zhou X-H, Obuchowski NA, McClish DK. Statistical methods in diagnostic medicine. New York: John Wiley, 2002.
29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-45.
30. Toledano AY, Gatsonis C. Generalized estimating equations for ordinal categorical data: arbitrary patterns of missing responses and missingness in a key covariate. Biometrics 1999;55:488-96.
31. Gail MH, Brinton LA, Byar DP, et al. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J Natl Cancer Inst 1989;81:1879-86.
32. Zhou X-H, Obuchowski N, McClish D. Statistical methods in diagnostic medicine. New York: Wiley Publishing, 2002.
33. Bloomquist AK, Yaffe MJ, Mawdsley GE, et al. Quality control for digital mammography in the ACRIN DMIST. Med Phys (in press).
34. Pisano ED, Cole EB, Major S, et al. Radiologists' preferences for digital mammographic display. Radiology 2000;216:820-30.
35. Pisano ED, Cole EB, Hemminger BM, et al. Image processing algorithms for digital mammography: a pictorial essay. Radiographics 2000;20:1479-91.
36. Ghate SV, Soo MS, Baker JA, Walsh R, Gimenez EI, Rosen EL. Comparison of recall and cancer detection rates for immediate versus batch interpretation of screening mammograms. Radiology 2005;235:31-5.