

Effectiveness of Computer-Aided Detection in Community Mammography Practice

Joshua J. Fenton, Linn Abraham, Stephen H. Taplin, Berta M. Geller, Patricia A. Carney, Carl D'Orsi, Joann G. Elmore, William E. Barlow; for the Breast Cancer Surveillance Consortium

Manuscript received July 15, 2010; revised April 18, 2011; accepted April 25, 2011.

Correspondence to: Joshua J. Fenton, MD, MPH, Department of Family and Community Medicine and Center for Healthcare Policy and Research, University of California, Davis, 4860 Y St, Ste 2300, Sacramento, CA 95817 (e-mail: joshua.fenton@ucdmc.ucdavis.edu).

Background Computer-aided detection (CAD) is applied during screening mammography for millions of US women annually, although it is uncertain whether CAD improves breast cancer detection when used by community radiologists.

Methods We investigated the association between CAD use during film-screen screening mammography and specificity, sensitivity, positive predictive value, cancer detection rates, and prognostic characteristics of breast cancers (stage, size, and node involvement). Records from 684 956 women who received more than 1.6 million film-screen mammograms at Breast Cancer Surveillance Consortium facilities in seven states in the United States from 1998 to 2006 were analyzed. We used random-effects logistic regression to estimate associations between CAD and specificity (true-negative examinations among women without breast cancer), sensitivity (true-positive examinations among women with breast cancer diagnosed within 1 year of mammography), and positive predictive value (breast cancer diagnosed after positive mammograms) while adjusting for mammography registry, patient age, time since previous mammography, breast density, use of hormone replacement therapy, and year of examination (1998–2002 vs 2003–2006). All statistical tests were two-sided.

Results Of 90 total facilities, 25 (27.8%) adopted CAD and used it for an average of 27.5 study months. In adjusted analyses, CAD use was associated with statistically significantly lower specificity (OR = 0.87, 95% confidence interval [CI] = 0.85 to 0.89, $P < .001$) and positive predictive value (OR = 0.89, 95% CI = 0.80 to 0.99, $P = .03$). A non-statistically significant increase in overall sensitivity with CAD (OR = 1.06, 95% CI = 0.84 to 1.33, $P = .62$) was attributed to increased sensitivity for ductal carcinoma in situ (OR = 1.55, 95% CI = 0.83 to 2.91; $P = .17$), although sensitivity for invasive cancer was similar with or without CAD (OR = 0.96, 95% CI = 0.75 to 1.24; $P = .77$). CAD was not associated with higher breast cancer detection rates or more favorable stage, size, or lymph node status of invasive breast cancer.

Conclusion CAD use during film-screen screening mammography in the United States is associated with decreased specificity but not with improvement in the detection rate or prognostic characteristics of invasive breast cancer.

J Natl Cancer Inst 2011;103:1152–1161

Computer-aided detection (CAD) is a radiological device designed to assist radiologists during mammography interpretation. CAD software algorithms analyze data from mammogram images to identify patterns associated with underlying breast cancers (1). After a radiologist completes an initial mammogram assessment, CAD marks potential abnormalities on the image for the consideration of the radiologist before making a final recommendation. In film mammography, CAD is coupled with a device to convert film mammograms to digital images and a viewing board while CAD is integrated directly into digital mammography environments.

CAD received Food and Drug Administration approval in 1998 on the basis of small studies, which suggested that CAD could increase breast cancer detection without undue increases in recall rates. Congress subsequently amended the Medicare statute to

include supplemental coverage for CAD use, and CAD is now applied on nearly three of four screening mammograms performed on American women (2). Analogous CAD systems are under development to assist interpretation of computed tomography and magnetic resonance imaging of the breast, chest, colon, liver, and prostate (3).

Despite broad acceptance and use, it is unclear if the benefits of CAD during screening mammography outweigh its potential risks and costs (4,5). Ideally, CAD would lead to earlier detection of high-risk cancers, particularly invasive tumors, by improving sensitivity for these cancers and reducing the incidence of advanced stage breast cancer (6,7). However, the high sensitivity of CAD for mammographic calcifications may shift diagnostic attention to relatively indolent cancers, such as ductal carcinoma in situ

(DCIS), which often present with calcifications (8–13). In contrast, CAD is less likely to mark invasive cancers presenting as uncalcified masses (8,9) and may therefore have little or no impact on the detection of higher-risk invasive breast cancer (14,15). Although meta-analyses suggest that CAD statistically significantly increases recall rates (16,17), clinical studies have generally included too few women with breast cancer to clarify whether CAD differentially affects detection of in situ vs invasive breast cancer, or whether CAD is associated with improved prognostic characteristics of breast cancer such as more localized stage or smaller tumor size (11–13,18,19). Such data are critical for understanding the potential of CAD to decrease breast cancer mortality—the ultimate goal of breast cancer screening.

In a study of CAD implementation from 1998 to 2002 within the national Breast Cancer Surveillance Consortium (BCSC), CAD was associated with decreased specificity and positive predictive value (PPV₁), and increased recall and biopsy rates, but no difference in cancer detection rates was observed (10). However, this early study did not include all BCSC sites, and of the 43 facilities, only seven used CAD for an average of 18 months from 1998 to 2002. The study also did not evaluate the prognostic characteristics of breast cancers such as tumor stage and size. In this study, we therefore examined performance impacts, cancer detection rates, and breast cancer prognostic characteristics associated with CAD implementation within 25 of 90 BCSC facilities from 1998 to 2006. We hypothesized that CAD would be associated with reduced specificity and that any increase in sensitivity associated with CAD would be attributable to increased detection of DCIS rather than invasive breast cancer. We further hypothesized that the prognostic characteristics of invasive breast cancer would be similar with and without CAD.

Methods

Setting

We studied mammography facilities participating in the BCSC, a federally supported collaboration that links mammography data to cancer outcomes in California, Colorado, North Carolina, New Hampshire, New Mexico, Washington, and Vermont (20). BCSC data quality is rigorously monitored, and BCSC sites have received institutional review board approval for active or passive consenting processes or a waiver of consent to enroll participants, link data, and perform analytical studies. All procedures are compliant with the Health Insurance Portability and Accountability Act, and BCSC sites have received a Federal Certificate of Confidentiality to protect the identities of patients, physicians, and facilities. We used standard BCSC definitions to identify screening mammogram samples (as opposed to mammograms performed for diagnostic purposes) and to define mammogram performance outcomes (eg, specificity, sensitivity, PPV₁, and cancer detection rates) (20). Practice patterns and outcomes within the BCSC are broadly representative of the practice of mammography in the United States (21).

Facilities and Patients

We included mammograms from facilities that averaged at least 100 film-screen screening mammograms per month of clinical activity from January 1, 1998, to December 31, 2006. Of 93 poten-

CONTEXT AND CAVEATS

Prior knowledge

Computer-aided detection (CAD) is used by many radiologists to analyze and interpret mammograms before making a final recommendation. Although three out of four screening mammograms include CAD in the United States, it is unclear if CAD use improves breast cancer detection and justifies the potential associated risks and costs.

Study design

Using data from 684 956 women and more than 1.6 million mammograms administered at Breast Cancer Surveillance Consortium facilities that implemented CAD between 1998 and 2006, the relationships between performance, cancer detection rates, and breast cancer prognostic characteristics and CAD implementation were investigated.

Contributions

CAD implementation was associated with decreased specificity and positive predictive value. CAD use was not associated with increased detection rates or more favorable prognostic characteristics.

Implications

The health benefits of CAD use during screening mammograms remain unclear, and the data indicate that the associated costs may outweigh the potential health benefits.

Limitations

Training of radiologists in CAD use and interpretation could vary widely. For this study, digitized mammography data were not available, thus film-screen mammograms were used for analysis. Film-screen mammograms are digitized before CAD, potentially introducing noise and adversely affecting performance. Also, the analysis assumed that after a facility implemented CAD, all subsequent mammograms were analyzed with CAD, introducing potential bias.

From the Editors

tially eligible facilities, we excluded three because CAD use could not be definitively determined. Within the remaining 90 facilities, we included screening mammograms for women aged 40 years or older, defined as bilateral mammograms designated by radiologists as routine screening. Although 17 (19%) facilities adopted digital mammography during the study period, we excluded digital mammograms because too few were performed to distinguish the performance impacts of digital mammography and CAD ($n = 23\ 830$ digital mammograms of 1 644 936 total mammograms, 1.45%).

Mammography Outcomes

We defined mammograms with Breast Imaging Reporting and Data System (BI-RADS; American College of Radiology, Reston, VA) assessments of 0, 4, or 5 as positive and mammograms with BI-RADS assessments of 1 and 2 as negative. Mammograms with a BI-RADS assessment of 3 were considered positive if the radiologist also recommended immediate evaluation but were considered negative if the radiologist did not recommend immediate evaluation (22). We ascertained breast cancer diagnoses (either DCIS or invasive breast cancer) for 1 year after the screening

mammography by linkage with regional Surveillance, Epidemiology, and End Results registries or with local or statewide tumor registries. This methodology captures both incident breast cancers diagnosed by mammography and interval breast cancers that arise within 1 year of screening mammography.

We defined specificity as the proportion of screening mammograms that were negative among women who were not diagnosed with breast cancer. We defined sensitivity as the proportion of screening mammograms that were positive among women diagnosed with breast cancer (including interval breast cancers) within 1 year of screening mammography. We defined PPV_1 as the proportion of women diagnosed with breast cancer after a positive screening mammogram (23).

We computed the recall rate as the proportion of all screening mammograms that were positive and the cancer detection rate as the number of cancers that were preceded by positive mammograms per 1000 screening mammograms. The cancer detection rate can also be computed as the incidence rate of breast cancer (per 1000 screening mammograms) multiplied by sensitivity. For each mammogram, we determined whether radiologists recommended a breast biopsy, a fine needle aspiration, or a surgical consultation within 90 days. We classified invasive breast cancers by the following prognostic characteristics: American Joint Committee on Cancer stage (24), tumor diameter (in millimeters), and presence of regional lymph node involvement. On the basis of cancer characteristics associated with increased mortality (7) and decreased survival (25,26), we defined tumors (including both DCIS and invasive cancer) as advanced if they were designated stage II–IV, if tumors measured greater than 15 mm in diameter, or if lymph node involvement was observed.

Computer-Aided Detection

Of 90 total facilities, 25 (27.8%) implemented CAD during the study period. Because a previous survey indicated that CAD was used on most if not all mammograms after implementation (10), we classified mammograms performed after facility implementation of CAD as exposed to CAD. We defined mammograms interpreted at facilities that never implemented CAD or before CAD implementation as non-CAD. We excluded mammograms interpreted during the initial 3 months of CAD use at each facility because radiologists may have been adapting to the technology during this period. Excluding this 3-month period, facilities used CAD for an average of 27.5 months (range = 4–60 months). The BCSC does not collect data on the specific CAD product used by individual facilities.

Patient Variables

We included patient variables associated with interpretive accuracy including age, breast density, and time since previous mammogram (27,28). Radiologists report BI-RADS breast density on each mammogram. We also specified whether patients were using hormone therapy (HT) because HT has been associated with increased risk of abnormal mammography, breast biopsy, and breast cancer (29–31). Patients report HT use on self-administered questionnaires before mammography.

Statistical Analyses

We performed bivariate analyses to compare patient characteristics and unadjusted outcomes by CAD use. We stratified descriptive

analyses by examination year (January 1, 1998, to December 31, 2002, vs January 1, 2003, to December 31, 2006), because US breast cancer incidence declined in mid-2002 when many women stopped using HT following the publication of Women's Health Initiative findings demonstrating a link between HT and breast cancer (31–33). For mammograms interpreted at facilities that implemented CAD, we stratified by CAD use rather than year. However, because the median year of CAD implementation was 2003, most post-CAD mammograms (88.5%) were interpreted from January 1, 2003, to December 31, 2006.

We performed random-effects logistic regression to model binary outcomes for each performance measure as functions of CAD use while adjusting for patient characteristics, year (1998–2002 vs 2003–2006), and BCSC registry. For specificity, the model estimated the adjusted relative odds of a true-negative mammogram among women without breast cancer diagnosed within 1 year of mammography. Similarly, for sensitivity, the model estimated the adjusted relative odds of a true-positive mammogram among women diagnosed with breast cancer within 1 year of mammography. For PPV_1 , the models estimated the adjusted relative odds of a breast cancer diagnosis within 1 year of a positive mammogram. For breast cancer detection rates, the model estimated the adjusted relative odds of a breast cancer diagnosis within 1 year of a positive mammogram among all mammograms. Patient covariates were included as categorical variables to allow adjustment for potentially nonlinear relationships between outcomes and covariates. Patient age on the mammogram date was categorized as 40–44 years, 45–49 years, 50–54 years, 55–59 years, 60–64 years, 65–69 years, 70–74 years, or 75 years or older. Breast density was categorized by BI-RADS density category as follows: almost entirely fat, scattered fibroglandular tissue, heterogeneously dense, and extremely dense. Time since previous mammography was categorized as no previous mammogram, 9–15 months, 16–20 months, 21–27 months, or 28 months or longer. Models included facility-level random effects to account for correlation of outcomes within facility and varying baseline performance within facilities independent of CAD and patient factors.

We obtained similar results with generalized estimation equations, conditional logistic regression (conditioning outcomes on facility), and after restricting the random-effects analyses to the 78.6% of mammograms that were interpreted by radiologists who interpreted mammograms at facilities before and after CAD implementation (or from 1998–2002 to 2003–2006 at non-CAD facilities). We repeated the analyses for cancer detection outcomes after excluding mammograms performed more than 12 months after CAD introduction. We performed this sensitivity analysis to address concerns that an initial increase in cancer detection rates with CAD (eg, during the first year of CAD use) may be obscured in analyses including many years of post-implementation data (34).

We assessed the potential impact of outliers by creating scatterplots and computing extreme studentized deviates of pre- to post-CAD changes in specificity and sensitivity within facilities (35), which suggested that one or both measures were outlying within four facilities. When we repeated principal regression models of specificity, sensitivity, and PPV_1 after excluding mammograms from these four facilities, the CAD-associated odds ratio (OR) for specificity was slightly attenuated, but CAD-associated

odds ratios for sensitivity and PPV₁ were essentially unchanged. We therefore report results from analyses including mammograms from all facilities.

We conducted a secondary post hoc analysis to assess whether initial performance changes with CAD subsequently moderated over time and whether performance impacts differed across facilities. Across all facilities that implemented CAD, we computed unadjusted specificity, sensitivity, and PPV₁ across three time periods encompassing the transition to CAD (pre-CAD, 3–12 months post-CAD, and >12 months post-CAD). We made similar comparisons within the subset of facilities that exhibited statistically significant declines in specificity after CAD implementation in a previous report ($n = 6$) (10) and the remaining CAD facilities that are new to this report ($n = 19$). Analyses were conducted with SAS, Version 9.2 (SAS Institute, Cary, NC). Statistical tests were two-sided and P less than .05 was considered significant.

Results

Study Samples

From January 1, 1998, to December 31, 2006, 684 956 women received more than 1.6 million screening mammograms at 90 BSCS facilities, of which 25 (27.8%) facilities adopted CAD and used it for an average of 27.5 study months. The mammograms were interpreted by 793 radiologists, including 154 (19.4%) at facilities with CAD. Of the 684 956 women included in the study, 7722 (1.1%) were diagnosed with breast cancer within 1 year of screening, including 1013 women who were diagnosed with breast cancer within 1 year of screening at facilities with CAD (13.1% of all patients). The proportion of women who used HT declined substantially across the study period in all facilities (Table 1).

Interpretive Performance

At facilities that never implemented CAD, specificity increased slightly but statistically significantly from 91.1% between 1998 and 2002 to 91.3% between 2003 and 2006 (difference = 0.2%, 95% confidence interval [CI] of the difference = 0.04% to 0.28%, $P = .008$) (Table 2). The recall rate correspondingly decreased statistically significantly from 9.3% between 1998 and 2002 to 9.1% between 2003 and 2006 (difference = -0.2%, 95% CI of the difference = -0.30% to -0.05%, $P = .005$). In contrast, at facilities that implemented CAD, specificity decreased statistically significantly from 91.9% to 91.4% after CAD implementation (difference = -0.5%, 95% CI of the difference = -0.67% to -0.39%, $P < .001$), whereas the recall rate increased from 8.4% to 8.9% (difference = 0.5%, 95% CI of the difference = 0.36% to 0.64%, $P < .001$). From the 1998–2002 to 2003–2006 time periods, sensitivity increased statistically significantly from 80.7% to 84.0% at facilities that did not implement CAD (difference = 3.3%, 95% CI of the difference = 1.1% to 5.6%, $P = .005$). However, at facilities that implemented CAD, sensitivity did not statistically significantly change from 79.7% before CAD to 81.1% after CAD implementation (difference = 1.4%, 95% CI of the difference = -1.6% to 4.5%, $P = .35$). At facilities that did not implement CAD, PPV₁ changed little across the time periods (difference = -0.1%, 95% CI of the difference = -0.4% to 0.2%, $P = .63$). However, PPV₁ statistically significantly decreased from

4.3% to 3.6% after the transition to CAD use at facilities that implemented CAD (difference = -0.7%, 95% CI of the difference = -1.0% to -0.3%, $P < .001$). The number of biopsies recommended per 1000 mammograms declined statistically significantly both at facilities that never implemented CAD (difference = -1.3 biopsy recommendations per 1000 mammograms, 95% CI of the difference = -1.8 to -0.8 biopsy recommendations per 1000 mammograms, $P < .001$) and at facilities that implemented CAD (difference = -2.2 biopsy recommendations per 1000 mammograms, 95% CI of the difference = -2.7 to -1.7 biopsy recommendations per 1000 mammograms, $P < .001$).

Breast Cancer Detection Rates and Clinical Characteristics

Between 1998–2002 and 2003–2006, the breast cancer detection rate at facilities that did not implement CAD was stable (4.2 cancers detected per 1000 mammograms in 1998–2002 vs 4.0 cancers detected per 1000 mammograms in 2003–2006), as were the detection rates of invasive breast cancer (3.3 invasive cancers detected per 1000 mammograms in 1998–2002 vs 3.2 invasive cancers detected per 1000 mammograms in 2003–2006) and DCIS (0.9 DCIS detected per 1000 mammograms in 1998–2002 vs 0.9 DCIS detected per 1000 mammograms in 2003–2006) (Table 2). Among invasive cancers diagnosed in women screened at facilities that never implemented CAD, stage distribution, tumor size, and lymph node status were similar among cancers diagnosed in 1998–2002 and 2003–2006. Among all breast cancers diagnosed in women screened at these facilities, the proportion of advanced cancers was similar in both time periods.

At facilities that implemented CAD, the overall cancer detection rate declined statistically significantly from before CAD implementation to after CAD implementation (difference = -0.4 cancers detected per 1000 mammograms, 95% CI of the difference = -0.7 to -0.1 cancers detected per 1000 mammograms, $P = .01$) (Table 2). The decline in the overall cancer detection rate was attributed mainly to a statistically significant decline in the detection rate of invasive breast cancer, whereas the detection rate of DCIS remained stable. Among invasive cancers, tumor size and lymph node status were similar before and after CAD implementation, as were the proportions of all cancers that were advanced (including both DCIS and invasive cancer). Although the stage distributions of invasive cancers were generally similar before and after CAD implementation, an overall test found a statistically significantly less favorable stage distribution after CAD implementation ($P = .036$), with a greater proportion of stage III invasive cancers and smaller proportions of stage I and II cancers compared with the period before CAD implementation.

Adjusted Analyses

After adjusting for BCSC registry, patient characteristics, HT use, and interpretation year, CAD use was associated with statistically significantly lower specificity (OR = 0.87, 95% CI = 0.85 to 0.89, $P < .001$), a non-statistically significant increase in sensitivity (OR = 1.06, 95% CI = 0.84 to 1.33, $P = .62$), and statistically significantly lower PPV₁ (OR = 0.89, 95% CI = 0.80 to 0.99, $P = .03$) (Figure 1). When the analysis of sensitivity was restricted to invasive cancers only, CAD use was no longer associated with increased sensitivity (OR = 0.96, 95% CI = 0.75 to 1.24, $P = .77$). In contrast,

Table 1. Patient characteristics by computer-aided detection (CAD) or year of mammogram*

Patient characteristic†	Facilities that never implemented CAD				Facilities that implemented CAD				Total (N = 1 621 106, 684 956 women), No. (%)‡
	1998–2002 (n = 635 083), No. (%)	2003–2006 (n = 328 287), No. (%)	1998–2002 (n = 390 543), No. (%)	2003–2006 (n = 267 193), No. (%)	Before CAD (n = 404 310), No. (%)	After CAD (n = 253 426), No. (%)‡			
Age, y									
40–44	90 323 (14.2)	39 734 (12.1)	63 623 (16.3)	36 432 (13.6)	64 393 (15.9)	35 662 (14.1)	230 112 (14.2)		
45–49	101 853 (16.0)	47 527 (14.5)	66 488 (17.0)	43 097 (16.1)	67 304 (16.6)	42 281 (16.7)	258 965 (16.0)		
50–54	111 752 (17.6)	55 043 (16.8)	66 760 (17.1)	43 949 (16.5)	67 998 (16.8)	42 711 (16.9)	277 504 (17.1)		
55–59	84 166 (13.3)	51 485 (15.7)	51 439 (13.2)	40 092 (15.0)	53 644 (13.3)	37 887 (14.9)	227 182 (14.0)		
60–64	64 140 (10.1)	37 515 (11.4)	39 917 (10.2)	30 759 (11.5)	42 270 (10.5)	28 406 (11.2)	172 331 (10.6)		
65–69	58 443 (9.2)	30 242 (9.2)	34 949 (9.0)	25 521 (9.6)	37 121 (9.2)	23 349 (9.2)	149 155 (9.2)		
70–74	53 414 (8.4)	26 715 (8.1)	30 296 (7.8)	20 933 (7.8)	32 094 (7.9)	19 135 (7.6)	131 358 (8.1)		
≥75	70 992 (11.2)	40 026 (12.2)	37 071 (9.5)	26 410 (9.9)	39 486 (9.8)	23 995 (9.5)	174 499 (10.8)		
Breast density									
Almost entirely fat	48 717 (7.7)	21 815 (6.6)	36 235 (9.3)	22 794 (8.5)	35 945 (8.9)	23 084 (9.1)	129 561 (8.0)		
Scattered fibroglandular tissue	279 472 (44.0)	147 653 (45.0)	178 098 (45.6)	130 354 (48.8)	183 681 (45.4)	124 771 (49.2)	735 577 (45.4)		
Heterogeneously dense	257 309 (40.5)	131 485 (40.1)	146 758 (37.6)	98 094 (36.7)	153 879 (38.1)	90 973 (35.9)	633 646 (39.1)		
Extremely dense	49 585(7.8)	27 334 (8.3)	29 452 (7.5)	15 951 (6.0)	30 805 (7.6)	14 598 (5.8)	122 322 (7.5)		
Previous mammography									
No previous mammogram	31 532 (5.0)	10 747 (3.3)	18 816 (4.8)	7211 (2.7)	20 313 (5.0)	5714 (2.3)	68 306 (4.2)		
9–15 mo	299 210 (47.1)	160 474 (48.9)	213 462 (54.7)	164 324 (61.5)	220 699 (54.6)	157 087 (62.0)	837 470 (51.7)		
16–20 mo	74 853 (11.8)	33 995 (10.4)	55 665 (14.3)	33 028 (12.4)	57 030 (14.1)	31 663 (12.5)	197 541 (12.2)		
21–27 mo	124 649 (19.6)	72 409 (22.1)	40 197 (10.3)	24 830 (9.3)	41 758 (10.3)	23 269 (9.2)	262 085 (16.2)		
≥28 mo	104 839 (16.5)	50 662 (15.4)	62 403 (16.0)	37 800 (14.2)	64 510 (16.0)	35 693 (14.1)	255 704 (15.8)		
Current hormone therapy user	227 761 (35.9)	45 774 (13.9)	109 419 (28.0)	49 894 (18.7)	109 094 (27.0)	50 219 (19.8)	432 848 (26.7)		

* The number of mammograms is defined by n.

† Patient characteristics on dates of screening mammograms. Patients may have received more than one mammogram.

‡ Mammograms interpreted during the initial 3 months of CAD use were excluded.

Table 2. Unadjusted performance, biopsy recommendation, breast cancer detection rates, and breast cancer prognostic characteristics by computer-aided detection (CAD) or year of mammogram*

Characteristic	Never implemented CAD			Implemented CAD		
	1998–2002	2003–2006	<i>P</i>	Before CAD	After CAD	<i>P</i>
Screening mammograms, No.						
Total	635 083	328 287	—	404 310	253 426	—
Without breast cancer	631 790	326 708	—	402 473	252 413	—
With breast cancer†	3293	1579	—	1837	1013	—
Invasive breast cancers	2671	1276	—	1469	761	—
Ductal carcinoma in situ	622	303	—	368	252	—
Performance measure, % (95% CI)						
Specificity	91.1 (91.0 to 91.2)	91.3 (91.2 to 91.4)	.008	91.9 (91.8 to 92.0)	91.4 (91.3 to 91.5)	<.001
Recall rate	9.3 (9.2 to 9.3)	9.1 (9.0 to 9.2)	.005	8.4 (8.3 to 8.5)	8.9 (8.8 to 9.0)	<.001
Sensitivity	80.7 (79.3 to 82.0)	84.0 (82.1 to 85.8)	.005	79.7 (77.8 to 81.5)	81.1 (78.6 to 83.5)	.35
Positive predictive value	4.5 (4.3 to 4.7)	4.4 (4.2 to 4.7)	.63	4.3 (4.1 to 4.5)	3.6 (3.4 to 3.9)	<.001
Breast biopsy recommendation‡						
Biopsies recommended, No.	9068	4271	—	5116	2647	—
Biopsies recommended per 1000 screens (95% CI)	14.3 (14.0 to 14.6)	13.0 (12.6 to 13.4)	<.001	12.7 (12.3 to 13.0)	10.4 (10.1 to 10.8)	<.001
Cancer detection rate per 1000 mammograms (95% CI)						
All breast cancers	4.2 (4.0 to 4.3)	4.0 (3.8 to 4.3)	.30	3.6 (3.4 to 3.8)	3.2 (3.0 to 3.5)	.01
Invasive breast cancers	3.3 (3.2 to 3.5)	3.2 (3.0 to 3.4)	.27	2.8 (2.7 to 3.0)	2.3 (2.1 to 2.5)	<.001
Ductal carcinoma in situ	0.9 (0.8 to 0.9)	0.9 (0.8 to 1.0)	.91	0.8 (0.7 to 0.9)	0.9 (0.8 to 1.0)	.13
Ductal carcinoma in situ, %	18.9	19.2	.80	20.0	24.9	.003
Invasive cancer stage (n = 5504), No. (%)						
1	1483 (63.2)	737 (62.8)	.19	773 (60.1)	405 (58.2)	.036
2	738 (31.4)	353 (30.1)	—	420 (32.6)	216 (31.0)	—
3	103 (4.4)	67 (5.7)	—	75 (5.8)	65 (9.3)	—
4	23 (1.0)	17 (1.5)	—	19 (1.5)	10 (1.4)	—
Invasive tumor size (n = 5613)						
Mean (SD), mm	18.1 (14.4)	18.0 (15.4)	.85	18.6 (14.1)	19.4 (15.0)	.25
Small (<15 mm), No. (%)	1460 (60.2)	724 (61.8)	.36	765 (58.1)	397 (56.8)	.56
Lymph node involvement of invasive cancers (n = 5845)						
Nodes negative, No. (%)	1893 (74.4)	904 (74.6)	.92	993 (72.5)	531 (73.7)	.59
Advanced cancers (n = 7521), No. (%)	1247 (38.8)	570 (37.3)	.31	712 (40.0)	369 (37.1)	.14

* CI = confidence interval; SD = standard deviation. *P* values are from two-sided χ^2 or paired *t*-tests for categorical or continuous outcomes respectively.

† Includes invasive breast cancer and ductal carcinomas in situ.

‡ Radiologist recommended biopsy, surgical consultation, or fine needle aspiration within 90 days of the screening mammogram.

when the analysis was restricted to DCIS, CAD use was associated with greater sensitivity (OR = 1.55, 95% CI = 0.83 to 2.91, *P* = .17), although the association was non-statistically significant.

Furthermore, after adjusting for BCSC registry, patient characteristics, HT use, and interpretation year, CAD use was not associated with statistically significant differences in the odds of biopsy recommendation (OR = 0.99, 95% CI = 0.93 to 1.05, *P* = .66), overall breast cancer detection (OR = 1.01, 95% CI = 0.92 to 1.12, *P* = .79), the detection of invasive breast cancer (OR = 0.97, 95% CI = 0.86 to 1.08, *P* = .54), the diagnosis of stage I invasive cancer compared with later stage invasive cancer (OR = 0.90, 95% CI = 0.73 to 1.11, *P* = .34), or the diagnosis of invasive tumors of 15 mm or less in size compared with greater than 15 mm (OR = 0.92, 95% CI = 0.74 to 1.15, *P* = .46) (Figure 1). Also, CAD use was not associated with statistically significant differences in the diagnosis of invasive tumors without vs with nodal involvement (OR = 1.05, 95% CI = 0.84 to 1.30, *P* = .68), or the diagnosis of advanced vs non-advanced breast cancer (OR = 0.96, 95%

CI = 0.80 to 1.15, *P* = .67) (Figure 1). However, CAD was associated with a non-statistically significant increase in the rate of detection of DCIS (OR = 1.16, 95% CI = 0.95 to 1.42; *P* = .14). In an analysis that excluded mammograms performed more than 12 months after CAD implementation, there remained no statistically significant association between CAD and the detection of any breast cancer (OR = 1.06, 95% CI = 0.93 to 1.22, *P* = .36).

Variability in CAD Impacts by Time and Facility

In post hoc analyses assessing variability in CAD impacts over time and across the 25 facilities that implemented CAD, specificity declined from 91.9% to 91.5% from pre-CAD compared with 3–12 months post-CAD (difference = -0.4%, 95% CI of the difference = -0.7 to -0.3%, *P* < .001) and remained similarly decreased more than 12 months after CAD implementation (Table 3). Although sensitivity increased from 79.7% pre-CAD to 83.1% during the 3–12 months after CAD implementation, the change was non-statistically significant (difference = 3.4%, 95% CI of the

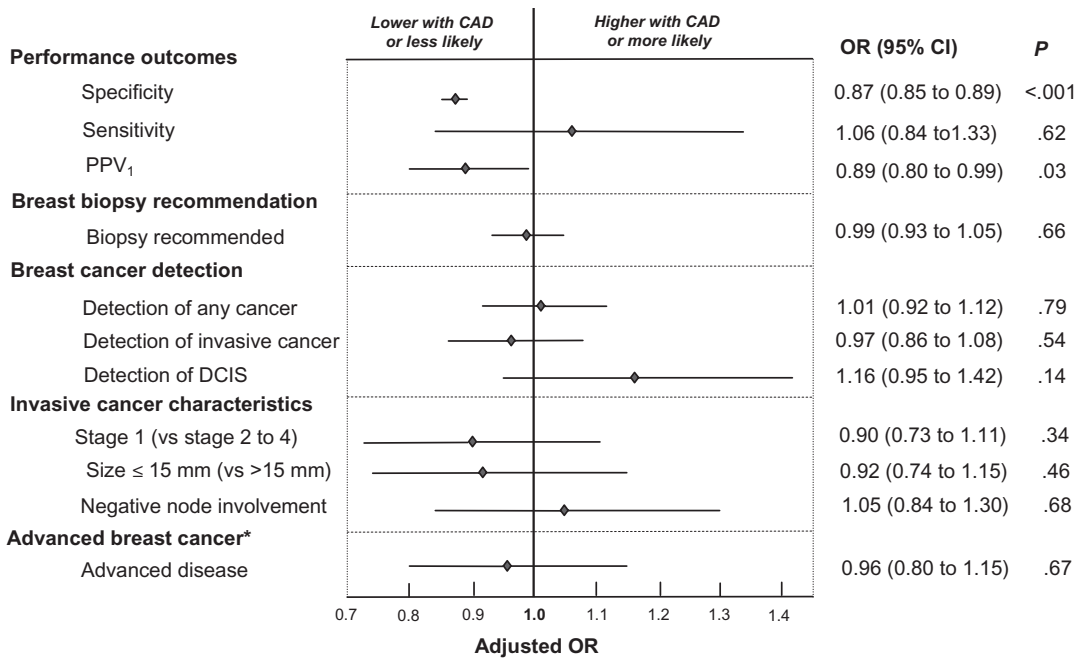


Figure 1. Adjusted association between Computer-Aided Detection (CAD) and performance, biopsy recommendation, and breast cancer prognostic characteristics. Odds ratios (ORs) and 95% confidence intervals (CIs) of CAD use vs non-CAD use as estimated using random-effects logistic regression are presented. The 95% confidence intervals are represented by error bars. Odds ratios were adjusted for mammography registry, patient age (40–44, 45–49, 50–54, 55–59, 60–64, 65–69, 70–74, and ≥75 years), breast density (almost entirely fat, scattered fibroglandular tissue, and heterogeneously and extremely dense), time since prior mammography (no prior mammogram, 9–15 months, 16–20 months, 21–27 months, and ≥28 months), current hormone replacement therapy, and year of examination (1998–2002 or 2003–2006). Analyses for each outcome include the following numbers of mammograms and women: specificity (n = 1 613 384 mammograms among 681 421 women

without breast cancer); sensitivity (n = 7722 mammograms among 7722 women with breast cancer); positive predictive value (PPV₁) (n = 145 293 positive mammograms among 127 647 women); biopsy recommendations and breast cancer detection (n = 1 621 106 mammograms among 684 956 women); invasive cancer outcomes (n = 6177 mammograms among 6177 women with invasive cancers, excluding from 332 [5.4%] to 673 [10.9%] because of missing stage, size, or node involvement data). The model for advanced cancer includes both invasive cancers and ductal carcinomas in situ (DCIS). Advanced cancer (*) was defined as invasive breast cancers with either stage II, III, or IV; tumor size greater than 15 mm; or positive node involvement. Variance of facility-level random-effects was statistically significantly different from zero in all models (P < .01), except for models of stage I invasive breast cancer (P = .02) and negative node involvement (P = .99).

difference = -0.9% to 7.8%, P = .14), and sensitivity after 12 months of CAD use was similar to the pre-CAD period. PPV₁ declined from 4.3% in the pre-CAD period to 3.7% during the period of 3–12 months post-CAD implementation (difference = -0.6%, 95% CI of the difference = -1.1% to -0.1%, P = .02) and remained similarly decreased after 12 months of CAD use.

The interpretive impact of CAD, however, varied across facilities. Of seven CAD facilities included in a previous report (10), six were included in the current study, whereas one was excluded because it averaged no more than 100 mammograms per month. Among these six facilities (Group 1), specificity decreased from 90.3% pre-CAD to 87.5% during the 3–12 months post-CAD (difference = -2.8%, 95% CI of the difference = -3.2% to -2.3%, P < .001) before increasing to 88.8% after 12 months of CAD use (difference vs 3–12 months post-CAD = 1.3%, 95% CI of the difference = 0.8% to 1.8%, P < .001) (Table 3). At these six facilities, sensitivity increased statistically significantly from 79.8% pre-CAD to 88.9% during the 3–12 months after CAD implementation (difference = 9.1%, 95% CI of the difference = 2.2% to 16.0%, P = .02) but then decreased to 81.6% after 12 months of CAD use (difference vs pre-CAD period = 1.8%, 95% CI of the difference = -4.3% to 7.8%, P = .57). PPV₁ decreased at these facilities from 4.1% pre-CAD to 3.6% during

the 3–12 months post-CAD (difference = -0.5%, 95% CI of the difference = -1.3% to 0.4%, P = .30) but decreased further to 3.0% after 12 months of CAD implementation (difference vs 3–12 months = -0.6%, 95% CI of the difference = -1.4% to 0.2%, P = .11). Compared with the pre-CAD period, PPV₁ was statistically significantly lower after 12 months of CAD use at these facilities (difference = -1.1%, 95% CI of the difference -1.7% to -0.5%, P < .001).

In contrast, among 19 other facilities that implemented CAD (Group 2), specificity, sensitivity, and PPV₁ changed little from the pre- to the post-CAD period. Indeed, unadjusted specificity increased slightly from 92.3% pre-CAD to 92.8% during the 3–12 months after CAD implementation (difference = 0.5%, 95% CI of the difference = 0.2% to 0.7%, P < .001) and remained similarly increased (92.9%) after 12 months of CAD use (difference vs pre-CAD period = 0.6%, 95% CI of the difference = 0.4% to 0.8%, P < .001).

Discussion

Within a large geographically diverse sample of mammography facilities, CAD was associated with statistically significant decreases in specificity and PPV₁, a non-statistically significant increase in

Table 3. Unadjusted performance at facilities by months since implementation of Computer-Aided Detection (CAD)*

Sample size or performance outcome	All CAD facilities, 1998–2006 (n = 25)						Group 1 (n = 6)†			Group 2 (n = 19)‡		
	Pre-CAD	3–12 mo post-CAD	>12 mo post-CAD	Pre-CAD	3–12 mo post-CAD	>12 mo post-CAD	Pre-CAD	3–12 mo post-CAD	>12 mo post-CAD	Pre-CAD	3–12 mo post-CAD	>12 mo post-CAD
	No. of total mammograms	404 310	88 556	164 870	79 465	22 121	62 480	324 845	66 435	102 390	324 845	66 435
No. of mammograms for women with cancer	1837	350	663	411	117	266	1426	233	397	1426	233	397
Specificity, % (95% CI)	91.9 (91.8 to 92.0)	91.5 (91.3 to 91.7)	91.4 (91.2 to 91.5)	90.3 (90.1 to 90.5)	87.5 (87.1 to 87.9)	88.8 (88.5 to 89.0)	92.3 (92.2 to 92.4)	92.8 (92.6 to 93.0)	92.9 (92.8 to 93.1)	92.3 (92.2 to 92.4)	92.8 (92.6 to 93.0)	92.9 (92.8 to 93.1)
Sensitivity, % (95% CI)	79.7 (77.8 to 81.5)	83.1 (78.8 to 86.9)	80.1 (76.8 to 83.1)	79.8 (75.6 to 83.6)	88.9 (81.7 to 93.9)	81.6 (76.4 to 86.1)	79.7 (77.5 to 81.7)	80.3 (74.6 to 85.2)	79.1 (74.8 to 83.0)	79.7 (77.5 to 81.7)	80.3 (74.6 to 85.2)	79.1 (74.8 to 83.0)
Positive predictive value, % (95% CI)	4.3 (4.1 to 4.5)	3.7 (3.3 to 4.2)	3.6 (3.3 to 3.9)	4.1 (3.7 to 4.5)	3.6 (3.0 to 4.4)	3.0 (2.6 to 3.4)	4.4 (4.1 to 4.6)	3.8 (3.3 to 4.3)	4.2 (3.7 to 4.6)	4.4 (4.1 to 4.6)	3.8 (3.3 to 4.3)	4.2 (3.7 to 4.6)

* CI = confidence interval.

† Includes Breast Cancer Surveillance Consortium facilities from a previous report that implemented CAD from 1998 to 2002.

‡ Includes additional Breast Cancer Surveillance Consortium facilities that implemented CAD from 1998 to 2006.

sensitivity, but no statistically significant improvements in either cancer detection rates or the prognostic characteristics of incident breast cancers. The non-statistically significant increase in sensitivity was attributed to greater detection of DCIS with CAD rather than increased detection of invasive breast cancer.

This study offers important new insights regarding the effectiveness of CAD when used in real-world practice. First, the results suggest a limited impact of CAD on breast cancer detection, particularly with respect to invasive breast cancer detection. In adjusted analyses, CAD was not associated with improved detection of invasive breast cancer, increased early stage diagnosis, or smaller-sized invasive breast cancers. These findings raise concerns that CAD, as currently implemented in clinical practice, may have little or no impact on breast cancer mortality, which may depend on earlier detection of invasive breast cancer (6,7). Second, this study of CAD adoption within 25 BCSC facilities indicates that CAD has a modest impact in typical practice compared with a previous analysis within seven BCSC facilities from 1998 to 2002 (10). The current findings are consistent with a meta-analysis that suggests a modest increase in recall rates with CAD with little or no impact on cancer detection rates (17).

Nishikawa and Pesce (33) argue that studies with matched designs are the most accurate means of assessing CAD clinical impacts (ie, in which outcomes are compared before and after CAD application on the same mammogram) (34). However, matched studies typically impose the restriction that radiologists can only upgrade final BI-RADS assessments after viewing CAD output, resulting in the recall of women who would not have been recalled in the absence of CAD (11,13,36,37). In our view, this design assesses the efficacy of CAD, or its clinical impact when used under optimal conditions. In contrast, this study assesses the effectiveness of CAD use in actual everyday practice conditions, in which radiologists with variable experience and expertise may use CAD in a nonstandardized idiosyncratic fashion (5). Some community radiologists, for example, may decide not to recall women because of the absence of CAD marks on otherwise suspicious lesions.

This large-scale population-based observational study also enabled assessment of the impact of CAD on important breast cancer outcomes such as DCIS detection and the stage distribution of invasive cancers. These outcomes may be impossible to assess with adequate statistical power within matched studies or even randomized trials (38). Although our analyses may lack sufficient power to exclude a small benefit of CAD in terms of invasive breast cancer detection, the principal contribution of CAD may be increased detection for DCIS—a precancerous lesion with an ill-defined long-term prognosis (39). Point estimates and 95% confidence intervals from this study may be useful in statistical models of long-term breast cancer outcomes among women screened with and without CAD (40). Such models could quantify the potential for CAD to improve breast cancer mortality, possible overdiagnosis of DCIS, patient preferences for earlier treatment of DCIS vs later treatment of invasive cancer, the harms of additional false-positive mammograms, and societal costs (including both supplemental fees for CAD use and the costs of added diagnostic testing).

We performed post hoc analyses to explore potential heterogeneity across facilities and over time. Because the analyses were

conducted post hoc, these results should be interpreted cautiously, particularly the resultant confidence intervals and *P* values. These analyses suggest that substantial initial changes in specificity and sensitivity after CAD implementation subsequently attenuated within six BCSC facilities that were included in a previous report (10). The performance impact of CAD at these facilities may have diminished as radiologists gained experience with CAD. Similar adjustment effects, however, were not apparent within 19 other CAD facilities that are reported on here for the first time. It is possible that more recent versions of CAD software have a less potent influence on interpretation, although a recent comparative study suggests that earlier compared with later versions of CAD software do not differ greatly in performance (41). It is also possible that training in CAD use may vary across radiologists and facilities, leading to variable interpretive impacts of CAD. In addition, overall performance changes associated with CAD may be attributed to substantive impacts within some facilities, whereas CAD may have little interpretive impact when implemented in most community facilities. Variability in the impact of CAD on recall rates was also observed in a meta-analysis (17) and across three sites in a trial comparing mammogram interpretation by a single radiologist with CAD versus interpretation by two radiologists (double reading) (42).

Although prior analysis of BCSC data indicated increased biopsy risk with CAD (10), this analysis reveals a decline in the rate of biopsy recommendation over time regardless of CAD use. In the Women's Health Initiative, women randomly assigned to HT had nearly two-thirds greater cumulative breast biopsy risk (29), and reduced HT use after publication of the Women's Health Initiative results in July 2002 may largely explain the observed reduction in biopsy recommendations (32). Biopsy recommendations declined despite reduced specificity within facilities that implemented CAD, suggesting that most CAD-induced recalls were resolved without biopsy.

A limitation of this study is the absence of digital mammography data. Whereas CAD algorithms perform a similar alerting function in the film-screen and digital environments, film-screen mammograms must be digitized before CAD analysis, and digitization may introduce noise and adversely affect performance. However, small retrospective studies suggest that the performance impacts of CAD are similar when used in digital (43–46) and film-screen environments (8,11,19).

Because prior research suggests that facilities apply CAD on nearly all mammograms after implementation (10), these analyses assumed that all mammograms were interpreted with CAD after implementation—another limitation of this study. To the extent that facilities did not use CAD on all mammograms, results may be biased toward the null. As the analyses account for salient patient factors, unmeasured radiologist or facility characteristics may affect results. Results were similar, however, in analyses in which CAD effects were conditional on the basis of the facility, which would control to some extent for potentially confounding facility factors. Similarly, results were unchanged when analyses were restricted to mammograms interpreted by radiologists who were present at facilities before and after CAD implementation, therefore controlling to some extent for changes in radiologist staffing over time. Although the number of women with breast cancers

diagnosed after CAD implementation (>1000 cancers) is greater than that observed in previous samples, larger samples may be needed to detect small increases in sensitivity or cancer detection with CAD. Finally, another limitation of this study is the lack of data on the CAD products that each facility used, so the potentially distinct impacts of different products could not be investigated.

Among a large sample of US mammography facilities, CAD was associated with statistically significantly decreased specificity and PPV₁. CAD was not associated with improved sensitivity for invasive breast cancer, increased rates of breast cancer detection, or more favorable stage or size of invasive breast cancers. CAD is now applied to the large majority of screening mammograms in the United States with annual direct Medicare costs exceeding \$30 million (2). As currently implemented in US practice, CAD appears to increase a woman's risk of being recalled for further testing after screening mammography while yielding equivocal health benefits.

References

1. Chan HP, Doi K, Vyborny CJ, Lam KL, Schmidt RA. Computer-aided detection of microcalcifications in mammograms. Methodology and preliminary clinical study. *Invest Radiol*. 1988;23(9):664–671.
2. Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? *J Am Coll Radiol*. 2010;7(10):802–805.
3. Harvey D. Navigating through the data—CAD applications in liver, prostate, breast, and lung cancer. *Radiol Today*. 2010;11(3):18.
4. Birdwell RL. The preponderance of evidence supports computer-aided detection for screening mammography. *Radiology*. 2009;253(1):9–16.
5. Philpotts LE. Can computer-aided detection be detrimental to mammographic interpretation? *Radiology*. 2009;253(1):17–22.
6. Duffy SW, Tabar L, Vitak B, et al. The relative contributions of screen-detected in situ and invasive breast carcinomas in reducing mortality from the disease. *Eur J Cancer*. 2003;39(12):1755–1760.
7. Autier P, Hery C, Haukka J, Boniol M, Byrnes G. Advanced breast cancer and breast cancer mortality in randomized controlled trials on mammography screening. *J Clin Oncol*. 2009;27(35):5919–5923.
8. Birdwell RL, Ikeda DM, O'Shaughnessy KF, Sickles EA. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology*. 2001;219(1):192–202.
9. Brem RF, Rapelyea JA, Zisman G, Hoffmeister JW, Desimio MP. Evaluation of breast cancer with a computer-aided detection system by mammographic appearance and histopathology. *Cancer*. 2005;104(5):931–935.
10. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med*. 2007;356(14):1399–1409.
11. Freer TW, Ulisse MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*. 2001;220(3):781–786.
12. Gromet M. Comparison of computer-aided detection to double reading of screening mammograms: review of 231,221 mammograms. *AJR Am J Roentgenol*. 2008;190(4):854–859.
13. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: interpretation with computer-aided detection—prospective evaluation. *Radiology*. 2006;239(2):375–383.
14. Alberdi E, Povyakalo AA, Strigini L, et al. Use of computer-aided detection (CAD) tools in screening mammography: a multidisciplinary investigation. *Br J Radiol*. 2005;78(Spec No 1):S31–S40.
15. Taplin SH, Rutter CM, Lehman C. Testing the effect of computer-assisted detection on interpretive performance in screening mammography. *AJR Am J Roentgenol*. 2006;187(6):1475–1482.
16. Noble M, Bruening W, Uhl S, Schoelles K. Computer-aided detection mammography for breast cancer screening: systematic review and meta-analysis. *Arch Gynecol Obstet*. 2009;279(6):881–890.

17. Taylor P, Potts HW. Computer aids and human second reading as interventions in screening mammography: two systematic reviews to compare effects on cancer detection and recall rate. *Eur J Cancer*. 2008; 44(6):798–807.
18. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. *Radiology*. 2005; 236(2):451–457.
19. Warren Burhenne LJ, Wood SA, D’Orsi CJ, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology*. 2000;215(2):554–562.
20. 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(4):1001–1008.
21. Rosenberg RD, Yankaskas BC, Abraham LA, et al. Performance benchmarks for screening mammography. *Radiology*. 2006;241(1):55–66.
22. Taplin SH, Ichikawa LE, Kerlikowske K, et al. Concordance of breast imaging reporting and data system assessments and management recommendations in screening mammography. *Radiology*. 2002;222(2):529–535.
23. D’Orsi CJ, Bassett LW, Feig SA, et al. *American College of Radiology (ACR): Breast Imaging Reporting and Data System (BI-RADS)*. Reston, VA: American College of Radiology; 1998.
24. *AJCC Cancer Staging Manual*. 3rd ed. New York, NY: Springer; 1988.
25. Ries LAG, Melbert D, Krapcho M, et al., eds. *SEER Cancer Statistics Review, 1975-2004*. Bethesda, MD: National Cancer Institute; 2007. http://www.seer.cancer.gov/csr/1975_2004/#citation. Accessed June 7, 2011
26. Rosenberg J, Chia YL, Plevritis S. The effect of age, race, tumor size, tumor grade, and disease stage on invasive ductal breast cancer survival in the U.S. SEER database. *Breast Cancer Res Treat*. 2005;89(1):47–54.
27. 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(3):168–175.
28. Yankaskas BC, Taplin SH, Ichikawa L, et al. Association between mammography timing and measures of screening performance in the United States. *Radiology*. 2005;234(2):363–373.
29. Chlebowski RT, Anderson G, Pettinger M, et al. Estrogen plus progestin and breast cancer detection by means of mammography and breast biopsy. *Arch Intern Med*. 2008;168(4):370–377. quiz 45.
30. Chlebowski RT, Kuller LH, Prentice RL, et al. Breast cancer after use of estrogen plus progestin in postmenopausal women. *N Engl J Med*. 2009; 360(6):573–587.
31. Kerlikowske K, Miglioretti DL, Buist DS, Walker R, Carney PA. Declines in invasive breast cancer and use of postmenopausal hormone therapy in a screening mammography population. *J Natl Cancer Inst*. 2007;99(17):1335–1339.
32. Hersh AL, Stefanick ML, Stafford RS. National use of postmenopausal hormone therapy: annual trends and response to recent evidence. *JAMA*. 2004;291(1):47–53.
33. Ravdin PM, Cronin KA, Howlader N, et al. The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med*. 2007;356(16):1670–1674.
34. Nishikawa RM, Pesce LL. Computer-aided detection evaluation methods are not created equal. *Radiology*. 2009;251(3):634–636.
35. Rosner B. Percentage points for a generalized ESD many-outlier procedure. *Technometrics*. 1983;25(2):165–172.
36. Dean JC, Ilvento CC. Improved cancer detection using computer-aided detection with diagnostic and screening mammography: prospective study of 104 cancers. *AJR Am J Roentgenol*. 2006;187(1):20–28.
37. Ko JM, Nicholas MJ, Mendel JB, Slanetz PJ. Prospective assessment of computer-aided detection in interpretation of screening mammography. *AJR Am J Roentgenol*. 2006;187(6):1483–1491.
38. Jiang Y, Miglioretti DL, Metz CE, Schmidt RA. Breast cancer detection rate: designing imaging trials to demonstrate improvements. *Radiology*. 2007;243(2):360–367.
39. Ernster VL, Barclay J, Kerlikowske K, Wilkie H, Ballard-Barbash R. Mortality among women with ductal carcinoma in situ of the breast in the population-based surveillance, epidemiology and end results program. *Arch Intern Med*. 2000;160(7):953–958.
40. Fryback DG, Stout NK, Rosenberg MA, Trentham-Dietz A, Kuruchittham V, Remington PL. The Wisconsin Breast Cancer Epidemiology Simulation Model. *J Natl Cancer Inst Monogr*. 2006;36(3):37–47.
41. Kim SJ, Moon WK, Kim SY, Chang JM, Kim SM, Cho N. Comparison of two software versions of a commercially available computer-aided detection (CAD) system for detecting breast cancer. *Acta Radiol*. 2010; 51(5):482–490.
42. Gilbert FJ, Astley SM, Gillan MG, et al. Single reading with computer-aided detection for screening mammography. *N Engl J Med*. 2008;359(16):1675–1684.
43. Brancato B, Houssami N, Francesca D, et al. Does computer-aided detection (CAD) contribute to the performance of digital mammography in a self-referred population? *Breast Cancer Res Treat*. 2008;111(2):373–376.
44. Obenauer S, Sohns C, Werner C, Grabbe E. Computer-aided detection in full-field digital mammography: detection in dependence of the BI-RADS categories. *Breast J*. 2006;12(1):16–19.
45. Skaane P, Kshirsagar A, Stapleton S, Young K, Castellino RA. Effect of computer-aided detection on independent double reading of paired screen-film and full-field digital screening mammograms. *AJR Am J Roentgenol*. 2007;188(2):377–384.
46. Yang SK, Moon WK, Cho N, et al. Screening mammography-detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. *Radiology*. 2007;244(1):104–111.

Funding

This work was supported by the National Cancer Institute–funded Breast Cancer Surveillance Consortium co-operative agreement (grants U01CA63740, U01CA86076, U01CA86082, U01CA63736, U01CA70013, U01CA69976, U01CA63731, and U01CA70040), the National Cancer Institute (grant CA104699 to J.G.E.), and the American Cancer Society (grant MRSRG-05-214-01-CPPB to J.J.F.). The collection of cancer data used in this study was supported in part by several state public health departments and cancer registries throughout the U.S. For a full description of these sources, please see: <http://breastscreening.cancer.gov/work/acknowledgement.html>.

Notes

The funders did not have any involvement in the design of the study; the collection, analysis, and interpretation of the data; the writing of the article; or the decision to submit the article for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health. We thank the participating women, mammography facilities, and radiologists for the data they have provided for this study. A list of the BCSC investigators and procedures for requesting BCSC data for research purposes are provided at: <http://breastscreening.cancer.gov/>.

SHT was an employee of the National Cancer Institute (NCI) during the writing of this article but any opinions are his alone and not reflective of the opinions of the NCI or the US Federal Government. CD holds stock options in and serves as a medical consultant for Hologic, Inc (a manufacturer of CAD equipment), which had no role in data collection, analysis, or interpretation.

Affiliations of authors: Department of Family and Community Medicine and Center for Healthcare Policy and Research, University of California, Davis, Sacramento, CA (JJF); Group Health Research Institute, Seattle, WA (LA); Applied Research Program, Division of Cancer Control and Population Sciences, National Cancer Institute, Bethesda, MD (SHT); Department of Family Medicine, Office of Health Promotion Research, University of Vermont, Burlington, VT (BMG); Departments of Family Medicine and Public Health and Preventive Medicine, Oregon Health and Science University, Portland, OR (PAC); Department of Radiology, Emory University, Atlanta, GA (CD); Department of Medicine and Department of Epidemiology, University of Washington, Seattle, WA (JGE); Cancer Research and Biostatistics, Seattle, WA (WEB).