Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection

Constance D. Lehman, MD, PhD; Robert D. Wellman, MS; Diana S. M. Buist, PhD; Karla Kerlikowske, MD; Anna N. A. Tosteson, ScD; Diana L. Miglioretti, PhD; for the Breast Cancer Surveillance Consortium

IMPORTANCE After the US Food and Drug Administration (FDA) approved computer-aided detection (CAD) for mammography in 1998, and the Centers for Medicare and Medicaid Services (CMS) provided increased payment in 2002, CAD technology disseminated rapidly. Despite sparse evidence that CAD improves accuracy of mammographic interpretations and costs over $400 million a year, CAD is currently used for most screening mammograms in the United States.

OBJECTIVE To measure performance of digital screening mammography with and without CAD in US community practice.

DESIGN, SETTING, AND PARTICIPANTS We compared the accuracy of digital screening mammography interpreted with (n = 495,818) vs without (n = 129,807) CAD from 2003 through 2009 in 323,973 women. Mammograms were interpreted by 271 radiologists from 66 facilities in the Breast Cancer Surveillance Consortium. Linkage with tumor registries identified 3159 breast cancers in 323,973 women within 1 year of the screening.

MAIN OUTCOMES AND MEASURES Mammography performance (sensitivity, specificity, and screen-detected and interval cancers per 1000 women) was modeled using logistic regression with radiologist-specific random effects to account for correlation among examinations interpreted by the same radiologist, adjusting for patient age, race/ethnicity, time since prior mammogram, examination year, and registry. Conditional logistic regression was used to compare performance among 107 radiologists who interpreted mammograms both with and without CAD.

RESULTS Screening performance was not improved with CAD on any metric assessed. Mammography sensitivity was 85.3% (95% CI, 83.6%-86.9%) with and 87.3% (95% CI, 84.5%-89.7%) without CAD. Specificity was 91.6% (95% CI, 91.0%-92.2%) with and 91.4% (95% CI, 90.6%-92.0%) without CAD. There was no difference in cancer detection rate (4.1 in 1000 women screened with and without CAD). Computer-aided detection did not improve intraradiologist performance. Sensitivity was significantly decreased for mammograms interpreted with vs without CAD in the subset of radiologists who interpreted mammograms both with and without CAD (odds ratio, 0.53; 95% CI, 0.29-0.97).

CONCLUSIONS AND RELEVANCE Computer-aided detection does not improve diagnostic accuracy of mammography. These results suggest that insurers pay more for CAD with no established benefit to women.
Computer-aided detection (CAD) for mammography is intended to assist radiologists in identifying subtle cancers that might otherwise be missed. Computer-aided detection marks potential areas of concern on the mammogram, and the radiologist determines whether the area warrants further evaluation. Although CAD for mammography was approved by the US Food and Drug Administration (FDA) in 1998, by 2001, less than 5% of screening mammograms were interpreted with CAD in the United States. However, in 2002, the Centers for Medicare and Medicaid Services (CMS) increased reimbursement for CAD, and by 2008, 74% of all screening mammograms in the Medicare population were interpreted with CAD.2,3

Measuring the true impact of CAD on the accuracy of mammographic interpretation has proved challenging. Findings on potential benefits and harms are inconsistent and contradictory.4-19 Study designs include reader studies4-7 of enriched case sets; prospective “sequential reading” clinical studies8-12 in which a radiologist records a mammogram interpretation without CAD assistance, then immediately reviews and records an interpretation with CAD assistance; and retrospective observational studies13-16 using historical controls. One large European trial17 used a randomized clinical trial design to compare mammographic interpretations by a single reader with CAD compared with double readings without CAD.

Comparisons of mammography interpretations with vs without CAD in US community practice have not supported improved performance with CAD.18,19 However, these studies were limited by relatively small numbers and a focus on older women. Another limitation was that CAD technology was studied relatively early in its adoption, so examinations were interpreted during the early part of radiologists’ learning curves and included examinations with outdated film screen mammography. Our study addresses these limitations by using a large database of more than 495 000 full-field digital screening mammograms interpreted with CAD, accounting for radiologists’ early learning curves, and adjusting for patient and radiologist variables. We also assessed performance within a subset of radiologists who interpreted with and without CAD during the study period.

Methods

Data Source

Data were pooled from 5 mammography registries that participate in the Breast Cancer Surveillance Consortium (BCSC)20 funded by the National Cancer Institute: (1) the San Francisco Mammography Registry, (2) the New Mexico Mammography Advocacy Project, (3) the Vermont Breast Cancer Surveillance System, (4) the New Hampshire Mammography Network, and (5) the Carolina Mammography Registry. Each mammography registry links women to a state tumor registry or regional Surveillance Epidemiology and End Results program that collects population-based cancer data. Each registry and the BCSC Statistical Coordinating Center have institutional review board approval for either active or passive consenting processes or a waiver of consent to enroll participants, link data, and perform analytic studies. All procedures are Health Insurance Portability and Accountability Act compliant, and all registries and the Statistical Coordinating Center have received a Federal Certificate of Confidentiality and other protection for the identities of women, physicians, and facilities that participate in this research.

Participants

We included digital screening mammography examinations interpreted by 271 radiologists with (n = 495 818) or without CAD (n = 129 807) between January 1, 2003, and December 31, 2009, among 323 973 women aged 40 to 89 years with information on race, ethnicity, and time since last mammogram. Of the radiologists, 82 never used CAD, 82 always used CAD, and 107 sometimes used CAD. The latter 107 radiologists contributed 45 990 examinations interpreted without using CAD and 337 572 interpreted using CAD. The median percentage of examinations interpreted using CAD among the 107 radiologists was 93%, and the interquartile range was 31%.

Data Collection

Methods used to identify and assess screening mammograms, patient characteristics, and outcomes have been described previously.20,21 Briefly, screening mammograms were defined as bilateral mammograms designated as “routine screening” by the radiologist. Mammographic assessments followed the Breast Imaging Reporting and Data System (BI-RADS) of 0, additional imaging; 1, negative; 2, benign finding; 3, probably benign finding; 4, suspicious abnormality; or 5, abnormality highly suspicious for malignant neoplasm.

Woman-level characteristics including menopausal status, race/ethnicity, and first-degree family history were captured through self-administered questionnaires at each examination. Breast density was recorded by the radiologist at the time of the mammogram using the BI-RADS standard terminology of almost entirely fat, scattered fibroglandular densities, heterogeneously dense, and extremely dense.

Outcomes

We calculated sensitivity, specificity, cancer detection rates, and interval cancer rates. We defined positive mammograms as those with BI-RADS assessments of 0, 4 or 5, or 3 with a recommendation for immediate follow-up. Negative mammogram results were defined as BI-RADS assessments 1 or 2, or 3 without a recommendation for immediate follow-up. All women were followed for breast cancer from their mammogram up until their next screening mammogram or 12 months, whichever came first. Breast cancer diagnoses included ductal carcinoma in situ (DCIS) or invasive breast cancer within this follow-up period.

False-negative examination results were defined as mammograms with a negative assessment but a breast can-
Mammography performance measures were modeled using logistic regression, including normally distributed, radiologist-specific random effects to account for the correlation among examinations read by the same radiologist. Random effects were allowed to vary by CAD use or nonuse during the reading. Performance measures were estimated at the median of the random effects distribution. Adjusted, radiologist-specific relative performance was measured by an odds ratio (OR) with 95% CIs comparing CAD use to no CAD, adjusting for patient age at diagnosis, time since last mammogram and year of examination, and the BCSC registry.

Receiver operating characteristic (ROC) curves were estimated from 135 radiologists who interpreted at least 1 mammogram associated with a cancer diagnosis using a hierarchical logistic regression model that allowed the threshold and accuracy parameters to depend on whether CAD was used during examination interpretation. We assumed a constant accuracy among radiologists for examinations interpreted under the same condition (with or without CAD) and allowed the threshold for recall to vary across radiologists through normally distributed, radiologist-specific random effects that varied by whether the radiologist used CAD during the reading. We estimated the normalized partial area under the summary ROC curves across the observed range of false-positive rates from this model. We plotted the true-positive rate vs the false-positive rate and superimposed the estimated ROC curves.

Two separate main sensitivity analyses were conducted in subsets of total examinations: (1) to account for a possible learning curve for using CAD, we excluded the first year of each radiologist’s CAD use; and (2) to estimate the within-radiologist effect of CAD, we limited analysis to the 107 radiologists who interpreted mammograms during the study period with and without CAD, using conditional logistic regression and adjusting for patient age, time since last mammogram, and race/ethnicity.

Two-sided statistical tests were used with P < .05 considered statistically significant. All analyses were conducted by one of us (R.D.W.) using SAS statistical software (version 9.2; SAS Institute Inc for Windows 7).

Results

Increase in Digital Screening Mammography and CAD Use

Digital screening mammography and CAD use increased from 2000 to 2012. In 2003, only 5% of all screening mammograms in the BCSC were digital with CAD; by 2012, 83% of all screening mammograms were acquired digitally and interpreted with CAD assistance (Figure 1).

Among 323,973 women ages 40 to 89 years, 625,625 digital screening mammography examinations were performed (495,818 interpreted with CAD and 129,807 without CAD) between 2003 and 2009 by 271 radiologists. Breast cancer was diagnosed in 3159 women within 12 months of the screening mammogram and prior to the next screening mammogram. Women undergoing screening mammography with and without CAD assistance were similar in age, characteristics of examinations, and mammographic density and assessments were computed separately by CAD use vs no use.

We evaluated the diffusion of digital screening mammography with and without CAD in the larger BCSC population from 2002 through 2012 including 5.2 million screening mammograms.

Statistical Analysis

All analyses were conducted using the screening examination as the unit of analysis and allowing women to contribute multiple examinations during the study period; however, only 1 screening examination was associated with a breast cancer diagnosis. Distributions of breast cancer risk factors, demographic characteristics of examinations, and mammographic density and assessments were computed separately by CAD use vs no use.

Data are provided from the larger BCSC population including all screening mammograms (5.2 million mammograms) for the indicated time period.

Figure 1. Screening Mammography Patterns From 2000 to 2012 in US Community Practices in the Breast Cancer Surveillance Consortium (BCSC)

Two-sided statistical tests were used with P < .05 considered statistically significant. All analyses were conducted by one of us (R.D.W.) using SAS statistical software (version 9.2; SAS Institute Inc for Windows 7).

Results

Increase in Digital Screening Mammography and CAD Use

Digital screening mammography and CAD use increased from 2000 to 2012. In 2003, only 5% of all screening mammograms in the BCSC were digital with CAD; by 2012, 83% of all screening mammograms were acquired digitally and interpreted with CAD assistance (Figure 1).

Among 323,973 women ages 40 to 89 years, 625,625 digital screening mammography examinations were performed (495,818 interpreted with CAD and 129,807 without CAD) between 2003 and 2009 by 271 radiologists. Breast cancer was diagnosed in 3159 women within 12 months of the screening mammogram and prior to the next screening mammogram. Women undergoing screening mammography with and without CAD assistance were similar in age,
menopausal status, family history of breast cancer, time since last mammogram, and breast density. Women undergoing screening mammography with CAD were more likely to be non-Hispanic white than women whose mammograms were interpreted without CAD (Table 1).

### Performance Measures for Mammography Interpreted With and Without CAD

#### Overall

Diagnostic accuracy was not improved with CAD on any performance metric assessed. Sensitivity of mammography was 85.3% (95% CI, 83.6%-86.9%) with and 87.3% (95% CI, 84.5%-89.7%) without CAD. Sensitivity of mammography for invasive cancer was 82.1% (95% CI, 80.0%-84.0%) with and 85.0% (95% CI, 81.5%-87.9%) without CAD; for DCIS, sensitivity was 93.2% (95% CI, 91.1%-94.9%) with and 94.3% (95% CI, 89.4%-97.1%) without CAD. Specificity of mammography was 91.6% (95% CI, 91.0%-92.2%) with and 91.4% (95% CI, 90.6%-92.0%) without CAD. There was no difference in overall cancer detection rate (4.1 cancers per 1000 women screened with CAD and without CAD) or in invasive cancer detection rate (2.9 vs 3.0 per 1000 women screened with CAD and without CAD). However, the DCIS detection rate was higher in patients whose mammograms were assessed with CAD compared with those whose mammograms were assessed without CAD (1.2 vs 0.9 per 1000; 95% CI, 1.0-1.9; P < .03) (Table 2).

To allow for the possibility that performance improved after the first year of CAD use by a radiologist, and to account for any possible learning curve, we excluded the first year of mammographic interpretations with CAD for individual radiologists and found no differences for any of our performance measurements (data not shown).
From the ROC analysis, the accuracy of mammographic interpretations with CAD was significantly lower than for those without CAD ($P = .002$). The normalized partial area under the summary ROC curve was 0.84 for interpretations with CAD and 0.88 for interpretations without CAD (Figure 2). In this subset of 135 radiologists who interpreted at least 1 mammogram associated with a cancer diagnosis, sensitivity of mammography was 84.9% (95% CI, 82.9%-86.9%) with and 89.3% (95% CI, 86.9%-91.7%) without CAD. Specificity of mammography was 91.1% (95% CI, 90.4%-91.8%) with and 91.3% (95% CI, 90.5%-92.1%) without CAD.

Differences by Age, Breast Density, Menopausal Status, and Time Since Last Mammogram
We found no differences in diagnostic accuracy of mammographic interpretations with and without CAD in any of the subgroups assessed, including patient age, breast density, and time since last mammogram (Table 3).

Intraradiologist Performance Measures for Mammography With and Without CAD
Among 107 radiologists who interpreted mammograms both with and without CAD, intraradiologist performance was not improved with CAD, and CAD was associated with decreased sensitivity. Sensitivity of mammography was 83.3% (95% CI, 81.0%-85.6%) with and 89.6% (95% CI, 86.0%-91.3%) without CAD. Specificity of mammography was 90.7% (95% CI, 89.8%-91.7%) with and 89.6% (95% CI, 88.6%-91.1%) without CAD. The OR for specificity between mammograms interpreted with CAD and those interpreted without CAD by the same radiologist was 1.02 (95% CI, 0.99-1.05). Sensitivity was significantly decreased for mammograms interpreted with CAD ($P = .002$).

Table 2. Performance Measures of Digital Screening Mammography With and Without CAD

<table>
<thead>
<tr>
<th>Measure</th>
<th>CAD, No.</th>
<th>No CAD, No.</th>
<th>Mean (95% CI)</th>
<th>AOR (95% CI)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancers detected per 1000 exams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2145</td>
<td>558</td>
<td>4.1 (3.8-4.4)</td>
<td>0.99 (0.84-1.15)</td>
<td>.86</td>
</tr>
<tr>
<td>Invasive</td>
<td>1485</td>
<td>408</td>
<td>2.9 (2.7-3.1)</td>
<td>0.92 (0.77-1.08)</td>
<td>.30</td>
</tr>
<tr>
<td>DCIS</td>
<td>660</td>
<td>150</td>
<td>1.2 (1.0-1.3)</td>
<td>1.39 (1.03-1.87)</td>
<td>.03</td>
</tr>
<tr>
<td>Interval cancers per 1000 exams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>375</td>
<td>81</td>
<td>0.8 (0.7-0.8)</td>
<td>1.14 (0.87-1.50)</td>
<td>.33</td>
</tr>
<tr>
<td>Invasive</td>
<td>327</td>
<td>72</td>
<td>0.7 (0.6-0.7)</td>
<td>1.09 (0.82-1.46)</td>
<td>.54</td>
</tr>
<tr>
<td>DCIS</td>
<td>48</td>
<td>9</td>
<td>0.1 (0.1-0.1)</td>
<td>1.59 (0.72-3.51)</td>
<td>.25</td>
</tr>
<tr>
<td>Sensitivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2145</td>
<td>558</td>
<td>85.3 (83.6-86.9)</td>
<td>0.81 (0.60-1.10)</td>
<td>.18</td>
</tr>
<tr>
<td>Invasive</td>
<td>1485</td>
<td>408</td>
<td>82.1 (80.0-84.0)</td>
<td>0.83 (0.59-1.17)</td>
<td>.28</td>
</tr>
<tr>
<td>DCIS</td>
<td>660</td>
<td>150</td>
<td>93.2 (91.1-94.9)</td>
<td>0.88 (0.37-2.07)</td>
<td>.76</td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>444356</td>
<td>118025</td>
<td>91.6 (90.1-92.2)</td>
<td>1.02 (0.94-1.11)</td>
<td>.58</td>
</tr>
<tr>
<td>Invasive</td>
<td>444404</td>
<td>118034</td>
<td>91.5 (90.9-92.1)</td>
<td>1.02 (0.94-1.11)</td>
<td>.58</td>
</tr>
<tr>
<td>DCIS</td>
<td>444683</td>
<td>118097</td>
<td>91.4 (90.7-92.0)</td>
<td>1.04 (0.96-1.13)</td>
<td>.36</td>
</tr>
<tr>
<td>Recall rate per 100 exams</td>
<td>51087</td>
<td>11701</td>
<td>8.7 (8.1-9.4)</td>
<td>0.96 (0.89-1.04)</td>
<td>.35</td>
</tr>
</tbody>
</table>

Abbreviations: AOR, adjusted odds ratio; CAD, computer-aided detection; DCIS, ductal carcinoma in situ; exam, examination.
* Odds ratio for CAD vs No CAD adjusted for site, age group, race/ethnicity, time since prior mammogram, and calendar year of the examination using mixed-effects model with random effect for examination reader and varying with CAD use.
+ The 95% CIs for sensitivity and specificity are given as percentages.
### Table 3. Performance Measures of Digital Screening Mammography With and Without CAD, by Examination-Level Patient Characteristics

<table>
<thead>
<tr>
<th>Measure</th>
<th>CAD Events</th>
<th>CAD Exams</th>
<th>No CAD Events</th>
<th>No CAD Exams</th>
<th>Mean (95% CI)</th>
<th>AOR (95% CI)*</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>By Age, y</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancers detected per 1000 exams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>419</td>
<td>147,486</td>
<td>107</td>
<td>36,503</td>
<td>2.7 (2.4–3.1)</td>
<td>2.6 (2.0–3.4)</td>
<td>1.12</td>
</tr>
<tr>
<td>50-73</td>
<td>1358</td>
<td>295,392</td>
<td>383</td>
<td>82,000</td>
<td>4.3 (4.0–4.7)</td>
<td>4.5 (4.0–5.2)</td>
<td>0.94</td>
</tr>
<tr>
<td>Sensitivityª</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>419</td>
<td>515</td>
<td>107</td>
<td>126</td>
<td>81.6 (77.4–85.2)</td>
<td>89.9 (74.2–96.5)</td>
<td>0.62</td>
</tr>
<tr>
<td>50-73</td>
<td>1358</td>
<td>1581</td>
<td>383</td>
<td>437</td>
<td>85.9 (84.1–87.6)</td>
<td>87.6 (84.2–90.4)</td>
<td>0.87</td>
</tr>
<tr>
<td>Specificityª</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>127,519</td>
<td>146,971</td>
<td>32,228</td>
<td>36,377</td>
<td>88.7 (87.8–89.6)</td>
<td>89.1 (88.1–90.0)</td>
<td>0.99</td>
</tr>
<tr>
<td>50-73</td>
<td>267,865</td>
<td>293,811</td>
<td>75,251</td>
<td>81,563</td>
<td>92.3 (91.7–92.9)</td>
<td>92.2 (91.5–92.8)</td>
<td>1.01</td>
</tr>
<tr>
<td><strong>By BI-RADS Breast Density</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancers detected per 1000 exams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost entirely fat</td>
<td>147</td>
<td>52,875</td>
<td>34</td>
<td>88,33</td>
<td>2.8 (2.4–3.3)</td>
<td>3.8 (2.7–5.4)</td>
<td>0.58</td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>717</td>
<td>175,579</td>
<td>135</td>
<td>33,473</td>
<td>3.8 (3.3–4.2)</td>
<td>4 (3.4–4.8)</td>
<td>0.86</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>783</td>
<td>167,506</td>
<td>123</td>
<td>30,104</td>
<td>4.5 (4.1–5.0)</td>
<td>3.9 (3.1–4.9)</td>
<td>1.11</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>102</td>
<td>31,252</td>
<td>20</td>
<td>53,05</td>
<td>2.7 (1.9–3.7)</td>
<td>1.7 (0.5–5.4)</td>
<td>1.72</td>
</tr>
<tr>
<td>Sensitivityª</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost entirely fat</td>
<td>147</td>
<td>163</td>
<td>34</td>
<td>36</td>
<td>90.2 (84.4–94.0)</td>
<td>100 (91.4–100)</td>
<td>1.26</td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>717</td>
<td>810</td>
<td>135</td>
<td>151</td>
<td>89.0 (86.1–91.4)</td>
<td>89.4 (83.3–93.4)</td>
<td>0.73</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>783</td>
<td>949</td>
<td>123</td>
<td>143</td>
<td>82.5 (79.9–84.8)</td>
<td>86.0 (79.2–90.8)</td>
<td>0.86</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>102</td>
<td>144</td>
<td>20</td>
<td>26</td>
<td>72.1 (60.5–81.4)</td>
<td>77.8 (51.4–92.0)</td>
<td>0.85</td>
</tr>
<tr>
<td>Specificityª</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Almost entirely fat</td>
<td>49,864</td>
<td>52,712</td>
<td>8330</td>
<td>8797</td>
<td>95.8 (95.2–96.3)</td>
<td>94.7 (93.8–95.4)</td>
<td>0.77</td>
</tr>
<tr>
<td>Scattered fibroglandular densities</td>
<td>158,575</td>
<td>174,769</td>
<td>30,230</td>
<td>33,322</td>
<td>92.1 (91.4–92.7)</td>
<td>92.0 (91.2–92.7)</td>
<td>1.01</td>
</tr>
<tr>
<td>Heterogeneously dense</td>
<td>146,180</td>
<td>166,557</td>
<td>26,510</td>
<td>29,961</td>
<td>89.2 (88.3–90.1)</td>
<td>89.2 (88.1–90.2)</td>
<td>1.03</td>
</tr>
<tr>
<td>Extremely dense</td>
<td>27,930</td>
<td>31,108</td>
<td>4724</td>
<td>52,79</td>
<td>91.4 (90.2–92.4)</td>
<td>89.5 (88.0–90.8)</td>
<td>1.30</td>
</tr>
<tr>
<td><strong>By Menopausal Status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancers detected per 1000 exams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenopausal</td>
<td>401</td>
<td>120,559</td>
<td>117</td>
<td>34,688</td>
<td>3.3 (3.0–3.7)</td>
<td>2.9 (2.1–3.9)</td>
<td>1.16</td>
</tr>
<tr>
<td>Postmenopausal, currently taking HT</td>
<td>204</td>
<td>33,764</td>
<td>44</td>
<td>63,38</td>
<td>6 (5.3–6.9)</td>
<td>6.2 (3.9–10.0)</td>
<td>0.88</td>
</tr>
<tr>
<td>Postmenopausal, not currently taking HT</td>
<td>1217</td>
<td>243,105</td>
<td>304</td>
<td>64,335</td>
<td>4.7 (4.3–5.1)</td>
<td>4.6 (4.0–5.3)</td>
<td>0.96</td>
</tr>
<tr>
<td>Sensitivityª</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenopausal</td>
<td>401</td>
<td>484</td>
<td>117</td>
<td>141</td>
<td>83.0 (79.0–86.4)</td>
<td>84.2 (74.5–90.7)</td>
<td>1.06</td>
</tr>
<tr>
<td>Postmenopausal, currently taking HT</td>
<td>204</td>
<td>243</td>
<td>44</td>
<td>51</td>
<td>84.0 (78.7–88.1)</td>
<td>86.3 (73.6–93.4)</td>
<td>0.93</td>
</tr>
<tr>
<td>Postmenopausal, not currently taking HT</td>
<td>1217</td>
<td>1408</td>
<td>304</td>
<td>343</td>
<td>86.4 (84.5–88.1)</td>
<td>90.3 (84.7–94.0)</td>
<td>0.72</td>
</tr>
<tr>
<td>Specificityª</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Premenopausal</td>
<td>102,940</td>
<td>120,075</td>
<td>30,505</td>
<td>34,547</td>
<td>87.9 (86.9–88.8)</td>
<td>88.3 (87.4–89.2)</td>
<td>0.99</td>
</tr>
<tr>
<td>Postmenopausal, currently taking HT</td>
<td>30,129</td>
<td>33,521</td>
<td>5701</td>
<td>62,87</td>
<td>90.7 (89.8–91.6)</td>
<td>91.3 (90.1–92.3)</td>
<td>0.88</td>
</tr>
<tr>
<td>Postmenopausal, not currently taking HT</td>
<td>222,887</td>
<td>241,697</td>
<td>59,263</td>
<td>63,992</td>
<td>93.2 (92.6–93.7)</td>
<td>92.7 (92.1–93.3)</td>
<td>1.05</td>
</tr>
</tbody>
</table>

(continued)
**Discussion**

We found no evidence that CAD applied to digital mammography in US community practice improves screening mammography performance on any performance measure or in any subgroup of women. In fact, mammography sensitivity was decreased in the subset of radiologists who interpreted mammograms with and without CAD. This study builds on prior studies by demonstrating that radiologists’ early learning curve and patient characteristics do not account for the lack of benefit from CAD.

Whether CAD provides added value to women undergoing screening mammography is a topic of strong debate. The lack of consensus may be partly explained by wide variation in CAD use and inherent biases in the methods used to study the impact of CAD on screening mammography. Early studies supporting the efficacy of CAD were laboratory based and measured the ability of CAD programs to mark cancers on selected mammograms. The reported “high sensitivities” of CAD from these studies did not translate to higher cancer detection in clinical practice. In clinical practice, most positive marks by CAD must be reviewed and discounted by a radiologist to avoid unacceptably high rates of false-positive results and unnecessary biopsies, and to practice within acceptable performance parameters recommended by the American College of Radiology.

Concurrent with reports of a prior BCSC cohort study and Surveillance, Epidemiology, and End Results–Medicare data, which primarily evaluated film-screen mammography, we found higher rates of DCIS lesions detected with CAD on digital mammography, but no differences in sensitivity for cancer (whether for DCIS or invasive) and no differences in rates of invasive cancers detected. A meta-analysis in 2008 of 10 studies of CAD applied to screening mammography concluded that CAD significantly increased recall rates with no significant improvement in cancer detection rates compared with readings without CAD. The largest recent reader study of digital mammography obtained during the Digital Mammography Imaging Screening Trial (DMIST) found no impact of CAD on radiologist interpretations of mammograms. In that report, the authors concluded that radiologists overall were not...
fluenced by CAD markings and CAD had no impact, either benefi-
cial or detrimental, on mammography interpretations.

Our study had sufficiently large numbers to compare interpre-
tations of mammograms read by radiologists who practiced at some sites with CAD and at other sites without CAD. We are concerned that, in these comparisons, sensitiv-
ity was lower in CAD-assisted mammograms. Prior reports have
confirmed that not all cancers are marked by CAD and that cancers are overlooked more often if CAD fails to mark a
visible lesion. In a large reader study, Taplin et al7 reported
that visible, noncalcified lesions that went unmarked by CAD
were significantly less likely to be assessed as abnormal by
radiologists. However, our finding of lower sensitivity with
CAD was in a subgroup analysis and should be interpreted
with caution.

Given the observational methods of our study, we could
not compare mammography performance among women who
had their mammograms interpreted both with and without
CAD. It is possible that CAD was used preferentially in women
whose mammograms were more challenging. However, given
the large sample size we were able to control for multiple key
factors known to influence mammography performance, in-
cluding patient age, breast density, menopausal status, and
time since last mammogram. We also were not able to control
for radiologist characteristics, such as experience, and thus
compared performance with and without CAD in the same ra-
diologists, to address across-radiologist variability.

Our study found no beneficial impact of CAD on mam-
mography interpretation. However, CAD may offer advan-
tages beyond interpretation, such as improved workflow or re-
duced search time for faint calcifications. Future research on
potential applications of CAD may emphasize the contribu-
tion of CAD to guide decision-making about treatment of a ra-
diologist-detected lesion, with the worthy goals of reducing
unnecessary biopsy of a mammography lesion with specific
benign features or supporting biopsy of a lesion with specific
malignant features. Finally, CAD might improve mammogra-
phy performance when appropriate training is provided on how
to use it to enhance performance. Nevertheless, given that
the evidence of the current application of CAD in community prac-
tice does not show an improvement in diagnostic accuracy, we
question the policy of continuing to charge for a technology
that provides no established benefits to women.

Gross et al42 reported that the costs of breast cancer screen-
ing exceed $1 billion annually in the Medicare fee-for-service
population. Consistent with our findings, they found wide
variation in CAD use and very limited effectiveness and en-
couraged attention to more appropriate and evidence-based
application of new technologies in breast cancer screening
programs. Despite its overall lack of improvement on interpre-
tive performance, CAD has become routine practice in mam-
mography interpretations in the United States. Seventeen years
have passed since the FDA approved the use of CAD in screening
mammography, and 14 years have passed since Congress
mandated Medicare coverage of CAD. Ten years ago, the In-
stitute of Medicine stated that more information on CAD ap-
plied to mammography was needed before making conclu-
sions about its effect on interpretation.41 The US FDA estimates
that 38.8 million mammograms are performed each year in the
United States. In the BCSC database, 80% of mammograms are
performed for screening and by 2012, 83% of screening mam-
mograms in the BCSC were digital examinations interpreted
with CAD. Current CMS reimbursement for CAD is roughly $7
per examination, and many private insurers pay more than $20
per examination for CAD, translating to over $400 million per
year in current US health care expenditures, with no added
value and in some cases decreased performance.

Conclusions

In the era of Choosing Wisely and clear commitments to sup-
port technology that brings added value to the patient expe-
rience, while aggressively reducing waste and containing
costs,42 CAD is a technology that does not seem to warrant
added compensation beyond coverage of the mamma-
graphic examination. The results of our comprehensive study
lend no support for continued reimbursement for CAD as a
method to increase mammography performance or improve
patient outcomes.
Brandzel, MPH, Group Health Cooperative; Diana Buist, PhD, MPH, Group Health Cooperative; David Buzin, RA, University of California, San Francisco; Elisey Chiapello, BSc, University of California, San Francisco; Rachael Chicoine, BS, University of Vermont; Firas Dabbous, MS, University of Illinois at Chicago; Tammy Dodd, Group Health Cooperative; Therese Dolecek, PhD, MS, University of Illinois at Chicago; Scottie Elsasser, MS, Dartmouth College; Kevin Filcamo, Group Health Cooperative; Pete Frawley, Group Health Cooperative;Hongyan Gao, MS, Group Health Cooperative; Charlottet Gard, PhD, MS, Consultant New Mexico State University; Berta Geller, PhD, University of Vermont; Martha Goodrich, MS, Dartmouth College; Michael Anne Greenwood-Hickman, MPH, University of North Carolina, Chapel Hill; Cindy Groseclose, University of Vermont; Louise Henderson, PhD, MSPH, University of North Carolina, Chapel Hill; Deirdre Hill, PhD, University of New Mexico; Michael Hofmann, MS, University of California, San Francisco; Alla Hubbard, PhD, University of Pennsylvania; Erika Holden, Group Health Cooperative; Tiffany Hoots, University of North Carolina, Chapel Hill; Kathleen Howe, AA, University of Vermont; Laura Ichikawa, MS, Group Health Cooperative; Doug Kane, MS, Group Health Cooperative; Karla Kerlikowske, MD, University of California, San Francisco; Janna Khan, MPH, University of Illinois at Chicago; Gabe Knop, University of North Carolina, Chapel Hill; Suzanne Kolb, MPH University of Washington; Casey Luce, MSPh, Group Health Cooperative; Lin Ma, MA, University of California, San Francisco; Terry Macaro, Advocate Health Care; John Mace, PhD, University of Vermont; Jennifer Maeser, MS, University of Washington; Kathy Malvin, BA, University of California, San Francisco; Katie Marsh, MPH, University of North Carolina, Chapel Hill; Diana Miglioretti, PhD, University of California, Davis; Anne Marie Murphy, PhD, Metropolitan Chicago Breast Cancer Task Force; Ellen O’Meara, PhD, Group Health Cooperative; Tracy Onega, PhD, MA, Dartmouth College; Tiffany Kelley, BA, University of Vermont; Dusty Quick, University of Vermont; Garth Rauscher, PhD, University of Illinois at Chicago; KatekRose Richmire, Group Health Cooperative; Scott Savioli, MA, Dartmouth College; Deborah Seger, Group Health Cooperative; Jennette Sison, MPH, University of California, San Francisco; Brian Sprague, PhD, University of Vermont; Wm. Thomas Summersfelt, PhD, Advocate Health Care; Katherine Tossas-Milligan, MS, University of Illinois at Chicago; Anna Tosteson, ScD, Dartmouth College; Rod Walker, MS, Group Health Cooperative; Julie Weiss, MS, Dartmouth College; Rob Wellman, PhD, Group Health Cooperative; Karen Wernli, PhD, Group Health Cooperative; Hedy Whitson, MS, Group Health Cooperative; Bonnie Yanklas, PhD, University of North Carolina, Chapel Hill; Weimei Zhu, BS, Group Health Cooperative.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the NCIC or the National Institutes of Health.

Additional Contributions: 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.

REFERENCES

Is It Time to Stop Paying for Computer-Aided Mammography?

Joshua J. Fenton, MD, MPH

Computer-aided detection (CAD) is a technology designed to address the problem of screening mammography’s imperfect sensitivity. Now used on over 90% of US mammograms, CAD essentially acts like an automated second reader by marking potentially suspicious spots for radiologists to review before making final recommendations. Early studies suggested that CAD could increase cancer detection rates by 20%.1 But subsequent research suggested little, if any, impact of CAD on cancer detection and raised concerns that CAD may increase recall and biopsy rates.2,3

However, most clinical studies to date have assessed CAD impacts when used with film mammograms. Digital mammography has now largely supplanted film mammography in the United States. When used in the context of digital mammography, does CAD yield net benefits to women? A study by Lehman et al4 in this issue of JAMA Internal Medicine addresses this important question.

In an observational study of 323,973 women undergoing digital screening mammography in diverse US practices, Lehman et al4 found that CAD use was not associated with any improvement in sensitivity, specificity, positive predictive value, cancer detection rates, or other proximal screening outcomes. Indeed, among radiologists who interpreted digital mammograms with and without CAD, sensitivity was worse with CAD, contrary to CAD’s design.

While earlier evaluations suggested that community radiologists often overreacted to CAD output, leading to higher rates of diagnostic investigation,2,5 Lehman et al4 found little, if any, impact of CAD in modern digital mammography practice. It is possible that, with years of CAD use, many radiologists have learned that the yield of CAD is minimal so they now largely ignore CAD output. It is also conceivable that improvements in digital mammography technology have swamped any incremental impacts of CAD on interpretation that may have been previously detectable. This observational study may be confounded by unmeasured radiologist or mammography facility factors, although earlier research adjusting for these factors also found no benefits of CAD.2 Like all subgroup analyses, analyses of outcomes among subsets of radiologists must be interpreted cautiously. Despite these limitations, this study4 is another large-sample, real-world evaluation of CAD’s interpretive outcomes suggesting that CAD yields no clinically significant benefits in typical mammography practice.

The field of implementation science should take interest in interventions like CAD that are widely adopted in advance of strong evidence of effectiveness. What made CAD so alluring to patients, practitioners, or both, and why were payers willing (at least initially) to finance CAD? How is it that CAD is applied on 90% of US mammograms when it yields no clear benefits to patients?

The first essential step for broad CAD adoption was US Food and Drug Administration (FDA) approval in 1998. Because CAD is a device rather than a drug, the evidence bar for FDA approval was comparatively low. Its approval was based on small studies of CAD’s “safety” and “effectiveness.” Effectiveness, for example, was demonstrated by studies in which radiologists read sets of mammograms with enriched breast cancer prevalence, suggesting that CAD could prompt increased cancer detection. In addition, Congressional members, lobbied by industry, pressured the FDA to approve CAD.6

Even so, CAD was still a longshot. Use of CAD required film mammograms to be fed into machines to digitize images for computer analysis, and CAD output had to be viewed on dedicated devices separate from actual mammograms. Mammography was already a loss-leader for many radiology practices, yet CAD added unreimbursed technician and radiologist effort. Without reimbursement, few mammography facilities could justify the capital costs for CAD installation. At the time, establishing reimbursement for new preventive services, such as CAD, required Congressional amendment of the Medicare statute. While Congress had previously added Medicare benefits for preventive services, such as prostate cancer screening, these efforts required strong Congressional sponsorship and auspicious political winds.7

CAD ended up having both. Representing Silicon Valley (home of CAD’s leading manufacturer), California Congress-