Influence of Computer-Aided Detection on Performance of Screening Mammography

Joshua J. Fenton, M.D., M.P.H., **Stephen H. Taplin, M.D., M.P.H.**, **Patricia A. Carney, Ph.D.**, **Linn Abraham, M.S.**, **Edward A. Sickles, M.D.**, **Carl D'Orsi, M.D.**, **Eric A. Berns, Ph.D.**, **Gary Cutter, Ph.D.**, **R. Edward Hendrick, Ph.D.**, **William E. Barlow, Ph.D.**, and **Joann G. Elmore, M.D., M.P.H.**

University of California, Davis, Sacramento (J.J.F.); the National Cancer Institute, Bethesda, MD (S.H.T.); Oregon Health and Science University, Portland (P.A.C.); Group Health Cooperative, Seattle (L.A.); the University of California, San Francisco, San Francisco (E.A.S.); the Emory Clinic, Atlanta (C.D.); Northwestern University, Chicago (E.A.B., R.E.H.); the University of Alabama at Birmingham, Birmingham (G.C.); Cancer Research and Biostatistics, Seattle (W.E.B.); and the University of Washington, Seattle (J.G.E.)

Abstract

Background—Computer-aided detection identifies suspicious findings on mammograms to assist radiologists. Since the Food and Drug Administration approved the technology in 1998, it has been disseminated into practice, but its effect on the accuracy of interpretation is unclear.

Methods—We determined the association between the use of computer-aided detection at mammography facilities and the performance of screening mammography from 1998 through 2002 at 43 facilities in three states. We had complete data for 222,135 women (a total of 429,345 mammograms), including 2351 women who received a diagnosis of breast cancer within 1 year after screening. We calculated the specificity, sensitivity, and positive predictive value of screening mammography with and without computer-aided detection, as well as the rates of biopsy and breast-cancer detection and the overall accuracy, measured as the area under the receiver-operating-characteristic (ROC) curve.

Results—Seven facilities (16%) implemented computer-aided detection during the study period. Diagnostic specificity decreased from 90.2% before implementation to 87.2% after implementation (P<0.001), the positive predictive value decreased from 4.1% to 3.2% (P = 0.01), and the rate of biopsy increased by 19.7% (P<0.001). The increase in sensitivity from 80.4% before implementation of computer-aided detection to 84.0% after implementation was not significant (P = 0.32). The change in the cancer-detection rate (including invasive breast cancers and ductal carcinomas in situ) was not significant (4.15 cases per 1000 screening mammograms before implementation and 4.20 cases after implementation, P = 0.90). Analyses of data from all 43 facilities showed that the use of computer-aided detection was associated with significantly lower overall accuracy than was nonuse (area under the ROC curve, 0.871 vs. 0.919; P = 0.005).

Conclusions—The use of computer-aided detection is associated with reduced accuracy of interpretation of screening mammograms. The increased rate of biopsy with the use of computer-aided detection is not clearly associated with improved detection of invasive breast cancer.

COMPUTER-AIDED DETECTION WAS DEVeloped to assist radiologists in the interpretation of mammograms.[1] Detection programs analyze digitized mammograms and identify suspicious areas for review by the radiologist.[2] Promising studies of the application of computer-aided detection in mammogram test sets led to its approval by the Food and Drug Administration (FDA) in 1998,[3–5] and Medicare and many insurance companies now reimburse for the use of computer-aided detection. Within 3 years after FDA approval, 10% of mammography facilities in the United States had adopted computer-aided detection,[6] and undoubtedly more have done so since. The gradual adoption of digital mammography may foster even broader dissemination of computer-aided detection, since a digital platform facilitates its use. Analogous use of computer-aided detection for computed tomography of the lung and colon are at earlier stages of development and dissemination.[7,8]

Studies of the use of computer-aided detection in actual practice are limited by small numbers of patients or facilities, inability to control for confounding covariates associated with patients or radiologists, and lack of longitudinal follow-up to ascertain cancer outcomes, which precludes the estimation of sensitivity, specificity, and overall accuracy.[9–14] Using data for a large, geographically diverse group of patients, we assessed the effect of computer-aided detection on the performance of screening mammography in community-based settings. We evaluated the sensitivity, specificity, positive predictive value, cancer-detection rate, biopsy rate, and overall accuracy of screening mammography with and without the use of computer-aided detection. By combining data for patients with independent survey data from radiologists and facilities, we could adjust for characteristics of all three groups in our analyses.

Methods

Study Design

We linked data from surveys that were mailed to mammography facilities and affiliated radiologists to data on mammograms and cancer outcomes for women screened between 1998 and 2002 at Breast Cancer Surveillance Consortium facilities. The federally funded consortium facilitates research by linking mammogram registries to population-based cancer registries.[15] Three consortium registries participated in our study: the Group Health Cooperative Breast Cancer Surveillance System, a Washington State health plan with more than 100,000 female enrollees over the age of 40 years; the New Hampshire Mammography Network, which captures data for more than 85% of screening mammograms in New Hampshire; and the Colorado Mammography Program, which captures data for approximately half the screening mammograms in regional Denver. Study procedures were approved by institutional review boards at the University of Washington and the Group Health Cooperative in Seattle, Dartmouth College in New Hampshire, and the Cooper Institute in Colorado.

Study Data

The methods used to survey the facilities and radiologists have been described previously.[6,16–18] In brief, the surveys measured factors that may affect the interpretation of mammograms (e.g., procedures used in reading the images, use of computer-aided detection, years of experience of radiologists in mammography, and number of mammograms interpreted by radiologists in the previous year). Surveys and informed-consent materials were mailed in early 2002.

The consortium developed the methods for collecting and assessing the quality of mammographic and patient data.[15] We included bilateral mammograms designated by radiologists as obtained for "routine screening" of women 40 years of age or older who did not have a history of breast cancer. Mammographic data included assessments of the Breast Imaging Reporting and Data System (BI-RADS), recommendations by radiologists for further evaluation, ages of patients, breast density, time since most recent mammography, and the incidence of biopsy after screening (collected by two of the three registries). BI-RADS assessments were coded as follows: 0, additional imaging evaluation needed; 1, negative; 2, benign abnormality; 3, abnormality that is probably benign; 4, suspicious abnormality; or 5, abnormality highly suggestive of cancer.[19] We ascertained newly diagnosed invasive breast cancers and ductal carcinomas in situ through December 31, 2003, through linkage with regional Surveillance, Epidemiology, and End Results registries or with local or statewide tumor registries.

Performance Measures and Data Classification

We calculated specificity, sensitivity, positive predictive value, and overall accuracy. We defined mammograms with BI-RADS assessment scores of 0, 4, or 5 as positive and mammograms with BI-RADS assessment scores of 1 or 2 as negative. Mammograms with a BI-RADS assessment score of 3 were defined as positive if the radiologist also recommended immediate evaluation and as negative otherwise.[20] Specificity was defined as the percentage of screening mammograms that were negative among patients who did not receive a diagnosis of breast cancer within 1 year after screening. Sensitivity was defined as the percentage of screening mammograms that were positive among patients who received a diagnosis of breast cancer within 1 year after screening. The positive predictive value was defined as the probability of a breast-cancer diagnosis within 1 year after a positive screening mammogram.[19] Overall accuracy was assessed with the use of a receiver-operating-characteristic (ROC) curve, which plots the true positive rate (sensitivity) against the false positive rate (1 – specificity). The area under the ROC curve (AUC) estimates the probability that two hypothetical mammograms, one showing cancer and one not, will be classified correctly as positive and negative, respectively.[21] We also measured the recall rate (the percentage of screening mammograms that were positive) and the rates of biopsy and cancer detection (per 1000 screening mammograms).

After the initial survey, each registry provided additional information in 2005 regarding the use of computer-aided detection at affiliated facilities from 1998 through 2002. For facilities that used computer-aided detection, registry staff ascertained the date of implementation, the brand of computer-aided detection software used, and the estimated percentage of screening mammograms that were interpreted with the use of computer-aided detection after it was implemented. Among facilities that implemented computer-aided detection, all but one reported using computer-aided detection for 100% of screening mammograms after implementation. Thus, we represented the use of computer-aided detection as a binary variable, indicating whether or not it was used at facilities during each study month.

Women were classified on the basis of demographic and clinical covariates known to be associated with the accuracy of mammography,[17,22,23] including age (in 5-year categories), breast density, and months since the most recent mammography. Facilities were classified according to academic affiliation, relative frequencies of screening and diagnostic imaging, interpretation of screening mammograms in batches of 10 or more, availability of interventional services (e.g., core biopsy), number of radiologists who specialized in breast imaging, interpretation of screening mammograms by more than one radiologist (i.e., double-reading), and the frequency of feedback about performance and the method used to review it. Radiologists were classified on the basis of years of experience with mammography and the annual number of mammograms interpreted, which was self-reported

rather than collected from registry data because radiologists may interpret mammograms at nonconsortium facilities. Other characteristics of radiologists have not been associated with performance, so they were not included in the analyses.[17,18]

Statistical Analysis

We performed descriptive analyses to characterize facilities that did and those that did not implement computer-aided detection, as well as the patients and radiologists at these facilities. We used chi-square tests to compare unadjusted performance measures for screening mammography at facilities that adopted computer-aided detection with those that did not. Among the facilities that implemented computer-aided detection, we compared the performance of screening mammography before and after implementation. We examined the overall cancer-detection rates (per 1000 screening mammograms) as well as the rates of detection for invasive cancers and ductal carcinomas in situ.

To adjust for covariates associated with patients, facilities, or radiologists, we used mixed-effects logistic-regression analysis to model specificity, sensitivity, and positive predictive value as functions of the use of computer-aided detection, mammography registry, characteristics of patients (age, breast density, and time since most recent mammography), characteristics of radiologists (years of experience interpreting mammograms and number of mammograms interpreted annually), and four characteristics of facilities that were individually associated with specificity, sensitivity, or positive predictive value in separate analyses (P<0.10). For specificity, we modeled the odds of a true negative screening mammogram. For sensitivity, we modeled the odds of a true positive screening mammogram. For positive predictive value, we modeled the odds of a cancer diagnosis within 1 year after a positive screening mammogram. Models included a random effect at the facility level to account for correlation of mammography outcomes within each facility. We reran each model with an interaction term between the use of computer-aided detection and the study month to assess whether the effect of computer-aided detection on performance changed over time.

We used mixed-effects ordinal-regression analysis to fit an ROC model that included covariates associated with patients, radiologists, facilities, and registries as fixed effects and two random effects for the facility-level "threshold" (the likelihood that a mammogram would be interpreted as positive) and "accuracy" (the ability to discriminate cancer from noncancer).[24] We tested for a significant difference between the AUCs with and those without computer-aided detection, using a likelihood-ratio test. Hypothesis tests were two-sided, with an alpha level of 0.05.

Results

Study Data

Of 51 facilities that contributed mammographic data to registries in the period from 1998 through 2002, 43 (84%) responded to the survey. Within these 43 facilities, there were 159 radiologists who interpreted mammograms, of whom 122 (77%) provided complete responses and written informed consent for linkage to mammography and facility data. Radiologists who did and those who did not respond had similar performance measures for screening mammography.[17] Complete mammographic data were available for 222,135 women (a total of 429,345 screening mammograms), including 2351 women who received a diagnosis of breast cancer within 1 year after screening (Table 1). As in previous studies,[22,23] age, breast density, and time since most recent mammography for patients were associated with specificity, sensitivity, and positive predictive value.

Of the 43 facilities that responded to the survey, 7 (16%) implemented computer-aided detection during the study period. These seven facilities were staffed by 38 radiologists and used computer-aided detection for a total of 124 facility-months (mean, 18 months; range, 2 to 25), during which 31,186 screening mammograms (7% of the total) were interpreted, including 156 mammograms for women who received a diagnosis of breast cancer within 1 year after screening. Six of the seven facilities reported using computer-aided detection for all screening mammograms after implementation; the remaining facility reported using it for 80% of all screening mammograms. All facilities used the same commercial computer-aided detection product (ImageChecker system, R2 Technology).

Women screened at the 36 facilities that did not implement computer-aided detection during study period were older, had denser breasts, and were less likely to have undergone mammography within the previous 9 to 20 months than those screened at the 7 facilities that implemented computer-aided detection (Table 2), implying a higher overall risk of breast cancer among the women screened at the nonimplementing facilities.[25] On average, radiologists at facilities that did not implement computer-aided detection had more years of experience with mammography than did radiologists at facilities that implemented computer-aided detection. Characteristics of patients and radiologists at the facilities that adopted computer-aided detection were similar before and after its implementation. Facilities that did and those that did not implement computer-aided detection were similar across a range of characteristics, including the presence or absence of radiologists who specialized in breast imaging (Table 3).

Use of Computer-Aided Detection and Unadjusted Performance

Differences in characteristics of patients and radiologists (Table 2) would predict lower specificity, higher recall rates, and higher sensitivity at facilities that never implemented computer-aided detection than at facilities that did.[17,22,26] Indeed, at the 36 facilities that never implemented computer-aided detection, specificity was significantly lower (P<0.001) and recall rates were significantly higher (P<0.001) than at the 7 facilities that adopted computer-aided detection but had not yet implemented it (Table 4). After these 7 facilities implemented computer-aided detection, the opposite was true: the specificity was significantly lower and the recall rate was significantly higher than at the 36 facilities that never implemented computer-aided detection (P<0.001 for both comparisons), even though characteristics of patients and radiologists remained stable (Table 2).

As expected from characteristics of patients and radiologists, sensitivity was slightly higher at facilities that never implemented computer-aided detection than at facilities that adopted computer-aided detection but had not yet implemented it (Table 4). Although sensitivity increased from 80.4% before implementation to 84.0% after implementation, the change was not significant (P = 0.32). The positive predictive value was similar at facilities that never implemented computer-aided detection and facilities that adopted computer-aided detection but had not yet implemented it. After implementation at these facilities, the positive predictive value decreased significantly (P = 0.01).

Before the adoption of computer-aided detection at the 7 facilities, the biopsy rate was similar to that at the 36 facilities that never implemented computer-aided detection. After the seven facilities implemented computer-aided detection, the biopsy rate increased by 20% (from 14.7 biopsies per 1000 screening mammograms before implementation to 17.6 biopsies after implementation, P<0.001). As anticipated from risk factors of patients (i.e., older age, denser breasts, and less recent mammography) at the 36 facilities that never implemented computer-aided detection,[22,25] the cancer-detection rate was significantly higher at these facilities than at the 7 facilities that adopted computer-aided detection but had not yet implemented it (P = 0.03). Before and after the implementation of computer-

aided detection at the seven facilities, the cancer-detection rate was similar (4.15 and 4.20 cases per 1000 screening mammograms, respectively; P = 0.90), but the proportions of detected invasive cancers and ductal carcinomas in situ changed: the rate of detection of invasive breast cancer decreased by 12% (from 2.98 cases per 1000 screening mammograms before implementation to 2.63 cases after implementation, P = 0.32), whereas the rate of detection of ductal carcinomas in situ increased by 34% (from 1.17 to 1.57 cases per 1000 screening mammograms, P = 0.09). The percentage of detected cases of cancer that were ductal carcinomas in situ increased significantly after implementation of computer-aided detection as compared with before implementation (37.4% vs. 28.1%, P = 0.049).

Before the adoption of computer-aided detection at the 7 facilities, the biopsy rate was similar to that at the 36 facilities that never implemented computer-aided detection. After the seven facilities implemented computer-aided detection, the biopsy rate increased by 20% (from 14.7 biopsies per 1000 screening mammograms before implementation to 17.6 biopsies after implementation, P<0.001). As anticipated from risk factors of patients (i.e., older age, denser breasts, and less recent mammography) at the 36 facilities that never implemented computer-aided detection,[22,25] the cancer-detection rate was significantly higher at these facilities than at the 7 facilities that adopted computer-aided detection but had not yet implemented it (P = 0.03). Before and after the implementation of computer-aided detection at the seven facilities, the cancer-detection rate was similar (4.15 and 4.20 cases per 1000 screening mammograms, respectively; P = 0.90), but the proportions of detected invasive cancers and ductal carcinomas in situ changed: the rate of detection of invasive breast cancer decreased by 12% (from 2.98 cases per 1000 screening mammograms before implementation to 2.63 cases after implementation, P = 0.32), whereas the rate of detection of ductal carcinomas in situ increased by 34% (from 1.17 to 1.57 cases per 1000 screening mammograms, P = 0.09). The percentage of detected cases of cancer that were ductal carcinomas in situ increased significantly after implementation of computer-aided detection as compared with before implementation (37.4% vs. 28.1%, P = 0.049).

Adjusted Performance

Of the 429,345 mammograms in this analysis, 332,869 (78%) were interpreted by participating radiologists and thus could be included in analyses that adjusted simultaneously for characteristics of patients, facilities, and radiologists. After adjustment, the use of computer-aided detection as compared with nonuse remained associated with significantly lower specificity and positive predictive value, as well as nonsignificantly greater sensitivity (Table 5). Because mammograms are usually correctly interpreted, odds ratios do not accurately estimate percent changes in specificity, sensitivity, or positive predictive value. Thus, among women with breast cancer, the 46% greater adjusted odds of a positive mammogram with the use of computer-aided detection as compared with nonuse (odds ratio, 1.46) is consistent with the absolute increase in sensitivity of 3.6% associated with the implementation of computer-aided detection in unadjusted analyses (from 80.4% before implementation to 84.0% after implementation).

In secondary analyses, associations were generally similar to those in our principal analysis (Table 5). However, the association between the use of computer-aided detection and sensitivit was weaker when the analysis was restricted to women with invasive breast cancer and was stronger when restricted to women with ductal carcinoma in situ. The observed associations were similar to those in the principal analysis after the exclusion of mammograms that were interpreted within 3 months after the implementation of computer-aided detection, and interaction terms for time and use of computer-aided detection were nonsignificant, suggesting that the observed changes in performance with computer-aided detection persisted over time.

As shown in Figure 1, the modeled AUC was 0.919 without the use of computer-aided detection but was 0.871 with its use (P = 0.005). Because accuracy increases as the AUC approaches 1.0, the use of computer-aided detection was associated with significantly lower overall accuracy than was nonuse.

Discussion

The use of computer-aided detection in clinical practice has increased since the FDA approved the technology and Medicare began reimbursing for its use. In our observational study of large numbers of community-based mammography facilities and patients, the use of computer-aided detection was associated with increases in potential harms of screening mammography, including higher recall and biopsy rates, and was of uncertain clinical benefit.

As others have reported,[9,11–14] we found that the use of computer-aided detection was associated with higher recall rates than nonuse, implying that rates of false positive results were also higher with use, since most recalls do not result in a diagnosis of cancer. Increased recall rates and rates of false positive results may be logical consequences of the design of computer-aided detection software. With the goal of alerting radiologists to overlooked suspicious areas, computer-aided detection programs insert up to four marks on the average screening mammogram.[13,27,28] Thus, for every true positive mark resulting from computer-aided detection that is associated with an under-lying cancer, radiologists encounter nearly 2000 false positive marks.[29]

Increased recall rates could be a necessary cost of improved cancer detection. The use of computer-aided detection was associated with a nonsignificant trend toward increased sensitivity but with no substantive change in the overall detection of cancer. Use of the technology was, however, more strongly associated with the detection of ductal carcinoma in situ than with the detection of invasive breast cancer, a finding that may stem from the propensity of computer-aided detection software to mark calcifications.[5,30–32] To the extent that ductal carcinoma in situ is a precursor to invasive cancer,[33] the greater percentage of cancers found that were ductal carcinomas in situ after the implementation of computer-aided detection than before implementation may be viewed optimistically as a shift toward detecting breast cancer at an earlier stage with the use of computer-aided detection. On the other hand, the natural history of ductal carcinoma in situ is certainly more indolent than that of invasive cancer,[34] and the effect of computer-aided detection on mortality from breast cancer may be limited if it chiefly promotes the identification of ductal carcinoma in situ rather than invasive cancer.[35]

No single measure is sufficient to judge the effect of computer-aided detection on interpretive performance.[36] Rather, the benefits of true positive results must be weighed against the consequences of false positive results, including associated economic costs. Our results suggest that approximately 157 women would be recalled (and 15 women would undergo biopsy) owing to the use of computer-aided detection in order to detect one additional case of cancer, possibly a ductal carcinoma in situ (see the Supplementary Appendix, available with the full text of this article at www.nejm.org). After accounting for the additional fees for the use of computer-aided detection[37] and the costs of diagnostic evaluations after recalls resulting from the use of computer-aided detection,[38] we calculated that system-wide use of computer-aided detection in the United States could increase the annual national costs of screening mammography by approximately 18% ($550 million) (see the Supplementary Appendix).

Facilities that adopted computer-aided detection had performance measures before its implementation that differed from those at facilities that never adopted computer-aided detection; these differences were consistent with differences in the characteristics of patients and radiologists at the two groups of facilities. The use of computer-aided detection may have caused a regression toward mean levels of performance among radiologists whose interpretations of mammograms tended to differ from those of most radiologists, but its implementation was associated with changes in specificity and recall rates that overshot levels at facilities that never implemented computer-aided detection. Moreover, the use of computer-aided detection remained significantly associated with decreased specificity, decreased positive predictive value, and decreased overall accuracy in analyses that adjusted for differences in characteristics of patients, radiologists, and facilities. Nevertheless, the association between the use of computer-aided detection and the observed changes in performance could be explained by factors we did not measure.

Although six of the seven facilities that adopted computer-aided detection reported using it for 100% of mammograms after implementation, we did not measure the use of computer-aided detection at the level of the individual mammogram. In this respect, our estimates of the effects of computer-aided detection on performance may be conservative. All facilities that implemented computer-aided detection used the same commercial product. The manufacturer has updated its detection software since 2002, but we are unaware of any community-based studies that have found improved detection of breast cancer with recent versions of the software.

Even in our large study, only 156 cases of cancer developed during the study among women screened at facilities using computer-aided detection, resulting in wide confidence intervals around estimates of sensitivity and cancer-detection rates after implementation. Because of the rarity of breast cancer in community samples, very large samples would be needed to study the effect of the use of computer-aided detection on sensitivity with high statistical power (approximately 750,000 mammograms interpreted in total, half with the use of computer-aided detection and half without).

In conclusion, we found that, among large numbers of diverse facilities and radiologists, the use of computer software designed to improve the interpretation of mammograms was associated with significantly higher false positive rates, recall rates, and biopsy rates and with significantly lower overall accuracy in screening mammography than was nonuse. The nonsignificant trend toward greater sensitivity with the use of computer-aided detection as compared with nonuse may be largely explained by increased detection of ductal carcinoma in situ. As an FDA-approved technology whose use can be reimbursed by Medicare, computer-aided detection has been incorporated quickly into mammography practices, despite tentative evidence of clinical benefits. Now that computer-aided detection is used in the screening of millions of healthy women, larger studies are needed to judge more precisely whether benefits of routine use of computer-aided detection outweigh its harms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Chan HP, Doi K, Galhotra S, Vyborny CJ, MacMahon H, Jokich PM. Image feature analysis and computer-aided diagnosis in digital radiography. I. Automated detection of microcalcifications in mammography. Med Phys. 1987; 14:538–48. [PubMed: 3626993]

2. Doi K, Giger ML, Nishikawa RM, Schmidt RA. Computer-aided diagnosis of breast cancer on mammograms. Breast Cancer. 1997; 4:228–33. [PubMed: 11091604]

3. Brem RF, Baum J, Lechner M, et al. Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial. AJR Am J Roentgenol. 2003; 181:687–93. [PubMed: 12933460]

4. Warren Burhenne LJ, Wood SA, D'Orsi CJ, et al. Potential contribution of computer-aided detection to the sensitivity of screening mammography. Radiology. 2000; 215:554–62. Erratum, Radiology 2000:216:306. [PubMed: 10796939]

5. 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; 219:192–202. [PubMed: 11274556]

6. Hendrick RE, Cutter GR, Berns EA, et al. Community-based mammography practice: services, charges, and interpretation methods. AJR Am J Roentgenol. 2005; 184:433–8. [PubMed: 15671359]

7. Das M, Muhlenbruch G, Mahnken AH, et al. Small pulmonary nodules: effect of two computer-aided detection systems on radiologist performance. Radiology. 2006; 241:564–71. [PubMed: 17057074]

8. Summers RM, Jerebko AK, Franaszek M, Malley JD, Johnson CD. Colonic polyps: complementary role of computer-aided detection in CT colonography. Radiology. 2002; 225:391–9. [PubMed: 12409571]

9. Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology. 2001; 220:781–6. [PubMed: 11526282]

10. Gur D, Sumkin JH, Rockette HE, et al. Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. J Natl Cancer Inst. 2004; 96:185–90. [PubMed: 14759985]

11. Cupples TE, Cunningham JE, Reynolds JC. Impact of computer-aided detection in a regional screening mammography program. AJR Am J Roentgenol. 2005; 185:944–50. [PubMed: 16177413]

12. Birdwell RL, Bandodkar P, Ikeda DM. Computer-aided detection with screening mammography in a university hospital setting. Radiology. 2005; 236:451–7. [PubMed: 16040901]

13. Khoo LA, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom National Breast Screening Programme: prospective study. Radiology. 2005; 237:444–9. [PubMed: 16244252]

14. Morton MJ, Whaley DH, Brandt KR, Amrami KK. Screening mammograms: interpretation with computer-aided detection — prospective evaluation. Radiology. 2006; 239:375–83. [PubMed: 16569779]

15. Ballard-Barbash R, Taplin SH, Yankaskas BC, et al. Breast Cancer Surveillance Consortium: a national mammography screening and outcomes database. AJR Am J Roentgenol. 1997; 169:1001–8. [PubMed: 9308451]

16. D'Orsi C, Tu SP, Nakano C, et al. Current realities of delivering mammography services in the community: do challenges with staffing and scheduling exist? Radiology. 2005; 235:391–5. [PubMed: 15798153]

17. Barlow WE, Chi C, Carney PA, et al. Accuracy of screening mammography interpretation by characteristics of radiologists. J Natl Cancer Inst. 2004; 96:1840–50. [PubMed: 15601640]

18. Elmore JG, Taplin SH, Barlow WE, et al. Does litigation influence medical practice? The influence of community radiologists' medical malpractice perceptions and experience on screening mammography. Radiology. 2005; 236:37–46. [PubMed: 15987961]

19. D'Orsi, C.; Bassett, L.; Feig, S., et al. Breast imaging reporting and data system (BI-RADS). 3rd ed.. American College of Radiology; Reston, VA: 1998.

20. 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; 222:529–35. [PubMed: 11818624]

21. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982; 143:29–36. [PubMed: 7063747]

22. Yankaskas BC, Taplin SH, Ichikawa L, et al. Association between mammography timing and measures of screening performance in the United States. Radiology. 2005; 234:363–73. [PubMed: 15670994]

23. Miglioretti DL, Rutter CM, Geller BM, et al. Effect of breast augmentation on the accuracy of mammography and cancer characteristics. JAMA. 2004; 291:442–50. [PubMed: 14747501]

24. Tosteson AN, Begg CB. A general regression methodology for ROC curve estimation. Med Decis Making. 1988; 8:204–15. [PubMed: 3294553]

25. Barlow WE, White E, Ballard-Barbash R, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst. 2006; 98:1204–14. [PubMed: 16954473]

26. Smith-Bindman R, Chu P, Miglioretti DL, et al. Physician predictors of mammographic accuracy. J Natl Cancer Inst. 2005; 97:358–67. [PubMed: 15741572]

27. Castellino RA. Computer-aided detection: an overview. Appl Radiol Suppl. 2001; 30:5–8.

28. Food and Drug Administration. [Accessed March 9, 2007] Summary of safety and effectiveness data: MammoReader. http://www.fda.gov/cdrh/pdf/P010038b.pdf

29. Brenner RJ, Ulissey MJ, Wilt RM. Computer-aided detection as evidence in the courtroom: potential implications of an appellate court's ruling. AJR Am J Roentgenol. 2006; 186:48–51. [PubMed: 16357376]

30. Taplin SH, Rutter CM, Lehman C. Testing the effect of computer-assisted detection on interpretive performance in screening mammography. AJR Am J Roentgenol. 2006; 187:1475–82. [PubMed: 17114540]

31. Malich A, Sauner D, Marx C, et al. Influence of breast lesion size and histo-logic findings on tumor detection rate of a computer-aided detection system. Radiology. 2003; 228:851–6. [PubMed: 12869683]

32. Brem RF, Schoonjans JM. Radiologist detection of microcalcifications with and without computer-aided detection: a comparative study. Clin Radiol. 2001; 56:150–4. [PubMed: 11222075]

33. Evans AJ, Pinder SE, Ellis IO, Wilson AR. Screen detected ductal carcinoma in situ (DCIS): overdiagnosis or an obligate precursor of invasive disease? J Med Screen. 2001; 8:149–51. [PubMed: 11678555]

34. Sanders ME, Schuyler PA, Dupont WD, Page DL. The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. Cancer. 2005; 103:2481–4. [PubMed: 15884091]

35. Ernster VL, Ballard-Barbash R, Barlow WE, et al. Detection of ductal carcinoma in situ in women undergoing screening mammography. J Natl Cancer Inst. 2002; 94:1546–54. [PubMed: 12381707]

36. D'Orsi CJ. Computer-aided detection: there is no free lunch. Radiology. 2001; 221:585–6. [PubMed: 11719647]

37. The Centers for Medicare and Medicaid Services (CMS) increase reimbursement for computer aided detection (CAD) technology. R2 Technology; Los Altos, CA: 2004. http://www.r2tech.com/main/company/news_one_up.php?prID=81&offset=0&year_search=2001

38. Kerlikowske K, Salzmann P, Phillips KA, Cauley JA, Cummings SR. Continuing screening mammography in women aged 70 to 79 years: impact on life expectancy and cost-effectiveness. JAMA. 1999; 282:2156–63. [PubMed: 10591338]
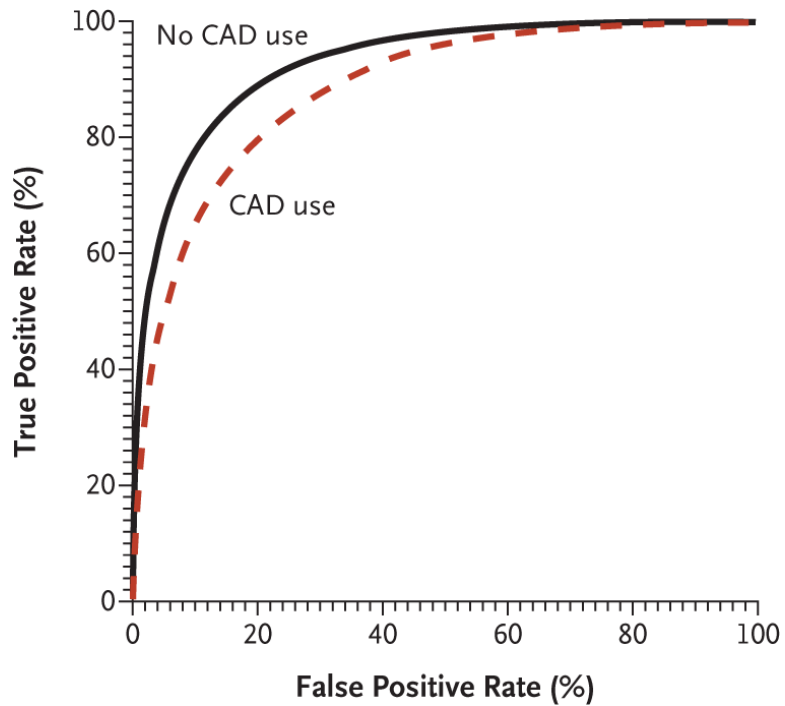
Figure 1. Overall Accuracy of Screening Mammography, According to the Use of Computer-Aided Detection (CAD)
Overall accuracy was assessed with the use of receiver-operating-characteristic (ROC) curves for 332,869 mammograms interpreted (308,099 without the use of CAD and 24,770 with the use of CAD) by participating radiologists. These curves plot the true positive rate of screening mammography (sensitivity) against the false positive rate (1 – specificity). The ROC curves and estimates of the areas under the curve (AUCs) were adjusted for patient, radiologist, and facility characteristics, as well as for mammography registry. The AUC was 0.919 for nonuse of CAD and 0.871 for use of CAD (P = 0.005).

Table 1

Unadjusted Performance of Screening Mammography from 1998 to 2002, According to Characteristics of Patients.

Variable	Mammograms for 222,135 Patients	Mammograms for 220,886 Patients without Breast Cancer*	Mammograms for 2351 Patients with Breast Cancer	Specificity†	Sensitivity‡	ppv§
		no. of mammograms (%)		percent		
All mammograms	429,345	426,994 (100)	2351 (100)	89.2	83.1	4.1
Age at mammography						
40–44 yr	62,637	62,486 (14.6)	151 (6.4)	87.2	72.2	1.3
45–49 yr	71,709	71,455 (16.7)	254 (10.8)	87.6	80.3	2.2
50–54 yr	79,221	78,870 (18.5)	351 (14.9)	88.6	82.3	3.1
55–59 yr	57,451	57,110 (13.4)	341 (14.5)	89.4	83.9	4.5
60–64 yr	41,737	41,464 (9.7)	273 (11.6)	90.0	84.6	5.3
65–69 yr	37,156	36,879 (8.6)	277 (11.8)	90.6	86.6	6.5
70–74 yr	33,928	33,643 (7.9)	285 (12.1)	91.6	84.6	7.9
≥75 yr	45,506	45,087 (10.6)	419 (17.8)	92.0	84.5	8.9
Breast density						
Almost entirely fat	41,203	41,084 (9.6)	119 (5.1)	95.2	91.6	5.2
Scattered fibroglandular tissue	184,714	183,786 (43.0)	928 (39.5)	90.6	87.9	4.5
Heterogeneously dense	166,303	165,212 (38.7)	1091 (46.4)	86.6	80.8	3.8
Extremely dense	37,125	36,912 (8.6)	213 (9.1)	87.7	69.0	3.1
Time since most recent mammogram						
No previous mammogram	16,862	16,763 (3.9)	99 (4.2)	81.3	86.9	2.7
9–15 mo	193,006	192,062 (45.0)	944 (40.2)	90.5	77.3	3.9
16–20 mo	45,718	45,486 (10.7)	232 (9.9)	89.6	81.0	3.8
21–27 mo	105,507	104,852 (24.6)	655 (27.9)	89.5	88.2	5.0
≥28 mo	68,252	67,831 (15.9)	421 (17.9)	87.0	88.4	4.0

*
Some women without breast cancer received a diagnosis of the disease later in the study period.

†
Specificity was defined as the percentage of screening mammograms that were negative (had a BI-RADS assessment score of 1 or 2, or of 3 with a recommendation of normal or short-interval follow-up) among patients who did not receive a diagnosis of breast cancer within 1 year after screening.

[‡] Sensitivity was defined as the percentage of screening mammograms that were positive (had a BI-RADS assessment score of 0, 4, or 5, or of 3 with a recommendation for immediate workup) among patients who received a diagnosis of breast cancer within 1 year after screening.

[§] The positive predictive value (PPV) was defined as the probability of a breast-cancer diagnosis within 1 year after a positive screening mammogram.

Table 2

Characteristics of Patients and Radiologists at Mammography Facilities, According to Use or Nonuse of Computer-Aided Detection (CAD). [*]

Characteristic	CAD Never Implemented	CAD Implemented	
		Before Implementation	After Implementation
Patients [†]			
Age (yr)	57.8±12.1	55.1±10.9	55.4±11.1
High breast density (%)[‡]	49.3	41.5	44.2
Mammography within previous 9–20 mo (%)	49.1	71.8	76.7
Radiologists [§]			
Experience performing mammography (%)			
<10yr	16.6	29.9	26.8
10–19 yr	57.0	54.2	53.2
≥20 yr	26.4	15.9	20.0
No. of mammograms interpreted in previous year (%)			
≤1000	11.9	8.0	7.9
1001–2000	30.1	45.1	40.9
>2000	58.0	46.9	51.2

[*] Plus–minus values are means ±SD. All characteristics differed significantly (P<0.001) across categories of CAD implementation.

[†] Characteristics of patients are provided for 429,345 mammograms (398,159 performed without CAD and 31,186 performed with CAD).

[‡] High density was defined as either heterogeneously or extremely dense.

[§] These characteristics are provided for 332,869 mammograms interpreted (308,099 without CAD and 24,770 with CAD) by radiologists who responded to a mailed survey and gave informed consent for linkage to mammographic data.

Table 3

Characteristics of 43 Mammography Facilities, According to Use or Nonuse of Computer-Aided Detection (CAD).

Facility Characteristic	CAD Never Implemented (N = 36)	CAD Implemented (N = 7)
	no. (%)	
General		
Associated with academic center	7 (19)	2 (29)
One or more staff radiologists specializing in breast imaging	13 (36)	3 (43)
For-profit	21 (58)	3 (43)
Feedback provided to a radiologist on performance		
Once a year	15 (42)	2 (29)
Twice or more a year	14 (39)	5 (71)
Uncertain	7 (19)	0
Interpretation of mammograms		
Usually interpreted in batches of 10 or more	27 (82)[*]	5 (83)[†]
Usually interpreted at another facility	15 (42)	2 (29)
Sometimes interpreted by >1 radiologist	20 (56)	2 (33)[†]
Diagnostic and screening services		
Diagnostic mammograms offered	26 (72)	6 (86)
Screening mammograms as a proportion of all mammograms		
1–74%	16 (44)	2 (29)
75–79%	7 (19)	4 (57)
80–100%	13 (36)	1 (14)
Interventional services offered[‡]	19 (54)[†]	6 (86)

[*] Data were missing for three facilities.

[†] Data were missing for one facility.

[‡] Interventional services included core biopsy, fine-needle aspiration, and other invasive procedures.

Table 4

Unadjusted Performance of Screening Mammography, Breast Biopsy, and Cancer Detection, According to Use or Nonuse of Computer-Aided Detection (CAD).

Variable	CAD Never Used	Before CAD Implementation	P Value[*]	After CAD Implementation	P Value[†]
Screening mammograms					
Total no. performed	313,259	84,900		31,186	
No. performed in patients without breast cancer	311,502	84,462		31,030	
No. performed in patients with breast cancer	1757	438		156	
Invasive cancer	1420	326		103	
Ductal carcinoma in situ	337	112		53	
Performance measure — % (95% CI)					
Specificity[‡]	89.2 (89.1–89.3)	90.2 (90.0–90.4)	<0.001	87.2 (86.8–87.5)	<0.001
Recall rate[§]	11.2 (11.1–11.4)	10.1 (9.9–10.3)	<0.001	13.2 (12.8–13.6)	<0.001
Sensitivity[¶]	83.7 (81.9–85.4)	80.4 (76.3–84.0)	0.09	84.0 (77.3–89.4)	0.32
PPV[‖]	4.2 (4.0–4.4)	4.1 (3.7–4.5)	0.71	3.2 (2.7–3.8)	0.01
Breast biopsy					
Total no. performed	3166	1248		550	
No. performed per 1000 screening mammograms (95% CI)	14.3 (13.8–14.8)[**]	14.7 (13.9–15.5)	0.41	17.6 (16.2–19.2)	<0.001
No. of cancers detected per 1000 screening mammograms (95% CI)					
All cancers	4.70 (4.46–4.94)	4.15 (3.72–4.60)	0.03	4.20 (3.51–4.98)	0.90
Invasive cancers	3.71 (3.50–3.93)	2.98 (2.62–3.37)	0.002	2.63 (2.09–3.26)	0.32
Ductal carcinomas in situ	0.98 (0.88–1.10)	1.17 (0.95–1.42)	0.14	1.57 (1.16–2.08)	0.09
Percentage of all cancers that were ductal carcinoma in situ	20.9 (18.9–23.1)	28.1 (23.5–33.1)	0.004	37.4 (29.1–46.3)	0.049

[*] P values were calculated with the use of the chi-square test for the comparison between facilities that never implemented CAD and those that adopted CAD but had not yet implemented it.

[†] P values were calculated with the use of the chi-square test for the comparison, among facilities that adopted CAD, between values before implementation and those after implementation.

[‡] Specificity was defined as the percentage of screening mammograms that were negative (had a BI-RADS assessment score of 1 or 2, or of 3 with a recommendation of normal or short-interval follow-up) among patients who did not receive a diagnosis of breast cancer within 1 year after screening.

[§] The recall rate was defined as the percentage of screening mammograms that were positive.

¶Sensitivity was defined as the percentage of screening mammograms that were positive (had a BI-RADS assessment score of 0, 4, or 5, or of 3 with a recommendation for immediate work-up) among patients who received a diagnosis of breast cancer within 1 year after screening.

‖The positive predictive value (PPV) was defined as the probability of a breast-cancer diagnosis within 1 year after a positive screening mammogram.

**The number of biopsies per 1000 screening mammograms at facilities that never implemented CAD was calculated on the basis of data from the two registries that maintain biopsy data (221,399 mammograms).

Table 5

Adjusted Relative Performance of Screening Mammography, According to Use of Computer-Aided Detection (CAD) as Compared with Nonuse.

Analysis*	No. of Mammograms	No. of Patients with Breast Cancer	Specificity		Sensitivity		PPV	
			Odds Ratio (95% CI)†	P Value	Odds Ratio (95%CI)‡	P Value	Odds Ratio (95% CI)§	P Value
Principal analysis	332,869	1822	0.68 (0.64–0.72)	<0.001	1.46 (0.81–2.63)	0.20	0.75 (0.58–0.97)	0.03
Secondary analysis								
Including mammograms interpreted by nonparticipating radiologists, without adjustment for radiologist characteristics	429,345	2351	0.72 (0.68–0.75)	<0.001	1.27 (0.77–2.12)	0.34	0.75 (0.59–0.94)	0.02
Excluding mammograms interpreted during first 3 months of CAD implementation at each facility	328,094	1800	0.62 (0.58–0.66)	<0.001	1.61 (0.84–3.09)	0.15	0.72 (0.54–0.95)	0.02
Detection of breast-cancer subtypes								
Invasive cancer	332,869	1427	—	—	1.18 (0.60–2.29)	0.62	—	—
Ductal carcinoma in situ	332,869	395	—	—	2.88 (0.54–15.2)	0.21	—	—

*
Except where noted, models were adjusted for characteristics of patients (age, breast density, and time since previous mammography) and of radiologists (number of years interpreting mammograms and number of mammograms interpreted in the past year), mammography registry, and four facility-level covariates that were associated (P<0.10) with one or more performance measures in preliminary analyses: the percentage of screening mammograms that are interpreted in batches of 10 or more, availability of diagnostic mammography, frequency of feedback about performance, and the method of review of the feedback.

†
The odds ratio gives the relative odds of a negative mammogram (one with a BI-RADS assessment score of 1 or 2, or of 3 with a recommendation of normal or short-interval follow-up) among patients who did not receive a diagnosis of breast cancer within 1 year after screening.

‡
The odds ratio gives the relative odds of a positive mammogram (one with a BI-RADS assessment score of 0, 4, or 5, or of 3 with a recommendation for immediate workup) among patients who received a diagnosis of breast cancer within 1 year after screening.

§
The odds ratio gives the relative odds of a breast-cancer diagnosis within 1 year after a positive screening mammogram.