Linda J. Warren Burhenne, MD
Susan A. Wood, PhD
Carl J. D'Orsi, MD
Stephen A. Feig, MD
Daniel B. Kopans, MD
Kathryn F. O'Shaughnessy, PhD
Edward A. Sickles, MD
Laszlo Tabar, MD
Carl J. Vyborny, MD, PhD
Ronald A. Castellino, MD

Index terms:
Breast neoplasms, 00.32
Breast neoplasms, diagnosis, 00.32
Cancer screening, 00.11, 00.1299
Computers, diagnostic aid

Radiology 2000; 215:554–562

Abbreviations:
BI-RADS = Breast Imaging Reporting and Data System
CAD = computer-aided detection
MQSA = Mammography Quality Standards Act

[1] From the Dept of Radiology, Univ of British Columbia, Vancouver, Canada (L.J.W.B.); R2 Technology, 325 Distel Circle, Los Altos, CA 94022 (S.A.W., K.F.O., R.A.C.); Dept of Radiology, Univ of Massachusetts Medical Center, North Worcester (C.J.D.); Breast Imaging Center, Thomas Jefferson Univ Hosp, Philadelphia, Pa (S.A.F.); Dept of Radiology, Massachusetts General Hosp, Boston (D.B.K.); UCSF Medical Center, San Francisco, Calif (E.A.S.); Dept of Mammography, Central Hosp, Falun, Sweden (L.T.); and Dept of Radiology, Univ of Chicago, Ill (C.J.V.). From the 1998 RSNA scientific assembly. Received Jul 19, 1999; revision requested Aug 26; revision received Sep 21; accepted Oct 21. **Address correspondence to** R.A.C. (e-mail: *rcastell@r2tech.com*).

L.J.W.B., member of Medical Advisory Board, chair of Mammography Advisory Panel, and stockholder, R2 Technology; S.A.W. and K.F.O., employees, R2 Technology; C.J.V., stockholder, R2 Technology; R.A.C., medical director and stockholder, R2 Technology; C.J.D., S.A.F., D.B.K., E.A.S., and L.T., members of Mammography Advisory Panel, R2 Technology.

Author contributions:
Guarantors of integrity of entire study, all authors; study concepts and design, L.J.W.B., C.J.D., S.A.F., D.B.K., E.A.S., L.T., C.J.V.; definition of intellectual content, L.J.W.B., C.J.D., S.A.F., D.B.K., E.A.S., L.T., C.J.V., R.A.C.; literature research, L.J.W.B., S.A.F., E.A.S., C.J.V.; clinical studies, K.F.O.; data acquisition, K.F.O.; data analysis, S.A.F., E.A.S., C.J.V., S.A.W., R.A.C., K.F.O.; manuscript preparation and review, all authors; manuscript editing, S.A.F., E.A.S., C.J.V., S.A.W., R.A.C.

Potential Contribution of Computer-aided Detection to the Sensitivity of Screening Mammography[1]

PURPOSE: To determine the false-negative rate in screening mammography, the capability of computer-aided detection (CAD) to identify these missed lesions, and whether or not CAD increases the radiologists' recall rate.

MATERIALS AND METHODS: All available screening mammograms that led to the detection of biopsy-proved cancer (n = 1,083) and the most recent corresponding prior mammograms (n = 427) were collected from 13 facilities. Panels of radiologists evaluated the retrospectively visible prior mammograms by means of blinded review. All mammograms were analyzed by a CAD system that marks features associated with cancer. The recall rates of 14 radiologists were prospectively measured before and after installation of the CAD system.

RESULTS: At retrospective review, 67% (286 of 427) of screening mammography–detected breast cancers were visible on the prior mammograms. At independent, blinded review by panels of radiologists, 27% (115 of 427) were interpreted as warranting recall on the basis of a statistical evaluation index; and the CAD system correctly marked 77% (89 of 115) of these cases. The original attending radiologists' sensitivity was 79% (427 of [427 + 115]). There was no statistically significant increase in the radiologists' recall rate when comparing the values before (8.3%) with those after (7.6%) installation of the CAD system.

CONCLUSION: The original attending radiologists had a false-negative rate of 21% (115 of [427 + 115]). CAD prompting could have potentially helped reduce this false-negative rate by 77% (89 of 115) without an increase in the recall rate.

Mammography is acknowledged as the single most effective method of screening for breast cancer and is credited with helping to reduce breast cancer mortality by approximately 30% (1–5). The interpretation of screening mammograms is challenging, however. Findings on a screening mammogram leading to further recall are identified in approximately 5%–10% of patients, even though breast cancer is ultimately confirmed in only three to 10 cases in every 1,000 women screened (6). The necessity of viewing a large number of images to detect a small number of cancers, the complex radiographic structure of the breast, the subtle nature of many mammographic characteristics of early breast cancer, and radiologist fatigue or distraction all contribute to false-negative mammographic interpretations.

There is compelling evidence that many breast cancers detected at screening mammography are, in retrospect, visible on the previously obtained mammograms but have been missed by the interpreting radiologist in the prior year (7–11). Although there is no clear consensus as to the actual sensitivity of radiologists in the interpretation of screening mammograms, delayed detection of cancer is a substantial and potentially costly problem because of the associated increased mortality and increased cost of care. Although prior studies (12–18) with limited patient sampling and older mammographic techniques have addressed this issue, to our knowledge there has been no published study with sufficient patient sampling and statistical power in which the false-negative rate of the radiologists was reported. One of our main goals was to perform a more comprehensive study with

sufficient statistical power to determine the false-negative rate of radiologists reading screening mammograms.

To overcome the known limitations of human observers, second (or double) reading of screening mammograms by another radiologist has been implemented at many sites. The results of studies (19–24) indicate a potential 4%–15% increase in the number of cancers detected with double reading. In a radiology practice that performs 10,000 screening examinations per year, generally between 30 and 100 cancers per year will be detected; thus, double reading in this practice could contribute to the diagnosis of 1–15 additional cancers per year.

Rapid and continuing advances in computer technology, as well as the ready adaptation of radiologic images to digital formats, have increased the interest in computer prompting to enable the attending radiologist to act as his or her own second reader (25). One very promising adaptation of computer-prompting technology is computer-aided detection (CAD) in screening mammography (26–28). The current CAD systems demonstrate a high rate of detecting cancerous features on mammograms.

This study of screening mammography was divided into three parts. The first part was a statistical evaluation of the radiologists' sensitivity in detecting action-warranting (ie, actionable), asymptomatic breast cancer on screening mammograms in a large patient population. The assessment of radiologists' sensitivity was determined with community practice–based radiologists in a blinded panel review of consecutive, asymptomatic, previously obtained mammograms with which biopsy-proved cancer was diagnosed at the subsequent mammographic screening. By means of independent quintuple reading of these prior mammograms, which was done without CAD, we determined the number of radiologists who recommended recall of the patient on the basis of sufficiently recognizable mammographic features of cancer. We therefore estimated the number of missed cancers on the basis of these radiologists' determination of what were actionable cancers on these prior mammograms. In the second part of the study, we determined the sensitivity of CAD in identifying cancerous lesions on these same prior screening mammograms.

The results of these two sections of the study led to a determination of the potential benefit of CAD to the radiologist. In defining this potential benefit, we assumed that all cases that would be recalled were those in which imaging depicted sufficiently recognizable mammographic features for cancer that, if prompted, would help the radiologist with his or her assessment. In the third part of the study, we addressed the concern that CAD prompting might lead to unnecessary recalls. Therefore, the change in radiologists' recall rate with CAD was measured. These studies were performed in part as clinical trials for the premarket approval application for a CAD system (ImageChecker M1000, version 1.2; R2 Technology, Los Altos, Calif) that was approved by the U.S. Food and Drug Administration in June 1998.

MATERIALS AND METHODS

Thirteen facilities, including community hospitals, health maintenance organizations, mammography centers, and multispecialty practices, provided case material for the study (listed at the end of the article). Institutional review board approval for this study was obtained at each institution. At none of these sites were mammograms routinely double read. All of the original attending radiologists who interpreted the mammograms met qualification standards specified by the Mammography Quality Standards Act (MQSA) of 1992. A definition of terms used in this study is provided in the Appendix.

Case Material

All of the available screening mammographic studies obtained from the 13 facilities from 1994 through 1996 from which asymptomatic, biopsy-proved cancer was diagnosed (hereinafter referred to as ''current mammograms'')—a total of 1,083 cases—were collected (Fig 1). All cases were in women with a mean age of 62.6 years (age range, 34.0–94.0 years); 14% (156 of 1,083) of these patients were younger than 50 years. All of the available corresponding screening mammograms that were obtained 9–24 months (mean, 14 months) before the current mammograms—a total of 493 cases—also were collected. If more than one prior screening mammogram was available, the most recently obtained one was used. These 493 studies were obtained in women with a mean age of 62.5 years (age range, 34.0–86.0 years); 9% (45 of 493) of these patients were younger than 50 years. Because corresponding prior mammograms were not available for more than half (590 of 1,083) of the current mammograms, the distribution of cancer types (ie, microcalcifications and masses), patient ages, and CAD performance on the current mammograms with (n = 493) and without (n = 590) corresponding prior mammograms were compared to evaluate whether the two data sets were similar (Fig 1).

Visibility in Retrospect

The original two standard screening views—that is, craniocaudal and mediolateral oblique—for each breast (a total of four films per examination) were used for all cases. A radiologist (hereinafter referred to as ''site radiologist'') assessed all of the current mammograms from his or her institution, with knowledge of the location of the biopsy-proved cancer, to (*a*) mark the location of the biopsy-proved lesion or lesions on an overlay, which served as the standard of reference in further evaluations; (*b*) document the lesion's characteristics (ie, calcifications, spiculated masses, nonspiculated masses, and other); and (*c*) provide an assessment recommendation according to the American College of Radiology's BIRADS categories.

Three community practice–based radiologists (hereinafter referred to as ''designated radiologists'') who met MQSA qualification standards were recruited to review the corresponding prior mammograms. These designated radiologists were not otherwise involved in the study. The 493 prior mammograms were divided into three case sets, and each designated radiologist independently reviewed one case set. By using the current mammograms with the overlays as a guide, the designated radiologist determined whether the subsequently diagnosed lesion was visible retrospectively on the prior mammogram. If so, the designated radiologist created a second standard-of-reference overlay, marked the location of the lesion on the prior mammogram, and indicated the lesion characteristics and BI-RADS assessment.

Of the 493 prior mammograms reviewed, 62 cases in which the cancer was found to be visible retrospectively were excluded from the subsequent study because mammographic evidence of previous breast surgery was thought to unduly influence the interpreter. Four other cases were excluded: For this study, the radiologists reviewed the original mammograms because copied images contain less diagnostic information. In these four cases, the original mammograms had to be returned to the originating facility before the last panel review. Thus, a total of 427 prior mammograms were used in this study (hereinafter referred to as ''prior mammograms''), 286 (67%) of which had evidence of the subsequently diagnosed cancer (hereinafter referred to as ''visible prior mammograms'') (Fig 1).

Blinded Assessment by Panel Radiologists

Four panels, each consisting of five community practice–based radiologists who met MQSA qualification standards (hereinafter referred to as ''panel radiologists''), also were recruited to participate in the study. The 20 panel radiologists had a mean of 17 years (range, 3–35 years) of experience practicing mammography and read a mean of 300 (range, 40–1,000) screening mammograms per month. The original films from the 286 visible prior mammograms were divided into four sets, with 64, 70, 71, and 81 cases in each set (Fig 2). Also included in each case set were an additional 45 mammograms—five nonvisible prior mammograms, 20 current mammograms (with biopsy-proved cancers) randomly chosen from those without obvious cancerous features (ie, no current mammograms with a BI-RADS assessment of 5 were included), and 20 mammograms randomly chosen from 100 screening mammograms that were obtained in other patients and confirmed to be normal by having at least one subsequent normal examination. Each panel radiologist reviewed only one case set.

Each case set was read independently in a blinded review by each radiologist to determine whether there was evidence of an ''actionable lesion,'' which was defined as a BI-RADS assessment score of 0, 4, or 5; these scores warrant patient callback for additional imaging evaluation or biopsy. If so, the actionable lesion location was identified on an overlay, and the lesion characteristics and BI-RADS assessment were documented. This assessment was performed without the benefit of comparison with the current mammogram and overlay and without CAD prompting. The number of panel radiologists (range, none to five) who identified the correct location and classification of the lesion on the visible prior mammogram was tabulated. The radiologists' sensitivity and specificity in reading the 20 current mammograms and 20 normal mammograms were calculated. If any radiologist had a sensitivity or specificity of less than 50% (10 of 20) during his or her reading session, he or she was removed from that panel and an additional community practice–based radiologist was recruited to read the case set.

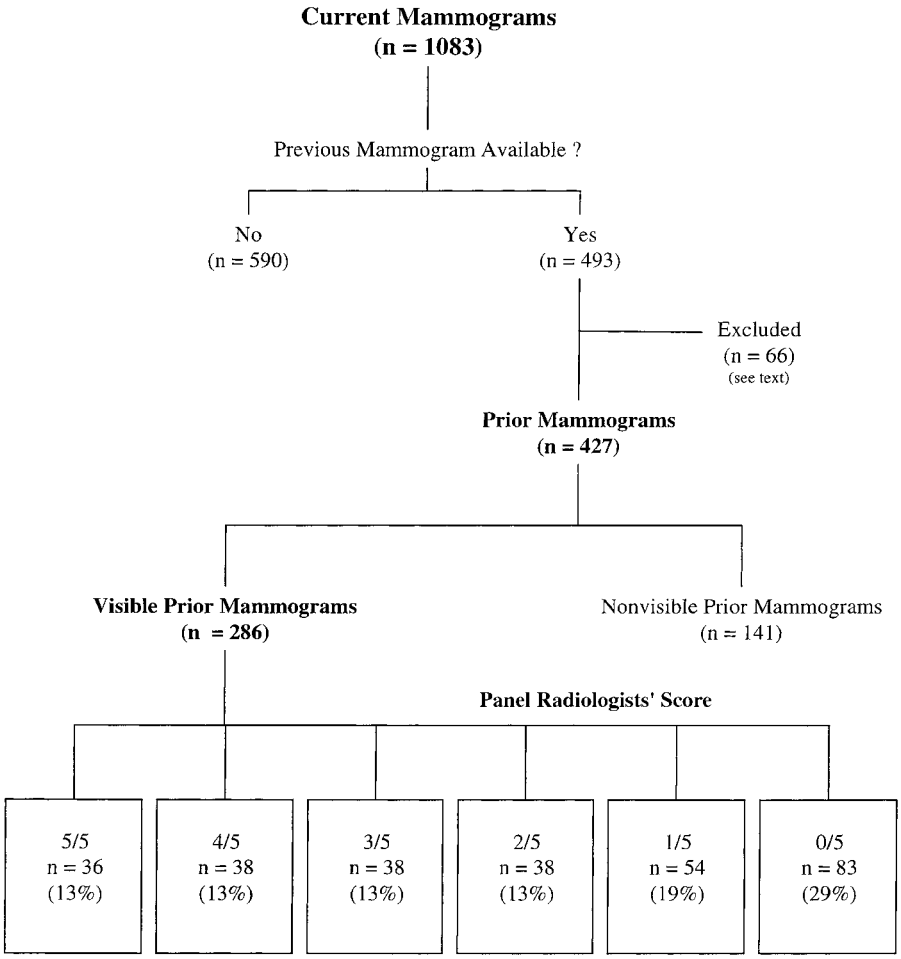

Figure 1. Diagrammatic breakdown of mammogram distribution. Of the 1,083 mammograms on which cancers were detected by using screening mammography (current mammograms), 590 did not have corresponding prior mammograms available. Of the remaining 493 cases with prior mammograms, 66 were excluded (see text). The remaining 427 prior mammograms were retrospectively reviewed with the knowledge of the cancer's subsequent location, of which 286 were considered to be visible retrospectively (visible prior mammograms). The 286 visible prior mammograms were reviewed by the panel radiologists.

	Set 1	Set 2	Set 3	Set 4	Total
Visible Priors	64	71	70	81	286
Invisible Priors	5	5	5	5	20
Normals*	20				20
Currents*	20				20

Figure 2. Composition of case sets read by the panel radiologists. Four panels of five radiologists each were set up to review the four case sets. The number of visible prior mammograms varied among the four case sets, but each of the 286 visible prior mammograms was read by five radiologists. Five different nonvisible prior mammograms were included in each case set as a quality control measure of the designated radiologist's assessment. The 20 biopsy-proved current mammograms and 20 normal mammograms were common to all case sets.

Analysis of Mammograms by the CAD System

The original films of all 1,083 current and 286 visible prior mammograms collected for the clinical study were analyzed by a CAD system (ImageChecker M1000) that digitized the mammograms at 50 μm with 12-bit resolution by using a film digitizer. The digital images were analyzed by using software that highlights regions of interest with the following characteristics: (*a*) clusters of bright spots (ie, regions suggestive of microcalcification clusters), which are marked by a triangle-shaped marker; and (*b*) dense regions, with or without radiating lines, and parenchymal distortions (ie, regions suggestive of spiculated masses, densities,

or architectural distortion), which are marked by an asterisk-shaped marker. These dense regions and distortions were all termed ''masses'' for the analysis described in this article.

The output of the CAD system was compared with the standard-of-reference location (or locations) of the cancer on the current and visible prior mammograms. A CAD mark on either the craniocaudal or mediolateral view in the correct area was scored as correctly marked. If the case had two biopsy-proved cancers, a mark on either cancer was considered to be correct. For lesions with both mass and calcification features that were documented by the radiologist, a marker of either type was counted as correct. The sensitivity of the CAD system was calculated overall for all current and visible prior mammograms.

Prospective Assessment of Change in Recall Rate with the CAD System

The change in radiologist recall rate before and after the installation of the CAD system was measured at five institutions (listed at the end of the article). The sites and participating radiologists all met MQSA qualification standards. The minimum acceptable site volume was 5,000 screening mammography cases per year, and the minimum radiologist reading volume that was acceptable for inclusion in the study was 100 screening mammograms per month. All mammograms included in the study were obtained in women seen for a screening examination.

A historical data review was conducted for a minimum of 4 months at each site before the installation of the CAD system, which tabulated the overall number of screening mammograms and the recall rate for the individual radiologists. The recall rate was defined as the percentage of patients who underwent screening mammography and were recalled immediately for additional imaging or biopsy.

The prospective study was initiated approximately 1 month after the installation of the CAD system. During this time, the medical and technical staffs were trained to operate the device, the system was integrated into the clinic's workflow, and study documentation procedures were implemented. The postinstallation phase of the study lasted a minimum of 4 months. All screening mammograms at a given site were processed on the CAD system, and the radiologist used the CAD information during his or her routine reading sessions before making a final diagnosis. The monthly screening volumes and recall rates were then calculated for each of 14 radiologists. All cases analyzed by the CAD system were recorded in CAD system logs, including the number of regions of interest found, which was used to determine the average number of marks per case. Pre- and postinstallation recall rates were measured. The pre- and postinstallation recall rates and 99% CIs were calculated on the basis of Clopper-Pearson exact likelihood limits, with equal probability in each tail. The 99% level was selected to partially adjust for the use of simultaneous inference involved in the multiple comparisons being performed and to avoid spurious significances that can result from unadjusted levels. In addition, data were analyzed by using the χ^2 test, with continuity corrections for the two-by-two contingency tables for the pre- and postinstallation recall rates, to test the null hypothesis that there was no difference between these rates. Equivalently, the statistical analysis was performed to test for the null hypothesis of independence of the pre- and postinstallation recall rates observed in each of these periods. To adjust for the multiple tests performed, a Bonferroni correction was used for the set of 14 radiologists, not including the nonindependent test on the aggregate numbers. A Bonferroni significance level of .00366 was used for each of the individual hypotheses.

RESULTS

Panel Radiologists' Sensitivity and Specificity

To ensure that the panel radiologists were not unduly influenced by the test environment, the sensitivity and specificity were measured for each panel radiologist by using the 20 current and 20 normal mammograms common to each case set. Three reading sessions were eliminated from the study because the participating radiologist failed to meet the inclusion criteria of having achieved a sensitivity and specificity of 50% or greater. Thus, there were 23 reading sessions to achieve, after exclusions, four panels of five radiologists each. The sensitivity and specificity achieved in the 23 reading sessions are illustrated in Figure 3. The average sensitivity and specificity of the 20 panel radiologists were 84% (334 of 400) and 82% (327 of 400), respectively (Fig 3).

Panel Radiologists' Blinded Review

The panel radiologists' blinded review of the 286 visible prior mammograms from the four case sets is summarized in Table 1. The number of radiologists (from none to five) who correctly identified the lesion's location and classification and considered the case to be actionable (ie, BI-RADS assessment of 0, 4, or 5) provided a likelihood estimate for each case that was interpreted as actionable. A score of 0/5 yielded a 0% probability for further recall; 1/5, a 20% probability; and so forth. By multiplying this probability by the number of visible prior mammograms that received the corresponding score, we calculated an adjusted likelihood value for further recall of the prior mammograms (Table 1). The summed result was 115 cases, which represented the aggregate number of visible prior mammograms that were deemed actionable by the panel radiologists at blinded review (hereinafter referred to as ''actionable prior mammograms'') but were missed by the original attending radiologist. Therefore, the sensitivity of the original radiologists working in their normal clinical environment was 79% (427 of [427 + 115]), with a corresponding false-negative rate of 21% (115 of [427 + 115]). The derivation of sensitivity and false-negative rate are detailed in the Appendix.

We compared the statistically derived result of 115 actionable prior mammograms with the number of visible prior mammograms that were correctly interpreted by the majority (ie, at least three of five) of the panel radiologists. The data in Table 1 indicate that 38, 38, and 36 visible prior mammograms were correctly identified by three (60%), four (80%), and five (100%) of the five panel radiologists, respectively. Therefore, the majority of the panel radiologists correctly identified, characterized, and judged to be actionable 112 (38 + 38 + 36) of the visible prior mammograms; these findings were close to our statistical result of 115 actionable prior mammograms. The CAD results for the majority (ie, at least three of five), supermajority (ie, at least four of five), and 100% consensus (ie, all of five) of the panel radiologists are presented in Table 2.

Performance of CAD System on Current and Visible Prior Mammograms

A total of 906 (84%) of the 1,083 current mammograms were correctly marked by the CAD system (Table 3). Four hundred (99%) of the 406 cancers that contained microcalcifications and 506 (75%) of the 677 cancers that were masses (spicu-

lated and other masses combined, plus other lesions) were correctly marked by the CAD system. Fourteen percent (157 of 1,083 mammograms) of the 177 remaining cases also were marked, but the marks missed the cancerous features designated by the site radiologist; 20 (2%) of the 1,083 mammograms had no marks. There was no perceptible difference in lesion characteristics or CAD sensitivity between the current mammograms with and those without available prior mammograms (Table 4).

One hundred seventy-one (60%) of the 286 visible prior mammograms were correctly marked by the CAD system (Tables 1 and 3). The CAD system successfully identified the majority of all the visible cancers that were detected by at least one of the five panel radiologists at blinded review and approximately one-third of the cancers that no panel radiologist detected. The percentage of correct markings increased as the panel radiologists' score increased (ie, 0/5 to 5/5); this indicated that the more obvious the lesion's characteristics were, the more likely that it would be detected by both the panel radiologists and the CAD system. If these 171 CAD-detected cancers were scaled to the actionability of the panel radiologists, the potential CAD benefit would be 89 cancers (Table 1). Thus, we calculated that 89 (77%) of 115 cancers that were considered to be actionable by the panel radiologists but were missed by the original attending radiologist at screening would be detected by the CAD system.

In this study, the CAD system placed an average of 1 mark per film on mammograms that did not depict cancerous lesions (4.1 marks per standard two-view bilateral screening mammogram).

Effect of CAD on Radiologists' Recall Rate

Fourteen radiologists from five facilities were involved in the study to measure radiologists' recall rate with the CAD system. According to historical data review, these 14 radiologists interpreted a total of 23,682 screening studies, with an overall recall rate of 8.3% (1,961 of 23,682). In the prospective portion of the study, these same radiologists, with the aid of CAD prompting, interpreted 14,817 screening mammograms and had an overall recall rate of 7.6% (1,126 of 14,817) (Table 5). There was no statistically significant increase in the radiologists' recall rate as a group or individually with the use of CAD at the threshold level used in this study. (See text regarding pre- and postinstallation rates at end of Materials and Methods section.)

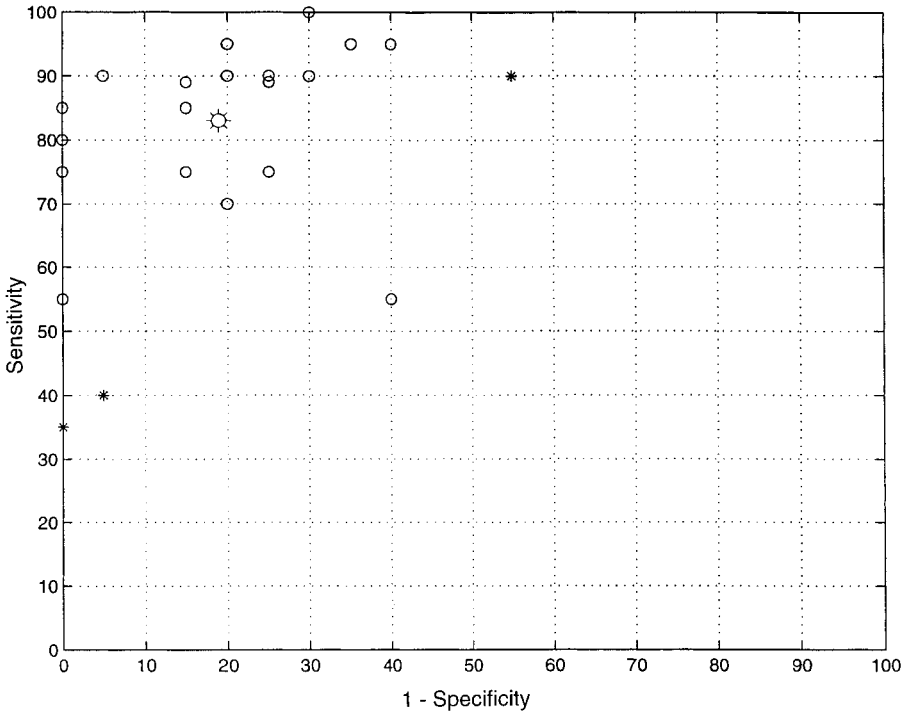

Figure 3. Sensitivity versus specificity of the panel radiologists in reviewing visible prior mammograms. ○ = values for the 20 panel radiologists. ∗ = values for additional three radiologists who were eliminated from the study because of low sensitivity or low specificity (see text). ⌑ = the average sensitivity (84%) and specificity (82%) of the 20 remaining radiologists.

TABLE 1
Panel Radiologists' Review of and CAD Performance on 286 Visible Prior Mammograms

Panel Radiologist Score*	Visible Prior Mammograms			Potential CAD Benefit‡
	No.	No. Actionable†	No. Correctly Marked with CAD	
0/5 (0)	83 (29)	0.0	28 (34)	0.0
1/5 (20)	53 (19)	10.6	28 (53)	5.6
2/5 (40)	38 (13)	15.2	24 (63)	9.6
3/5 (60)	38 (13)	22.8	28 (74)	16.8
4/5 (80)	38 (13)	30.4	30 (79)	24.0
5/5 (100)	36 (13)	36.0	33 (92)	33.0
Total	286 (100)	115	171 (60)	89

Note.—The numbers in parentheses are percentages.

* Data are the number of radiologists (of a total of five) who correctly identified the lesion's location and classification.

† The number of visible prior mammograms that were actionable was calculated by multiplying the panel radiologist score by the corresponding number of visible prior mammograms.

‡ The potential CAD benefit was calculated by multiplying the panel radiologist score by the number of visible prior mammograms correctly marked with CAD.

DISCUSSION

To our knowledge, the present study is the largest retrospective review of prior screening mammograms in patients whose subsequent (ie, current) screening mammogram led to a diagnosis of breast cancer. The case material was collected from a variety of mammography practices and interpreted by community practice–based radiologists with a variety of backgrounds who met MQSA qualification standards. All cases were performed no earlier than 1992, the year the MQSA was passed and 5 years after the initiation of the Ameri-

TABLE 2
Panel Radiologists Review of and CAD Performance on 286 Visible Prior Mammograms

Panel Radiologist Score*	Microcalcifications		Masses		Total	
	No.	CAD Marked	No.	CAD Marked	No.	CAD Marked
Majority	39	37 (95)	73	54 (74)	112	91 (81)
Supermajority	26	25 (96)	48	38 (79)	74	63 (85)
100% consensus	13	13 (100)	23	20 (87)	36	33 (92)

Note.—The numbers in parentheses are percentages.

* A majority score was that when three to five of the five radiologists correctly identified the lesion's location and classification. A supermajority score was that when four or five of the five radiologists correctly identified the lesion's location and classification. A 100% consesus score was that when all five radiologists correctly identified the lesion's location and classification.

TABLE 3
Summary of Mammographic Lesion Characteristics and CAD Performance

Lesion Characteristic	Current Mammograms		Visible Prior Mammograms	
	No.	No. Correctly Marked by CAD	No.	No. Correctly Marked by CAD
Microcalcifications	406 (37)	400 (99)	110 (38)	87 (79)
Masses	677 (63)	506 (75)	176 (62)	84 (48)
Total	1,083 (100)	906 (84)	286 (100)	171 (60)

Note.—The numbers in parentheses are percentages.

TABLE 4
Lesion Characteristics and CAD Performance on Current Mammograms

Factor	Current Mammograms		
	With Prior Mammograms	Without Prior Mammograms	Total
Lesion characteristics			
Microcalcifications	174/493 (35)	232/590 (39)	406/1,083 (37)
Masses	319/493 (65)	358/590 (61)	677/1,083 (63)
CAD detection			
Microcalcifications	170/174 (98)	230/232 (99)	400/406 (99)
Masses	227/319 (71)	279/358 (78)	506/677 (75)
Total	397/493 (81)	509/590 (86)	906/1,083 (84)

Note.—All data are the number of mammograms. The numbers in parentheses are percentages.

can College of Radiology Mammography Accreditation Program (29). The results of this study indicate that a large proportion of breast cancer was visible—on 286 (67%) of 427 studies—retrospectively on prior mammograms obtained 9–24 months (mean, 14 months) earlier and that an important percentage of these prior mammograms, 27% (115 of 427), were considered to be actionable by radiologists at blinded, independent review.

The topic of retrospectively visible screening mammography–detected breast cancers has been addressed extensively in the literature (Table 6). The data from the study by Harvey et al (11) most closely parallel those in the present study. They found that 55 (75%) of 73 mammograms depicted cancers that were retrospectively visible on the most recently obtained prior mammograms by at least one of three radiologists who knew the cancer's eventual location, whereas 26 (36%) of 73 of the prior mammograms were found to be actionable by two different radiologists at true prospective blinded readings. In the current study, the case sets analyzed by the panel radiologists were enriched with cases in which 9–24 months later (mean, 14 months) the patients were given a diagnosis of cancer, unlike the case distribution in a normal clinical environment. Yet our findings of retrospectively visible (on 286 [67%] of 427 mammograms) and actionable (on 115 [27%] of 427 mammograms) cancers were similar to those in the Harvey et al study (Table 6).

We considered that the enriched data set may have influenced the reading behavior of the panel radiologists. Thus, we measured the sensitivity and specificity of each panel radiologist (Fig 3) and had results that were similar to those in a previous study, in which the sensitivity was 73%–78% and the specificity was 88%–90% (30); these results were based on a much more dilute test set (302 cancers in 2,578 cases). In the current study, the largest number of visible prior mammograms, 83 (29%) of 286, was in the 0/5 actionability category, in which none of the five panel radiologists reviewing the studies considered the findings to be actionable. These results and a comparison of them with the results in the Harvey et al study suggest that the panel radiologists were able to nearly maintain their usual clinical threshold in what was clearly a ''cancer-enriched'' case set. We excluded 62 cases of visible prior mammograms that showed evidence of previous surgery. Our concern was that the mammographic evidence of previous surgery might influence the radiologist's interpretation because of knowledge that he or she was working with a cancer-enriched case set.

In our study, similar to that of Harvey et al (11), we found that although a majority of cancers could be seen in retrospect, a much smaller percentage of the retrospectively visible cancers were considered to be actionable at blinded review (Table 6). For this reason, we did not estimate the CAD benefit by comparing the CAD results for all visible prior mammograms, but rather we estimated its benefit by comparing the CAD results for only the smaller subset of actionable prior mammograms. Furthermore, the derived benefit of CAD was scaled to the actionability of the panel radiologists to produce a more conservative estimate. For example, if none of the five panel radiologists (score 0/5) identified the lesion as being suspicious, then none of the 28 (34%) of 83 correct CAD markings were considered in the calculation of CAD benefit. Similarly, if only one of the panel radiologists correctly identified the lesion as cancer, then only 20% (1/5) of the correct CAD markings were considered in the calculation of CAD benefit, and so

forth. It was only when the panel radiologists had a 100% consensus (5/5) interpretation of the lesion that we assumed the original radiologist would have acted on the CAD mark, if present.

We also determined that 112 of the visible prior mammograms were judged to be actionable by a majority (at least three of five) of the panel radiologists. Our statistically derived calculation of 115 actionable prior mammograms was very similar to these 112 cases. In summary, the CAD benefit was derived from only those cases that would have been worked up, if seen—not from the larger subset of retrospectively visible cases.

The reasons these retrospectively visible and actionable cancers did not result in a work-up by the original attending radiologists are unknown, but some have categorized them as oversights or errors in interpretation (31). Not surprisingly, such error rates have been shown to decrease with multiple readings (19–24). Another way to decrease these errors is to use CAD to prompt the radiologist to reassess those features exhibiting strong characteristics of cancer.

By collecting all the available prior mammograms obtained before the current mammograms, we were able to calculate the sensitivity of the original attending radiologists' interpretation by finding a subset of prior mammograms on which the panel radiologists found actionable cancer. In previous studies (12–18,32–40) involving smaller patient populations and older mammographic techniques with less controlled conditions, radiologists' sensitivity in reading mammograms has been estimated to be 83%–95% for first mammography screenings and 56%–86% for subsequent screenings.

The total number of mammographic examinations performed at the participating institutions during our study was calculated to be 290,000 cases, extrapolated from the 1,083 screening-detected current mammograms, by using a ratio for subsequent screening of 3.8 cancers detected per 1,000 screening mammograms (39), which is the expected incidence in a screened population. (Note: With three to 10 cancers detected per 1,000 screening mammograms, the original number of screening mammograms on which the 1,083 cancers were detected on the current mammograms ranged from roughly 110,000 to 360,000.) The 79% sensitivity of the original attending radiologists in interpreting mammograms in our study was calculated with a much larger study population and without case selection. It is likely that the majority of the current mammograms were not from the first screening examination, because we found no perceptible difference in the distribution of lesion characteristics, patient age, or CAD sensitivity in the subset of current mammograms with and the subset without corresponding prior mammograms (Table 4). First screening mammography studies should have a higher radiologist sensitivity, because the number of obvious cancers is likely to be greater in an initially screened population. In addition, only 14% of the current mammograms were obtained in women younger than 50 years.

We were concerned that a potential increase in false-positive interpretations might result from the use of CAD and lead to unnecessary recalls. Therefore, we measured the effect of the CAD system on radiologists' recall rate in the prospective clinical portion of the study (Table 5). The results of this analysis showed no significant increase in the recall rates of the 14 radiologists, who met MQSA qualification standards and were from a variety of clinical practices, before or after using the CAD system. This was not surprising, because the described CAD system does not detect cancer-compatible findings that are not visible to the radiologist on a mammogram. Instead, the CAD system simply marks findings that are at times overlooked during a reading session and that, once brought to the attention of the radiologist, will prompt reevaluation of that area on the mammogram. Thus, the false-positive marks by the CAD system are readily dismissed by the radiologist as being representative of an unimportant finding (ie, vascular or other benign calcification) or an area that, at repeated review, will show no evidence of anything that would prompt a recall.

The results of CAD analysis of the 1,083 current mammograms and 286 visible prior mammograms were very encouraging. The algorithms used by the CAD system were particularly sensitive in detecting microcalcifications that represented biopsy-proved cancer (99% sensitivity for the 406 current mammograms depicting microcalcifications), whereas the sensitivity for the detection of masses

TABLE 5
Effect of CAD on Radiologists' Recall Rate

Radiologist No.	Before CAD Installation		After CAD Installation		Net Change (%)
	No. of Cases	Recall Rate (%)	No. of Cases	Recall Rate (%)	
1	1,594	8.2	571	9.8	1.6
2	1,743	2.2	1,097	3.3	1.1
3	3,572	8.8	1,253	8.9	0.1
4	1,391	5.1	522	3.8	−1.3
5	1,293	10.0	644	9.5	−0.5
6	4,378	9.2	2,489	7.7	−1.5
7	574	5.2	889	4.9	−0.3
8	767	6.3	710	7.5	1.2
9	1,074	6.2	714	5.0	−1.2
10	619	5.2	1,170	3.5	−1.7
11	2,971	8.1	1,677	9.1	1.0
12	1,255	15.2	800	11.3	−3.9
13	971	13.0	675	10.5	−2.5
14	1,480	9.6	1,606	10.1	0.5
All	23,682	8.3	14,817	7.6	−0.7

TABLE 6
Retrospective Analyses of Screen-Detected Cancers

Reference Study	No. of Cases	No. of Cancers Visible in Retrospect*	No. of Actionable Cancers*
Harvey et al (11)	73	55 (75)	26 (36)†
van Dijck et al (9)	44	25 (57)	. . .
Current study	427	286 (67)	115 (27)

* The numbers in parentheses are percentages.

† The 26 actionable cancers were derived, by using the technique described in this study, from 22 cases in which cancers were detected by both radiologists (100% weight) and eight cases in which cancers were detected by one radiologist (50% weight).

was lower (75% sensitivity for the 677 current mammograms depicting masses) (Table 3). The CAD system correctly marked 171 (60%) of the 286 visible prior mammograms (Table 1), and, importantly, it correctly marked 81% of the visible prior mammograms that were found to be actionable by the majority (at least three of five) of the panel radiologists (Table 2). It is important to reemphasize that all visible (and nonvisible) prior mammograms were the most recent screening examinations performed in patients who had a diagnosis of cancer on the basis of subsequent screening findings.

Not surprisingly, the CAD system marked an increasing number of lesions on the visible prior mammograms as the score of the five panel radiologists increased (ie, 34% of lesions that 0/5 radiologists detected to 92% lesions that 5/5 radiologists detected) (Table 1); these findings indicate that the more obvious the lesion characteristics were, the more likely that the lesion would be detected by the CAD system and the radiologists. However, on 177 (16%) of the 1,083 current mammograms (20 cases with no marks and 157 cases with marks in the wrong location), CAD did not detect cancers that were diagnosed by the original radiologists (Table 3). Thus, this technology currently must be viewed as a prompting aid to the radiologist to initiate a second review and not as a primary screening method.

Of importance in our evaluation was that the 115 cancers that were considered to be actionable by the panel radiologists were detected without CAD prompting, which indicates that there was sufficient evidence of cancer that was either not identified or, if seen, not acted on when the visible prior mammograms were first interpreted by the original radiologist. These results also indicate that not all retrospectively visible cancers are deemed actionable at blinded review. However, if sufficient attention is directed to the visible and actionable lesions in true clinical settings, then these lesions might be addressed and more cancers diagnosed earlier.

In this study we showed that the described CAD system, by selecting mammographic findings that warrant immediate repeated reading by the radiologist, offers the potential for improved cancer detection. However, double reading of mammograms by two different radiologists or by the same radiologist in two separate reading sessions also has been shown to result in increased sensitivity compared with one-pass interpretation by a single radiologist (19–24). Therefore, an important follow-up to this study will be a comparison between a single-radiologist reading with CAD prompting and a double-radiologist reading, in which the sensitivity, specificity, performance time, and costs of these two alternative approaches to enhanced mammographic interpretation are measured. The approach that leads to superior clinical results or that does so at lower cost is likely to prevail.

We conducted what is, to the best of our knowledge, the largest retrospective review of prior mammograms in a consecutive population of patients who had no symptoms and subsequently had a diagnosis of breast cancer on the basis of screening mammography findings. The results indicate that on 286 (67%) of 427 prior mammograms, breast cancers were visible retrospectively and that by using multiple readers (ie, a panel of radiologists), the cancers on 115 (27%) of these 427 prior mammograms could have been detected. The original attending radiologists' sensitivity for the detection of actionable lesions was 79%, with a false-negative rate of 21%. It is cases such as these 115 detectable cancers that should be targeted to increase the yield of screening mammography with appropriate prompting techniques, such as CAD (25) or multiple readings (19). The described CAD system successfully identified a large portion (77%) of these detectable and actionable cancers. Importantly, the use of a CAD system does not increase the work-up rate of the radiologists. The contributions of CAD, which is in the early stages of implementation, to complex image interpretation in mammography and to all fields of medicine is likely to be substantial.

APPENDIX

Definitions of Terms

Breast Imaging Reporting and Data System (BI-RADS) scores.—0, need additional imaging evaluation; 1, negative mammogram; 2, benign finding; 3, probably benign finding, short-interval follow-up suggested; 4, suspicious abnormality, biopsy should be considered; 5, highly suggestive of malignancy, appropriate action should be taken.

*Current mammograms (*n = *1,083).*—Screening mammograms that were obtained between 1994 through 1996 from 13 facilities on which asymptomatic, biopsy-proved cancer was diagnosed.

*Prior mammograms (*n = *427).*—The most recently obtained mammograms that corresponded to the current mammograms and were obtained 9–24 months before the current mammograms.

*Visible prior mammograms (*n = *286).*—Prior mammograms on which a finding compatible with cancer was visible when retrospectively compared with the corresponding current mammogram on which the location of the cancer was known.

Actionable mammograms.—Mammograms interpreted by a radiologist to have a BI-RADS assessment score of 0, 4, or 5 (requiring further imaging evaluation or biopsy).

*Actionable prior mammograms (*n = *115).*—Prior mammograms that were interpreted by the panel radiologists at blinded review to be actionable.

Original radiologists.—The radiologists who initially interpreted the prior mammograms as having no evidence of cancer. The sensitivity calculation of 79% was based on these radiologists' original interpretation.

Site radiologists.—The radiologists who reviewed the current mammograms to determine the location of the biopsy-proved lesion on an overlay, documented the lesion's characteristics, and provided a BI-RADS assessment of the lesion.

Designated radiologists.—The radiologists who compared the prior mammograms with the current mammograms, with knowledge of the biopsy-proved cancer's location as defined on the overlay, to assess the visibility of the cancer retrospectively on the prior mammogram.

Panel radiologists.—The 20 radiologists (five each who independently reviewed four case sets) who reviewed the visible prior, normal, and current mammograms to assess the actionability of the mammogram.

Computation of Radiologists' Sensitivity

The 115 actionable prior mammograms represented 27% (115/427) of all prior mammograms. Recall that all 427 prior mammograms had corresponding subsequent current mammograms that were obtained 9–24 months later (mean, 14 months) and on which the cancer was detected in the current year. If we assume constancy in the demographics of the population they served and in the number of patients screened between the prior and current examinations, then we can reasonably assume that these radiologists detected 427 other cancers in the prior year. Because we know that these radiologists also missed cancers on the 115 actionable prior mammograms in the prior year, we can calculate the total number of detectable cancers in the prior year as 427 detected cancers plus 115 missed but actionable cancers, which equals 542 detectable cancers. Therefore, the sensitivity and false-negative rate of these radiologists, S_{rad} and FN_{rad}, respectively, in a normal clinical environment can be calculated from this data set

as follows: S_{rad} = 427 detected cancers/542 detectable cancers = 79% and $FN_{rad} = 1 - S_{rad} = 21\%$.

Facilities that provided case material: Alta Breast Center, Oakland, Calif; David Grant Medical Center, Travis Air Force Base, Calif; El Camino Hospital, Mountain View, Calif; Kaiser Permanente Medical Center, Sacramento, Calif; Kaiser Permanente Medical Center, South San Francisco, Calif; Park Nicollet Clinic, Minneapolis, Minn; Sequoia Hospital, Redwood City, Calif; Walnut Creek Radiology, Walnut Creek, Calif. The following facilities also provided case material and participated in the prospective assessment of radiologists' recall rate: Kaiser Permanente Medical Center, Redwood City, Calif; Kaiser Permanente Medical Center, San Francisco, Calif; Ochsner Clinic, New Orleans, La; Susan G. Komen Breast Center, Dallas, Tex; and Vanderbilt University Medical Center, Nashville, Tenn.

Acknowledgments: We acknowledge the help of the staffs at the participating institutions involved in the study. We also thank the following radiologists who participated in this study: Lurleen Brown, MD, Joanne Brumbaugh, MD, Vincent Burke, MD, Sherry Butler, MD, John Carroll, MD, Gunnar Cederbom, MD, Crandon Clark, MD, Richard Dorsay, MD, W. Phil Evans, MD, Jonathan Fish, MD, Richard Gabbay, MD, Bill Karras, MD, Deborah Kass, MD, Alan Kaufman, MD, Mary Lechner, MD, Christopher Merritt, MD, Michael Nelson, MD, John Rennick, MD, Michael Rizzo, MD, Daniel Seidler, MD, and Harold Wadley, MD. The authors also acknowledge the efforts of William Sacks, MD, PhD, of the U.S. Food and Drug Administration in developing, during the review of the premarket approval application for the ImageChecker, the method used to determine the probability that a radiologist would work up a lesion that is marked by the ImageChecker (Table 1) and the efforts of John Kennedy, MS, for statistical support and Roger H. Schneider, MS, Alan A. Stein, PhD, and S.P. Bob Wang, MS, for valuable input.

References

1. Shapiro S. Screening: assessment of current studies. Cancer Suppl 1994; 74:231–238.
2. Feig S. Increased benefit from shorter screening mammography intervals for women ages 40-49 years. Cancer 1997; 80:2035–2039.
3. Feig S, D'Orsi C, Hendrick R, et al. American College of Radiology guidelines for breast cancer screening. AJR Am J Roentgenol 1998; 171:29–33.
4. Leitch A, Dodd G, Costanza M, et al. American Cancer Society guidelines for the early detection of breast cancer: update 1997. CA Cancer J Clin 1997; 47:150–153.
5. Tabar L, Fagerberg G, Chen HH, et al. Efficacy of breast cancer screening by age: new results from the Swedish two-county trial. Cancer 1995; 75:2507–2517.
6. Sickles E. Quality assurance: how to audit your own mammography practice. Radiol Clin North Am 1992; 30:265–275.
7. Ikeda D, Andersson I, Wattsgard C, Janzon L, Linell F. Interval carcinomas in the Malmo mammographic screening trial: radiographic appearance and prognostic considerations. AJR Am J Roentgenol 1992; 159:287–294.
8. Duncan A, Wallis M. Classifying interval cancers. Clin Radiol 1995; 50:774–777.
9. van Dijck J, Verbeek A, Hendriks J, Holland R. The current detectability of breast cancer in a mammographic screening program: a review of the previous mammograms of interval and screen-detected cancers. Cancer 1993; 72: 1933–1938.
10. Burhenne H, Burhenne L, Goldberg F, et al. Interval breast cancers in the screening mammography program of British Columbia: analysis and classification. AJR Am J Roentgenol 1994; 162:1067–1071.
11. Harvey J, Fajardo L, Innis C. Previous mammograms in patients with impalpable breast carcinoma: retrospective vs blinded interpretation. AJR Am J Roentgenol 1993; 161:1167–1172.
12. Peeters P, Verbeek A, Hendricks J, van Bon M. Screening for breast cancer in Nijmegen: report on 6 screening rounds 1975–1986. Int J Cancer 1989; 43:226–230.
13. Miller A, Baines C, To T, Wall C. Canadian national breast screening study. II. Breast cancer detection and death rates among women aged 50-59 years. Can Med Assoc J 1992; 147:1477–1488.
14. Miller A, Baines C, To T, Wall C. Canadian national breast cancer screening study. I. Breast cancer detection and death rates among women aged 40 to 49 years. Can Med Assoc J 1993; 147:1459–1476.
15. Baines C, Miller A, Bassett A. Physical examination: its role as a single screening modality in the Canadian national breast screening study. Cancer 1989; 63:1816–1822.
16. Baines C, Miller A, Kopans D, et al. Canadian National Breast Screening Study: assessment of technical quality by external review. AJR Am J Roentgenol 1990; 155:743–747.
17. van Rosen A, Frisell J, Glas U, et al. Nonpalpable invasive breast carcinomas from the Stockholm screening project. Acta Oncol 1989; 28:23–27.
18. Lidbrink E, Tornberg S, Azavedo E, et al. The general mammography screening program in Stockholm: organization and first round results. Acta Oncol 1994; 33:353–358.
19. Thurfjell E, Lernevall K, Taube A. Benefit of independent double reading in a population-based mammography screening program. Radiology 1994; 191:241–244.
20. Anttinen I, Pamilo M, Soiva M, Roiha M. Double reading of mammography screening films—one radiologist or two? Clin Radiol 1993; 48:414–421.
21. Beam C, Sullivan D, Layde P. Effect of human variability on independent double reading in screening mammography. Acad Radiol 1996; 3:891–897.
22. Ciatto S, Del Turco M, Morrone D, et al. Independent double reading of screening mammograms. J Med Screen 1995; 2:99–101.
23. Warren RM, Duffy SW. Comparison of single reading with double reading of mammograms, and change in effectiveness with experience. Br J Radiol 1995; 68:958–962.
24. Brown J, Bryan S, Warren R. Mammography screening: an incremental cost effectiveness analysis of double versus single reading of mammograms. BMJ 1996; 312:809–812.
25. Doi K, Giger M, MacMahon H, et al. Computer-aided diagnosis: development of automated approaches for quantitative analysis of radiographic images. Semin Ultrasound CT MR 1992; 13:140–152.
26. Vyborny CJ, Giger ML. Computer vision and artificial intelligence in mammography. AJR Am J Roentgenol 1994; 162:699–708.
27. Vyborny CJ. Can computers help radiologists read mammograms? Radiology 1994; 191: 315–317.
28. Winsberg F, Elkin M, Macy J, Bordaz V, Weymouth W. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. Radiology 1967; 89:211–215.
29. Hendrick R. Quality assurance in mammography: accreditation, legislation, and compliance with quality assurance standards. Radiol Clin North Am 1992; 30:243–255.
30. Kerlikowske K, Grady D, Barclay J, et al. Variability and accuracy in mammographic interpretation using the American College of Radiology Breast Imaging Reporting and Data System. J Nat Cancer Inst 1998; 90:1801–1809.
31. Bird R, Wallace T, Yankaskas B. Analysis of cancers missed at screening mammography. Radiology 1992; 184:613–617.
32. Nystrom L, Rutqvist L, Wall S, et al. Breast cancer screening with mammography: overview of Swedish randomised trials. Lancet 1993; 341:973–978.
33. Tabar L, Fagerberg C, Gad A, et al. Reduction in mortality from breast cancer after mass screening with mammography: randomized trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet 1985; 1:829–832.
34. Tabar L, Fagerberg C, Duffy S, Day N. The Swedish two-county trial of mammographic screening for breast cancer: recent results and calculation of benefit. J Epidemiol Community Health 1989; 43:107–114.
35. Tabar L, Fagerberg G, Duffy S, Day N, Gad A, Grontoft O. Update of the Swedish two-county program of mammography screening for breast cancer. Radiol Clin North Am 1992; 30:187–210.
36. Tabar L, Fagerberg C, Day N, Duffy S. The Swedish two-county trial of mammographic screening of breast cancer: recent results on mortality and tumour characteristics. Pathol Biol 1992; 39:846.
37. Holmberg L, Tabar L, Adami H, Bergstrom R. Survival in breast cancer diagnosed between breast cancer screening examinations. Lancet 1986; 2:27–30.
38. Kerlikowske K, Grady D, Barclay J, Sickles E, Ernster V. Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 1996; 276: 33–38.
39. Kerlikowske K, Grady D, Barclay J, Sickles E, Ernster V. Likelihood ratios for modern screening mammography. JAMA 1996; 276: 39–43.
40. Mushlin A, Kouides R, Shapiro D. Estimating the accuracy of screening mammography: a meta-analysis. Am J Prev Med 1998; 14:143–153.