

Computer-Aided Detection with Full-Field Digital Mammography



Assessment
Program
Volume 21, No. 3
May 2006

Executive Summary

Background

Computer-aided detection (CAD) with full-field digital mammography (FFDM) is proposed as an adjunct to radiologists' reading of digital mammograms. In the fall of 2005, the results of the American College of Radiology Imaging Network's (ACRIN) Digital Mammographic Imaging Screening Trial (DMIST) were released. The trial showed with reasonable certainty that there was no difference in accuracy between FFDM and screen-film mammography (SFM) for asymptomatic women in general. For three subgroups of women, however, FFDM performed better than SFM: women under age 50; pre- or perimenopausal women; and women with heterogeneously dense or extremely dense breasts.

An Assessment on CAD in SFM (Vol. 17, No. 17; December 2002) concluded that CAD improves the accuracy of SFM. Specifically, it improves net health outcome compared with single-reader radiologist interpretation by increasing the true-positive rate without a disproportionate increase in the false-positive rate. The Assessment also concluded that there was insufficient evidence on the use of CAD with FFDM because of the lack of conclusive evidence at the time on the benefits of FFDM in general, as well as the fact that the use of CAD with FFDM was sufficiently different from its use with SFM that evidence on the latter cannot be extrapolated to the former. Now that the performance of FFDM itself has been demonstrated, the question is whether the use of CAD can improve the sensitivity and specificity of digital mammography, which continues to be lower than optimal.

While there are conceptual similarities between the application of CAD to a digitized screen-film mammogram and to a directly acquired digital mammogram, there are some important differences. On the one hand, screen-film mammograms must be digitized before the CAD algorithms are applied. This process can introduce artifacts as a result, (e.g., of dust particles on the film). Digital mammograms are captured directly, and therefore one might expect them to be more accurate. On the other hand, there are more data in a raw digital mammogram than can be shown in a single display format. This difference may permit interaction between the CAD software and the digital mammography data being displayed. Whether this flexibility provides similar, improved, or worsened diagnostic performance will depend on how these interactions are optimized in commercially available products.

Objective

To evaluate the impact of using CAD vs. single reading on the sensitivity, specificity, and biopsy rates for FFDM.



BlueCross
BlueShield
Association

An Association
of Independent
Blue Cross and
Blue Shield Plans



NOTICE OF PURPOSE: TEC Assessments are scientific opinions, provided solely for informational purposes. TEC Assessments should not be construed to suggest that the Blue Cross Blue Shield Association, Kaiser Permanente Medical Care Program or the TEC Program recommends, advocates, requires, encourages, or discourages any particular treatment, procedure, or service; any particular course of treatment, procedure, or service; or the payment or non-payment of the technology or technologies evaluated.

Search Strategy

A literature search was conducted on MEDLINE covering the period December 2005 through March 2006, using the following search terms: “(mammography(mh) OR mammogra*) AND (radiographic image interpretation, computer-assisted(mh) OR diagnosis, computer-assisted(mh) OR computer-assisted OR computer-aided OR CAD,” limited to English language. The search yielded 400 references, which were searched for studies on the impact on sensitivity, specificity, and/or biopsy rates of the use of CAD with digital mammography.

Selection Criteria

Articles had to compare the results of single reading of full-field digital mammography images with and without a subsequent review of the results of the CAD program. Of particular interest were articles that looked at the impact of CAD on the sensitivity, specificity, and biopsy rate associated with FFDM. Additionally, studies had to report on cancer detection based on histologic results for at least some of the patients in the sample.

Main Results

Unfortunately, no high-quality articles in peer-reviewed journals assessed the use of CAD as an adjunct to FFDM. Perhaps the strongest study was reported in an abstract from 2005, but there was not sufficient detail to assess the validity and generalizability of the results. For example, it is not clear how the sample of patients was amassed (and therefore how representative it is of the populations of interest) and whether the study was prospective or retrospective. A second study looked at the use of a commercial CAD system, but it did not report the incremental effect of CAD in addition to the initial reading by a radiologist. Several articles reported on the development of new CAD systems, but did not provide solid performance data. In addition, the primary focus of this Assessment is on commercially available CAD systems. A clinical trial on the use of CAD is just beginning in Taiwan, with plans to enroll 3,000 patients, although a focus on one ethnic group might limit its generalizability. It is not possible at this time to judge the impact of CAD on single reading of digital mammograms at this time. Therefore, the impact of CAD on cancer detection, treatment, and survival is unknown as well.

Author’s Conclusions and Comments

While evidence shows that the use of CAD with screen-film mammography is equal to or better than single reading of the SFM images, there is scant information on the performance of CAD with FFDM. Logically, it might seem that CAD should play the same role with FFDM as with SFM, but the differences between film and digital mammography—which show up in the greater accuracy of FFDM in certain populations—preclude extrapolating from the impact of CAD with SFM to CAD with FFDM. The large increase in the magnitude of the data collected by FFDM, the ability to fine-tune the digitally acquired images, and the elimination of the digitization step make FFDM sufficiently different from SFM that separate studies on the impact of CAD on FFDM are needed. Until results from better studies focusing on the use of CAD with FFDM become available, the benefits of CAD with FFDM cannot be determined.

Based on the available evidence, the Blue Cross and Blue Shield Association Medical Advisory Panel made the following judgments about whether the use of computer-aided detection (CAD) as an adjunct to single reading of full-field digital mammography meets the Blue Cross and Blue Shield Association’s Technology Evaluation Center (TEC) criteria.

1. The technology must have final approval from the appropriate governmental regulatory bodies.

Two CAD systems have received premarket application (PMA) approval by the U.S. Food and Drug Administration (FDA) for use with FFDM.

2. The scientific evidence must permit conclusions concerning the effect of the technology on health outcomes.

While there is a growing literature on the use of CAD with screen film mammography, there are very few articles that look at CAD with FFDM. Because the FDA approved the CAD devices for use with FFDM as a modification of their original approval of the CAD systems for use with SFM, whatever data were used to support the request for pre-market approval of these CAD systems with FFDM are not readily available. The articles on the use of CAD with FFDM are inadequate to determine the incremental impact on sensitivity, specificity, and biopsy rates.

In summary, the available evidence is considered insufficient to permit conclusions on the effect on relevant outcomes of using CAD after initial radiographic interpretation as a quality adjunct to single-reader mammography in patients having full-field digital mammography for screening or diagnostic purposes.

3. The technology must improve the net health outcome; and

4. The technology must be as beneficial as any established alternatives.

The available data from well-conducted studies are insufficient to make a determination of whether adding CAD to FFDM leads to diagnoses that are as or more accurate than reading the FFDM images alone. Given the lack of data on these intermediate outcomes, it also is not possible to determine the impact of CAD on health outcomes such as treatment success, recurrence rates, and survival. As a result, it is unclear whether health outcomes are the same or better with the use of CAD with FFDM compared to single reading of FFDM alone.

5. The improvement must be attainable outside the investigational settings.

Given the inability to determine whether the use of CAD improves health outcomes, considering the generalizability of the impact of CAD is premature.

For the above reasons, the use of CAD systems as an adjunct to single reading of full-field digital mammography images does not meet the TEC criteria.

Contents

Assessment Objective	5	Review of Evidence	11
Background	5	Conclusion	13
Methods	9	Summary of Application of the Technology Evaluation Criteria	14
Formulation of the Assessment	9	References	15

Published in cooperation with Kaiser Foundation Health Plan and Southern California Permanente Medical Group.

TEC Staff Contributors

Author—Barbara M. Rothenberg, Ph.D.; **TEC Executive Director**—Naomi Aronson, Ph.D.; **Managing Scientific Editor**—Kathleen M. Ziegler, Pharm.D.; **Research/Editorial Staff**—Claudia J. Bonnell, B.S.N., M.L.S.; Maxine A. Gere, M.S.

Blue Cross and Blue Shield Association Medical Advisory Panel

Allan M. Korn, M.D., F.A.C.P.—Chairman, *Senior Vice President, Clinical Affairs/Medical Director, Blue Cross and Blue Shield Association*; **Alan M. Garber, M.D., Ph.D.**—Scientific Advisor, *Staff Physician, U.S. Department of Veterans Affairs*; **Henry J. Kaiser, Jr., Professor**; and *Professor of Medicine, Economics, and Health Research and Policy, Stanford University*; **Steven N. Goodman, M.D., M.H.S., Ph.D.**—Scientific Advisor, *Associate Professor, Johns Hopkins School of Medicine, Department of Oncology*; *Division of Biostatistics (joint appointments in Epidemiology, Biostatistics, and Pediatrics)*—American Academy of Pediatrics Appointee. ■ **Panel Members** **Peter C. Albertsen, M.D.**, *Professor, Chief of Urology, and Residency Program Director, University of Connecticut Health Center*; **Edgar Black, M.D.**, *Vice President, Chief Medical Officer, Blue Cross Blue Shield of the Rochester Area*; **Sarah T. Corley, M.D.**, *Physician Consultant, NexGen Healthcare Information Systems, Inc.*—American College of Physicians Appointee; **Helen Darling, M.A.**, *President, National Business Group on Health*; **Josef E. Fischer, M.D., F.A.C.S.**, *Mallinckrodt Professor of Surgery, Harvard Medical School and Chair, Department of Surgery, Beth Israel Deaconess Medical Center*—American College of Surgeons Appointee; **Willard K. Harms, M.D., M.M.**, *Medical Director, Medical Policy and Adjudication, Blue Cross Blue Shield of Illinois*; **I. Craig Henderson, M.D.**, *Adjunct Professor of Medicine, University of California, San Francisco*; **Mark A. Hlatky, M.D.**, *Professor of Health Research and Policy and of Medicine (Cardiovascular Medicine), Stanford University School of Medicine*; **Bernard Lo, M.D.**, *Professor of Medicine and Director, Program in Medical Ethics, University of California, San Francisco*; **Barbara J. McNeil, M.D., Ph.D.**, *Ridley Watts Professor and Head of Health Care Policy, Harvard Medical School, Professor of Radiology, Brigham and Women's Hospital*; **Brent O'Connell, M.D., M.H.S.A.**, *Vice President and Medical Director, Pennsylvania Blue Shield/Highmark, Inc.*; **William R. Phillips, M.D., M.P.H.**, *Clinical Professor of Family Medicine, University of Washington*—American Academy of Family Physicians' Appointee; **Maren T. Scheuner, M.D., M.P.H.**, *Visiting Associate Professor, Department of Health Services, UCLA School of Public Health*—American College of Medical Genetics Appointee; **J. Sanford Schwartz, M.D.**, *Professor of Medicine, Department of Medicine, University of Pennsylvania School of Medicine and Professor, Health Care Systems, Health Management & Economics, The Wharton School*; **Earl P. Steinberg, M.D., M.P.P.**, *President, Resolution Health, Inc.*; **A. Eugene Washington, M.D., M.Sc.**, *Executive Vice Chancellor, University of California, San Francisco*; **Jed Weissberg, M.D.**, *Associate Executive Director for Quality and Performance Improvement, The Permanente Federation.*

CONFIDENTIAL: This document contains proprietary information that is intended solely for Blue Cross and Blue Shield Plans and other subscribers to the TEC Program. The contents of this document are not to be provided in any manner to any other parties without the express written consent of the Blue Cross and Blue Shield Association.

Assessment Objective

Computer-aided detection (CAD) with full-field digital mammography (FFDM) is proposed as an adjunct to radiologists' reading of digital mammograms. In the fall of 2005, the results of the American College of Radiology Imaging Network's (ACRIN) Digital Mammographic Imaging Screening Trial (DMIST) were released. The trial showed with reasonable certainty that there was no difference in the accuracy of FFDM and screen-film mammography (SFM) for asymptomatic women in general. For three subgroups of women, however, FFDM performed better than SFM: women under age 50; pre- or perimenopausal women; and women with heterogeneously dense or extremely dense breasts.

An Assessment on CAD in SFM (Vol. 17, No. 17; December 2002) concluded that CAD improves the accuracy of SFM. Specifically, it improves net health outcome compared with single-reader radiologist interpretation by increasing the true-positive rate without a disproportionate increase in the false-positive rate. The Assessment also concluded that there was insufficient evidence on the use of CAD with FFDM because of the lack of conclusive evidence at the time on the benefits of FFDM in general, as well as the fact that the use of CAD with FFDM was sufficiently different from its use with SFM that evidence on the latter cannot be extrapolated to the former. Now that the performance of FFDM itself has been demonstrated, the question is whether the use of CAD can improve the sensitivity and specificity of digital mammography, which continues to be lower than optimal.

The outcomes of primary interest in this Assessment are intermediate outcomes, including the effect of adding CAD on the true-positive rate (related measures include cancer detection rate or sensitivity) and false-positive rate (related measures include recall rate, biopsy rate, or specificity). Earlier detection of cancer through mammography screening is thought to relate to improvements in health outcomes such as mortality, though the available evidence is not definitive. This Assessment reviews the scientific evidence to determine whether the use of CAD as an adjunct to digital mammography improves these intermediate outcomes. It is assumed that improvement in these intermediate outcomes will be associated with improvements in health outcomes, as well.

Background

Summary of Previous Assessment

The use of computer-aided detection (CAD) with mammography was examined in a previous TEC Assessment, "Computer-Aided Detection (CAD) in Mammography" (Vol. 17, No. 17, December 2002). This Assessment looked at the use of commercial CAD systems where the radiologist interpreted the mammography first without CAD, made recommendations for follow-up, then looked at a second image with the CAD marks of possible lesions, and identified additional cases that might warrant biopsy or short-term follow-up. Because CAD is not 100% sensitive, possible lesions identified for follow-up before the review of the CAD-marked image were automatically sent for follow-up, even if they were not marked by the CAD system. As a result, CAD could increase the sensitivity but not the specificity of mammography. There are studies exploring the use of CAD for diagnosis, i.e., to determine whether a possible lesion is likely to be malignant or benign (e.g., Buchbinder et al. 2004), but they were not addressed in the previous Assessment, nor will they be addressed here.

The 2002 Assessment concluded that adding CAD to a single reading of screen-film mammography (SFM) improves the net health outcome by increasing the true positive rate of SFM without a disproportionate increase in the false positive rate. However, the Assessment also determined that there was insufficient evidence on the impact of CAD as an adjunct to digital mammography. The latter was based on (a) the lack of published clinical effectiveness studies using a commercially available CAD system applied directly to raw data derived from full-field digital mammography (FFDM) and (b) the differences between screen-film and full-field digital mammography that necessitate separate studies on the use of CAD with FFDM (especially the much greater amount of data collected with FFDM and the possible interaction between the CAD software and the digital mammography data being displayed). A recent TEC Assessment (Vol. 20, No. 16) written following the release of the ACRIN DMIST trial results found that digital mammography meets TEC criteria. The current Assessment will review any new evidence on the use of CAD with digital mammography to determine whether or not an equivalent or more beneficial effect on health outcomes can now be demonstrated.

Uses of Mammography

Screening mammography has improved detection of primary breast cancer. Mammographic findings are frequently reported according to the Breast Imaging Reporting and Data System (BIRADS) established by the American College of Radiology. BIRADS Category 1 is defined as a normal mammogram; Category 2, a benign finding; Category 3, a probably benign finding with short-term follow-up possibly recommended; Category 4, a suspicious biopsy for which biopsy should be considered; Category 5, highly suggestive of malignancy for which appropriate action should be taken; and Category 0 indicates the need for additional imaging information (Lieberman et al. 1998; D’Orsi et al. 1998). A meta-analysis prepared for the U.S. Preventive Services Task Force found that the relative risk (RR) of death from breast cancer following screening mammography was 0.84 overall (95% CI: 0.77–0.91), with a RR of 0.85 (95% CI: 0.73–0.99) for women 40 to 49 years old after 14 years of observation and 0.78 (95% CI: 0.70–0.87) for women 50 and older after 14 years of screening (Humphrey et al. 2002). A recent study by the Cancer Intervention and Surveillance Modeling Network (CISNET) used seven statistical models to assess the contribution of screening mammography to the reduction in breast cancer mortality between 1975 and 2000 (Berry et al. 2005). The researchers estimated that screening in the U.S. reduced the rate of death from breast cancer by 7% to 23%, with a median of 15%.

Patients with findings suspicious for breast cancer based on initial SFM, breast physical examination, or symptoms are referred for additional work-up and comprise the diagnostic mammography population. Digital mammography is also used in these patients.

Full-Field Digital Mammography

Description. Digital mammography uses solid state digital detectors that produce electrical signals when exposed to an X-ray source (Pisano 2000). Digital detectors offer improved detection because of improved efficiency of absorption of the incident X-ray photons, a linear response over a wide range of incident radiation intensities, and low system noise (Pisano et al. 1998; Feig and Yaffe 1996). Lesion conspicuity can be affected by contrast manipulations. Image processing has been shown to improve visualization of details within digital mammograms (Pisano et al. 2000). Since the steps of image acquisition and display are

separated, each can be optimized. Image storage, transmission and retrieval can also be improved. Computer-aided diagnosis (CAD) software can be utilized directly, without having to digitize the image first, as is the case with film mammography.

Performance. The largest and best study on the performance of FFDM in a screening population was the American College of Radiology Imaging Network’s (ACRIN) Digital Mammography Imaging Screening Trial (DMIST) (Pisano et al. 2005a,b). It compared the diagnostic accuracy of digital and film mammography for screening purposes by performing both tests in about 42,760 women. In looking at the entire DMIST study population, there was no statistically significant difference in the accuracy of FFDM vs. SFM, as measured by the area under the receiver operating characteristic (ROC) curve, sensitivity, and specificity. The sensitivity and specificity results are based on the BIRADS scoring and a 365-day follow-up period, while the ROC calculations were based on a seven-point malignancy scale and a 455-day follow-up period. Table 1 shows the test results for digital vs. film mammography. The number of false-positive results remains high for both tests. For the 177 cancers detected with digital mammography, 42,570 women were screened and 3,478 health women underwent additional testing because of false positive results. The comparable numbers for SFM were 167 cancers detected; 42,745 women screened; and 3,512 women undergoing additional testing unnecessarily.

DMIST planned a priori to examine the differences in test performance between FFDM and SFM for a number of subgroups. The results revealed a difference in diagnostic performance between FFDM and SFM for three subgroups of women: women under 50 years old; women who are pre- or perimenopausal; and women who have heterogeneously dense or extremely dense breasts (using a BIRADS scoring system). In each case, the area under the ROC curve was significantly larger for digital than for film mammography. The sensitivity was also higher for digital mammography, while the specificity was not significantly different. There does not appear to be any evidence that FFDM tends to detect the less-invasive or less-advanced cancers.

The sensitivity for detecting invasive cancers among all women is virtually the same for

Table 1. Findings of DMIST Trial on Test Accuracy (Based on Pisano et al. 2005b)

Measure	Full-Field Digital Mammography (95% confidence interval*)	Screen-Film Mammography (95% confidence interval*)	Statistical Significance (p value)**
Test Characteristics Using BIRADS Scores and 365-Day Follow-up			
All women			
– Sensitivity for all cancers	0.70 (0.64–0.76)	0.66 (0.60–0.72)	0.37
– Specificity	0.92 (0.918–0.922)	0.92 (0.918–0.922)	0.74
Women under 50 years old			
– Sensitivity	0.78 (0.68–0.88)	0.51 (0.37–0.65)	0.002
– Specificity	0.90 (0.89–0.91)	0.90 (0.89–0.91)	0.89
Pre- or perimenopausal women			
– Sensitivity	0.72 (0.62–0.82)	0.51 (0.39–0.63)	0.008
– Specificity	0.90 (0.896–0.904)	0.90 (0.896–0.904)	0.37
Women with heterogeneously or extremely dense breasts			
– Sensitivity	0.70 (0.62–0.78)	0.55 (0.47–0.63)	0.02
– Specificity	0.91 (0.906–0.914)	0.90 (0.896–0.904)	0.09

* The confidence intervals (CIs) were calculated using the formula: 95% CI = metric ± 1.96 × standard error.
**Results that are statistically significant at the p≤0.05 level are highlighted.

FFDM and SFM. It also is clear that neither technology detects all of the cancers. There are some women whose cancer would be detected using digital mammography but not film mammography and vice versa. DMIST did not look at the impact of screening mammography on breast cancer mortality with either digital or film mammography directly. The researchers assumed that screening mammography reduces breast cancer death rates, so that showing that digital mammography detects cancers as well or better than film mammography was sufficient.

Although DMIST was performed in a screening population, its findings appear to be applicable to diagnostic populations as well. A comparison of cancers detected in a screening population and a diagnostic population found that the latter tend to be larger both for invasive carcinoma and ductal carcinoma in situ (DCIS) and at a more advanced stage (Dee and Sickles 2001). An examination of the types of cancers detected by each type of mammogram in DMIST shows that FFDM was at least as likely as SFM to reveal invasive carcinomas and higher-grade DCIS. In addition, women undergoing diagnostic mammography tend to be younger than the screening population (Dee and Sickles 2001). Women under 50 are one of the subgroups where FFDM performs better than SFM in a screening population. There is

no evidence in the DMIST results to suggest that the difference between FFDM and SFM would be smaller or reversed in a diagnostic population. It is therefore reasonable to infer that digital mammography is at least as accurate as film mammography in a diagnostic population. There is also a small supporting study that compares digital and film mammography among 100 consecutive subjects with radiological abnormalities on an initial screening film mammography (Bonardi et al. 2005).

Computer-Aided Detection with Screen Film Mammography

Description. Commercially available CAD systems use computerized algorithms for identifying suspicious regions of interest. The intent of CAD is to aid in detection of potential abnormalities for the radiologist to re-review. The radiologist, not CAD, makes the diagnosis of whether a clinically significant abnormality exists and whether further diagnostic evaluation is warranted. CAD is proposed as an adjunct to mammography to decrease errors in perception (i.e., failure to see an abnormality).

Summary of Previous Assessment. In the prior Assessment on the use of CAD with mammography, it was concluded that the available evidence suggests that the use of CAD with film-screen mammography improves the net

health outcome compared with interpretation of the films by a single reader. Adding CAD increases the true-positive rate (related measures include cancer detection rate or sensitivity) without a disproportionate increase in the false-positive rate (related measures include recall rate, biopsy rate, or specificity). This conclusion was based in part on a study in which 12,860 eligible screening mammograms were prospectively interpreted by two experienced community practice mammographers (Freer and Ulisse 2001). Mammograms were analyzed by the ImageChecker M1000® v. 2.0 CAD system. Radiologists alone detected 41 cancers for a detection rate of 0.32% (41 of 12,860). The use of CAD resulted in detection of an additional 8 cancers, which were all stage 0 or 1 tumors, including 6 ductal carcinoma in situ lesions, and 2 invasive ductal cell carcinomas. Radiologists without CAD detected 96% of malignancies associated with a mass, but only 68% of malignancies associated with microcalcifications. CAD alone identified all malignancies associated with microcalcifications, but identified only 67% of malignancies associated with mass lesions. The overall detection rate of radiologists using CAD information was 0.36%, representing a relative increase of 19.5% (95% CI: 9.0–42.2%).

The CAD system placed a total of 14,214 marks on 5,204 cases, with an average of 2.8 marks placed per 4-view exam. The vast majority of CAD marks (97.4%) were dismissed by the radiologist, but an additional 156 cases were recommended for recall (for an 18.8% relative increase from 6.5% to 7.7% of subjects). The number of women recommended to have biopsy based on the radiologist's interpretation alone was 107; an additional 21 women were recommended to have a biopsy based on information provided by CAD. The positive biopsy yield was similar in both groups: 38% (41 of 107) and 38% (8 of 21), respectively.

Newer Studies. Around 50 articles on various aspects of the use of CAD with film mammography have been published since the last Assessment of CAD with mammography. One of the largest was a retrospective observational study of readings by 24 breast imaging specialists in an academic medical center practice who interpreted 115,571 screening mammograms, about half with and half without the aid of a CAD system (Gur et al. 2004). Most of the cases

without CAD were older than those with CAD, because a CAD system was added at the main clinical facility part way through the period covered in the study. The researchers did not detect any significant differences between readings with and without CAD in breast cancer detection rates (3.5%) or recall rates (11.4%). A variety of explanations have been offered for these findings, including the study design (Taylor and Given-Wilson 2005); changes in the composition of the patient population over time, which were not controlled for (Astley 2005; Roehrig 2005); and the fact that CAD systems may be more helpful for less-experienced readers. While the use of CAD with screen-film mammography is not the focus of this Assessment, it does suggest that the larger estimates of the impact of CAD may not be borne out in actual practice. Given the discrepancy between the Freer and Ulisse and Gur et al. findings, one researcher in the field has called for more prospective studies to “validate the clinical usefulness of CAD” (Giger 2004).

Computer-Aided Detection with Full-Field Digital Mammography

Description. Conceptually, the CAD systems used with digital mammography are very similar to those used with film mammography. The computer analyzes the digital images collected directly by the FFDM system, applies a set of algorithms that capture characteristics known to be associated potentially with malignancies, and produces an image with markings that show the site of suspicious findings. Sometimes, different marks are used for suspected masses and suspected microcalcifications. The major difference between CAD for FFDM and CAD for SFM is the extensive data set provided by the former and its interaction with the CAD algorithms.

FDA Status. The U.S. Food and Drug Administration (FDA) has given premarket application (PMA) approval to several computer-aided detection (CAD) systems for use with digital mammography. All of these are modifications to CAD systems that were originally approved for use with film mammography. The systems approved for use with digital mammography include the following: the ImageChecker® M1000-DM Computer Aided Detection (CAD) system manufactured by R2 Technology, Inc. (Sunnyvale, CA), which was approved for use with digital mammography

on November 13, 2001 (P970058); and Second Look™ Digital Computer-Aided Detection System, which is manufactured by ICAD, Inc. (Nashua, NH), and was approved for use in digital mammography on October 31, 2005 (P010038; also called at various times Mammoreader, ICAD System, and Securlook).

Methods

Search Methods

The literature search from the previous Assessment, which went through November 2002, was rerun through March 2006, using the following search terms: “(mammography(mh) OR mammo*) AND (radiographic image interpretation, computer-assisted(mh) OR diagnosis, computer-assisted(mh) OR computer-assisted OR computer-aided OR CAD),” limited to English language. The search yielded 443 references, which were searched for studies on the impact on sensitivity, specificity, and/or biopsy rates of the use of CAD with digital mammography. While there is now an extensive literature on the use of CAD with film mammography, there are very few studies on the use of CAD with digital mammography.

Study Selection

Because of the paucity of studies on the use of CAD with FFDM, all studies were reviewed that looked at the use of CAD with digital mammography, except for those dealing solely with the development of new CAD methodologies and systems. Only 2 articles and 3 abstracts reported on the performance of commercially available CAD systems; three more discussed the development of new CAD systems and provided limited performance data.

Medical Advisory Panel Review

This Assessment was reviewed by the Blue Cross and Blue Shield Association Medical Advisory Panel (MAP) on February 23, 2006. To maintain the timeliness of the scientific information in this Assessment, literature search updates were performed subsequent to the Panel’s review (see “Search Methods”). If the search updates identified any additional studies that met the criteria for detailed review, the results of these studies were included in the text where appropriate. There were no studies that would change the conclusions of this Assessment.

Formulation of the Assessment

Patient Indications

Two groups of patients undergo FFDM. The first group is comprised of individuals who present for screening mammography, based on age or other risk factors. The second group is patients referred for diagnostic mammography on the basis of an abnormal breast exam, including a palpable mass; breast symptoms such as spontaneous nipple discharge or unilateral focal pain; initial screening mammogram with a BIRADS classification of 0 or 3; and a history of breast cancer.

Technologies to Be Compared

Full-Field Digital Mammography with a Single Reading by a Radiologist.

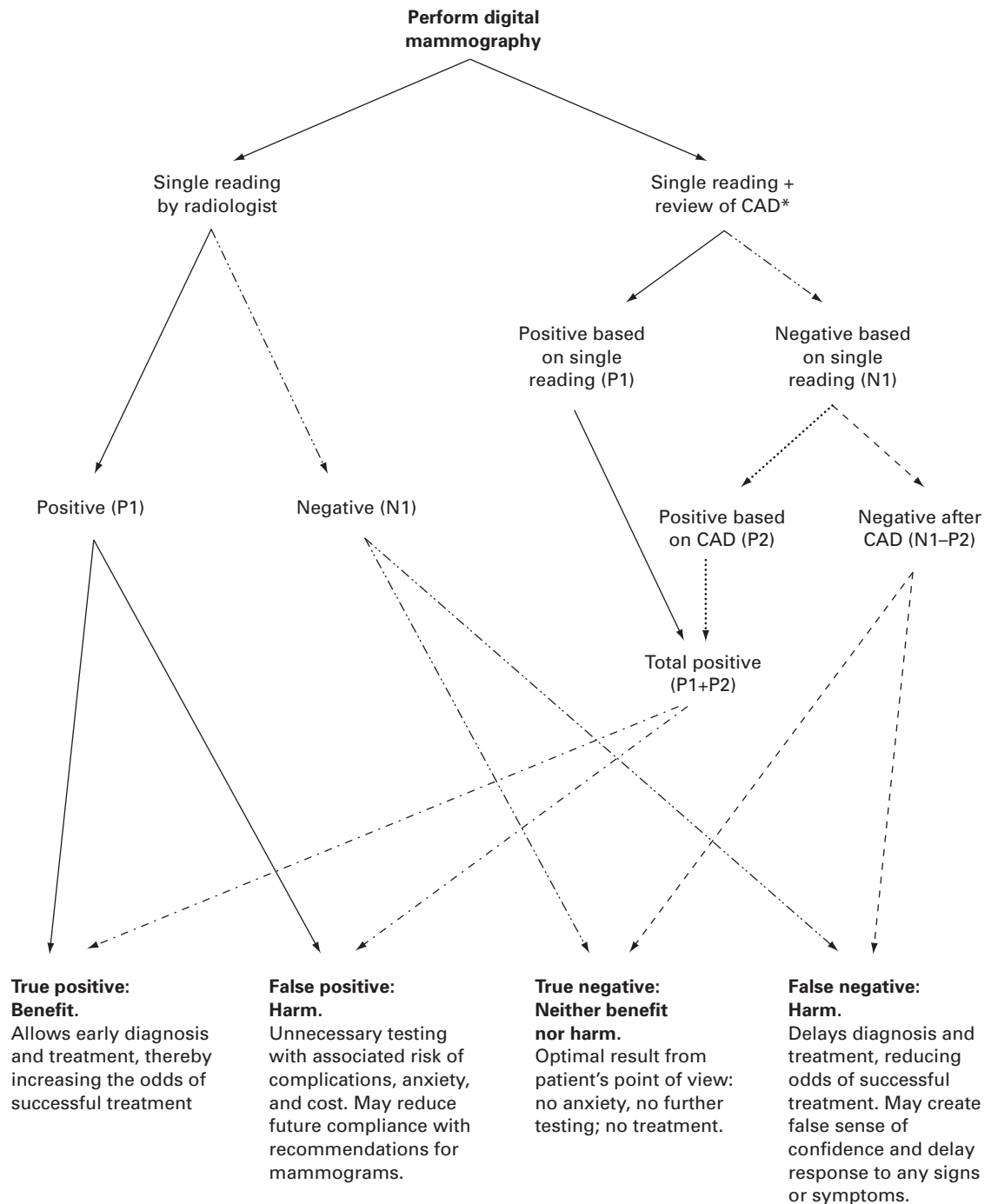
Full-Field Digital Mammography with a Single Reading Followed by the Radiologist’s Review of the Results of the Computer-Aided Detection System. The CAD system is used after a radiologist views the original images, and it can only be used to identify additional lesions that may require further work-up. In accordance with the stipulations of the U.S. Food and Drug Administration when the CAD systems received premarket approval, CAD will not be used to identify false positive results from the initial reading. Therefore, it may improve the sensitivity of FFDM, but not the specificity.

Health Outcomes

Benefits. Despite the improved performance in some populations (e.g., younger women and those with denser breasts), the sensitivity of full-field mammography is not optimal. The goal of CAD systems as currently used is to increase the sensitivity of FFDM by identifying lesions that the radiologists may have overlooked during the initial reading. It is also known that there is substantial variation in the accuracy of readings across radiologists. CAD systems are potentially helpful in bringing the accuracy of less skilled readers up to those of expert readers. If CAD does increase the sensitivity of FFDM, it will result in more true positive results, with earlier treatment and presumably better outcomes (see Figure 1). It will also reduce the false negative cases, which reduce the odds of successful treatments.

Harms. If the use of CAD increases the number of recalls and biopsies without an appreciable

Figure 1. Alternative Trajectories of Management, with and without CAD



*CAD = Computer-Aided Detection

increase in the number of cancers detected, then it may increase the burden of screening for women. There has been a concern that compliance with mammography screening may decline after a person has a false-positive result, although this has not always been borne out (Chiarelli et al. 2003; Currence et al. 2003).

The alternative trajectories are shown in Figure 1. As can be seen in the figure, adding CAD to the reading process is likely to increase the number of positive results (by adding P2 to P1), while decreasing the negative results (by subtracting P2 from N1, the initial number of negative results). The overall impact of adding CAD on the accuracy of digital mammography will depend on the number of true and false positive results within P2.

Specific Assessment Question

- For individuals undergoing either a screening or a diagnostic digital mammogram, how does adding a review of CAD results after the initial reading by the radiologist affect the number of true-positive and false-positive results, and the number of unnecessary additional tests or biopsies?

Review of Evidence

While there is a growing literature on the use of CAD with screen-film mammography, there are very few studies on the use of CAD with full-field digital mammography. This paucity of studies is confirmed by a recent review article (Pisano and Yaffe 2005), which notes that “there has been very little published in the peer-reviewed literature on the effect of CAD on the performance of radiologists interpreting digital mammograms.”

Only two published articles were found that assess the performance of a commercially available CAD system as an adjunct to digital mammography (Baum et al. 2002, Obenauer et al. 2006). The first is a retrospective study that examines the ability of CAD to detect histologically proven carcinomas on FFDM (GE’s Senographe 2000D) using R2’s ImageChecker V2.3 system. The sample consisted of 61 patients with 63 histologically verified carcinomas (including one multicentric case and one bilateral case), with a total of 187 images (two views were available on some breasts but not others, and mammograms of the contralateral breast were available for some patients).

Sensitivity and specificity were calculated by looking at whether the CAD system marked the histologically proven lesions. The sensitivity was 89% for all microcalcifications (95% CI: 70–98%), 100% for microcalcifications alone (without a suspicious mass), and 81% for masses alone (95% CI: 67–91%). There were 0.35 false-positive microcalcification marks per image and 0.26 false-positive marks per image for masses without associated microcalcifications. There are a number of weaknesses in this study, including the selected sample of cancer cases (whereas, there are very few cancers in a screening or even in a diagnostic population) and the fact that the results of the radiologists’ readings prior to the use of CAD were not reported. Therefore, the incremental value of CAD—of primary interest in this Assessment—cannot be assessed.

The second study (Obenauer et al. 2006), which was conducted by some of the same researchers, also was retrospective and looked only at cases with histopathologically proven cancers. It used the same systems: the Image Checker V2.3 CAD and General Electric’s Senographe 2000D system. There were 226 cancers viewed on a total of 412 images (there does not appear to be an overlap between these cases and those reported by Baum et al. [2002], although this was not confirmed). CAD marks placed over a histologically proven carcinoma were considered a true positive, while those placed in other areas were treated as false positives. The results are reported separately for cranio-caudal and mediolateral oblique views, which makes it difficult to assess the impact of using CAD at the patient level. Although the methods are not clearly described, it appears that these results were then compared to the BIRADS classification assigned by the radiologists. The number of true positive lesions marked by the CAD system increased with the BIRADS score, but the CAD system correctly identified some cancers even in BIRADS 1–3 cases, i.e., those that the radiologist had indicated had no lesions or benign ones. The CAD system “missed” fewer carcinomas at the higher BIRADS levels, and there were 0.5 false positive markers per image. This study is subject to the same weaknesses as the previous study by Baum et al.: retrospective design and a patient population consisting of all cancer cases. The authors point out that they cannot report the specificity and positive and negative predictive values because there are no true negatives included in the study. They also do not report

the sensitivity, which would not be very meaningful given the way in which true positives are defined. This study does not show the impact of CAD as it is meant to be used in a clinical setting, i.e., as a supplement to an initial reading by a radiologist.

Another team of researchers presented two abstracts on the use of the Image Checker (version 8) at the 2005 meeting of the Radiological Society of North America. One dealt with the possibility of using CAD to eliminate some cases with a high probability of being normal in order to reduce radiologists' workloads; it will not be discussed further here (Yaffe et al. 2005). The second abstract reported the sensitivity and specificity of CAD with FFDM (Jong et al. 2005). It included 220 mammograms (70 of which imaged biopsy-proven cancer) read by 10 radiologists (5 experienced and 5 trainee radiologists). The normal cases and those with benign lesions were followed for at least 1 year. The abstract does not describe how the cases were amassed, i.e., whether they were consecutive cases or whether the researchers deliberately selected a certain number of malignant and benign cases. The abstract also does not state whether the study was retrospective or prospective. The answers to these questions could have a major impact on the validity and generalizability of the results. The reported results are summarized in Table 2.

The reported sensitivity of CAD alone was 94.3%, while the reported specificity was only 39.3%. While these results are suggestive, insufficient detail is provided in the abstract to assess many aspects of the study, e.g., whether changes in sensitivity and specificity were statistically significant or whether the impact of CAD represents movement along the same ROC curve (i.e., simply a different operating point without any fundamental change in the tradeoff between sensitivity and specificity) or

movement to a higher and presumably better ROC curve where, for example, an increase in sensitivity can be achievable without as large a decrease in specificity as would occur without the use of CAD.

An abstract on the performance of ImageChecker was also presented at the 2001 RSNA meeting; it was written by employees of General Electric Medical Systems or R2 Technology, Inc., the maker of the CAD system (O'Shaughnessy et al. 2001). They took 90 biopsy-proven cancers and 71 normal cases (with 1-year follow-up) and applied the CAD system retrospectively. All of the cases had been imaged with FFDM; 82 of the cancers were among a diagnostic population, 8 among a screening population. The authors considered the CAD result positive if the location of the cancer was marked by the system. They estimated the sensitivity to be 89% overall, with 97% for microcalcifications and 84% for masses. There were 0.55 false marks per image among the normal cases. The authors also compared these results to the use of the same CAD system on film mammography and reported a similar performance. There are a number of limitations to this study from the point of view of this Assessment, in addition to the fact that abstracts provide limited detail on the study design. They include the use of a highly selected sample and the assumption that any lesion marked by CAD would have been picked up by the radiologist. A comparison of the radiologists' results before and after viewing the CAD results is needed to gauge the incremental impact of CAD on the accuracy of FFDM.

There are several articles that report on the development of CAD systems and contain data on their performance. Wei et al. (2005) report on the development of their own CAD system, which uses the raw FFDM image from any manufacturer. However, the paper appears to

Table 2. Summary of Results from Jong et al. 2005*

	Reported Sensitivity		Reported Specificity	
	Experienced	Trainee	Experienced	Trainee
Without CAD	78.9%	71%	73%	65.6%
With CAD	80.3%	74.3%	72%	62.5%
Difference in percentage points	+1.4	+3.3	-1	-3.1

*These data should be interpreted with caution, because the method of compiling the sample, and therefore its composition, is not well-defined.

report only on the performance of their system relative to that of an expert mammography reader. Li et al. (2002) report on a CAD system to detect masses. The sensitivity of the training data set was reported to be 91% with 3.21 false positive marks per image, while the sensitivity for the small testing data set was quite different at 60%, with a false positive rate of 2.47 marks per image. These retrospective data sets were also highly selected. Diekmann et al. (2004) report on the performance of their CAD system, which aims to improve the detection of microcalcifications. The true diagnosis for each case was based on histological confirmation or at least 1 year of follow-up. They found no difference in the ROC curves of the 4 readers in the study in interpreting FFDM with and without CAD; few specific details that would allow calculation of sensitivity and specificity are reported. Furthermore, their data set consists of 280 digital mammograms, with about the same number from each of the 5 BIRADS categories, so it is not representative of a real screening or diagnostic population. Finally, each reader interpreted images with and without CAD, but these were two separate steps. In other words, the reader did not view the image alone, record the findings, and then look at the CAD results for each mammography. Given the limitations of each of these studies, they will not be considered further in this Assessment.

Wei et al. (2005) note that “Several commercial CAD systems already obtained FDA approval for use with FFDMs. The commercial CAD systems generally reported similar performance on FFDMs and SFMs. However, their study [Li et al. 2002] was not reported in peer-reviewed journals so that the data set and algorithm are unknown.”

Conclusion

Now that the performance of full-field digital mammography has been established and the value of CAD as an adjunct to screen film mammography has been reviewed, the question is whether the addition of CAD to a single reading of digital mammograms will improve the sensitivity of FFDM. Studies are needed that compare the performance of FFDM alone with that of FFDM and CAD. Ideally, these studies would be prospective, would include only the population of interest (whether it be screening or diagnostic), would involve multiple centers

and readers, and would use histological analysis of biopsy specimens or extended follow-up (1 year or more) to determine whether the patient actually had breast cancer or not. Unfortunately, no studies were found that addressed this question adequately. Because of the differences between film and digital mammography, it is not sufficient to extrapolate the results of CAD from SFM to FFDM. Studies conducted with a convenience sample or with an enhanced number of cancer cases are not sufficient, because radiologists may apply different thresholds in diagnosing potential abnormalities if they know there is a higher-than-usual prevalence of cancer in the study population. Better quality studies would have radiologists reading mammograms without knowledge of the final diagnosis or knowledge of any special features of the study population.

Retrospective studies showing that cancers missed during screening would have been marked by a CAD system are informative but insufficient. Given the large number of false-positive marks that characterize CAD systems, it is possible that the CAD marks would have been dismissed by the radiologist if the CAD system had been in place at the time (Taylor and Given-Wilson 2005). Furthermore, looking at cancer cases is very different from reading screening cases, where the expected cancer rate is low (Astley 2005). Prospective studies of a screening or diagnostic patient sample that compare FFDM sensitivity and specificity before and after adding the CAD results are needed.

One prospective study on the use of CAD with FFDM is just starting and will be completed in 2008 (“The Diagnostic Efficacy of Computer-Aided Detection (CAD) in Full-Field Digital Mammography (FFDM)—A Prospective Study”; National Taiwan University Hospital; ClinicalTrials.gov Identifier NCT00173303; Jane Wang, MD, Principal Investigator). The planned enrollment is 3,000 women and includes both women undergoing screening and women with breast disease. In their justification for the study, the researchers state the following:

However, there are very few reports regarding the CAD application in FFDM, since FFDM can offer the post-acquisition processing on high-resolution review workstation for interpretation. Nevertheless, the spatial resolution of soft-copy reading on monitors for FFDM is slightly inferior to but the contrast resolution is slightly

superior to that of conventional SFM. Herein, the diagnostic efficacy and role of CAD in FFDM are still unclear. (www.clinicaltrials.gov, retrieved on 2 Feb 2006)

While the results of this study should be interesting and useful, their generalizability will need to be assessed. For example, earlier studies have shown that Asian women are less likely than some other groups to have dense breasts (Chan et al. 2004). Whether CAD has the same impact on FFDM across ethnic and racial groups may need to be evaluated.

Summary of Application of the Technology Evaluation Criteria

Based on the available evidence, the Blue Cross and Blue Shield Association Medical Advisory Panel made the following judgments about whether the use of computer-aided detection (CAD) as an adjunct to single reading of full-field digital mammography meets the Blue Cross and Blue Shield Association's Technology Evaluation Center (TEC) criteria.

1. The technology must have final approval from the appropriate governmental regulatory bodies.

Two CAD systems have received premarket application (PMA) approval by the U.S. Food and Drug Administration (FDA) for use with FFDM.

2. The scientific evidence must permit conclusions concerning the effect of the technology on health outcomes.

While there is a growing literature on the use of CAD with screen film mammography, there are very few articles that look at CAD with FFDM. Because the FDA approved the CAD devices for use with FFDM as a modification of their original approval of the CAD systems

for use with SFM, whatever data were used to support the request for premarket approval of these CAD systems with FFDM are not readily available. The articles on the use of CAD with FFDM are inadequate to determine the incremental impact on sensitivity, specificity, and biopsy rates.

In summary, the available evidence is considered insufficient to permit conclusions on the effect on relevant outcomes of using CAD after initial radiographic interpretation as a quality adjunct to single-reader mammography in patients having full-field digital mammography for screening or diagnostic purposes.

3. The technology must improve the net health outcome; and
4. The technology must be as beneficial as any established alternatives.

The available data from well-conducted studies are insufficient to make a determination of whether adding CAD to FFDM leads to diagnoses that are as or more accurate than reading the FFDM images alone. Given the lack of data on these intermediate outcomes, it also is not possible to determine the impact of CAD on health outcomes such as treatment success, recurrence rates, and survival. As a result, it is unclear whether health outcomes are the same or better with the use of CAD with FFDM compared to single reading of FFDM alone.

5. The improvement must be attainable outside the investigational settings.

Given the inability to determine whether the use of CAD improves health outcomes, considering the generalizability of the impact of CAD is premature.

For the above reasons, the use of CAD systems as an adjunct to single reading of full-field digital mammography images does not meet the TEC criteria.

NOTICE OF PURPOSE: TEC Assessments are scientific opinions, provided solely for informational purposes. TEC Assessments should not be construed to suggest that the Blue Cross Blue Shield Association, Kaiser Permanente Medical Care Program or the TEC Program recommends, advocates, requires, encourages, or discourages any particular treatment, procedure, or service; any particular course of treatment, procedure, or service; or the payment or non-payment of the technology or technologies evaluated.

CONFIDENTIAL: This document contains proprietary information that is intended solely for Blue Cross and Blue Shield Plans and other subscribers to the TEC Program. The contents of this document are not to be provided in any manner to any other parties without the express written consent of the Blue Cross and Blue Shield Association.

References

- Astley SM. (2005).** Evaluation of computer-aided detection (CAD) prompting techniques for mammography. *Br J Radiol*, 78:S20-S25.
- Baum F, Fischer U, Obenauer S et al. (2002).** Computer-aided detection in direct digital full-field mammography: initial results. *Eur Radiol*, 12:5015-7.
- Berry DA, Cronin KA, Plevritis SK et al. (2005).** Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med*, 355:1784-92.
- Bonardi R, Ambrogetti D, Ciatto S. (2005).** Conventional versus digital mammography in the analysis of screen-detected lesions with low positive predictive value. *Eur J Radiol*, 55:258-65.
- Buchbinder SS, Leichter IS, Lederman RB et al. (2004).** Computer-aided classification of BI-RADS category 3 breast lesions. *Radiology*, 250(5):820-5.
- Chan ZS, Wu AH, Gauderman WJ. (2004).** Does mammographic density reflect ethnic differences in breast cancer incidence rates? *Am J Epidemiol*, 159(2):140-7.
- Chiarelli AM, Moravan V, Halapy E et al. (2005).** False positive result and reattendance in the Ontario Breast Screening Program. *J Med Screening*, 10(3):129-35.
- Currence BV, Pisano ED, Earp JA et al. (2005).** Does biopsy, aspiration or six-month follow-up of a false-positive mammogram reduce future screening or have large psychological effects? *Acad Radiol*, 10(11):1257-66.
- Dee KE, Sickles EA. (2001).** Medical audit of diagnostic mammography examinations: comparison with screening outcomes obtained concurrently. *AJR Am J Roentgenol*, 176:729-35.
- Diekmann F, Diekmann S, Bollow M et al. (2004).** Evaluation of a wavelet-based computer assisted detection system for identifying microcalcifications in digital full-field mammography. *Acta Radiol*, 45(2):136-41.
- D'Orsi CJ, Bassett LW, Feig SA et al. (1998).** *Illustrated Breast Imaging Reporting and Data System (Illustrated BI-RADS), Third Edition*, American College of Radiology, Reston, VA.
- Feig SA, Yaffe MJ. (1996).** Current status of digital mammography. *Semin Ultrasound CT MR*, 17(5): 424-45.
- Freer TW, Ulisse MJ. (2001).** Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*, 220(3):781-6.
- Giger ML. (2004).** Computerized analysis of images in the detection and diagnosis of breast cancer. *Semin Ultrasound CT MRI*, 25:411-8.
- Gur D, Sumkin JH, Rockette HE et al. (2004).** Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system. *J Natl Cancer Inst*, 96(3):185-90.
- Humphrey LL, Helfand M, Chan BKS et al. (2002).** Breast cancer screening: A summary of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med*, 137(5 Part1):547-60.
- Jong R, Umranikar S, Wolfman J et al. (2005).** Performance of CAD on a FFDM system. Abstract presented at the 2005 annual meeting of the Radiological Society of North America, Chicago, IL.
- Li L, Clark RA, Thomas JA. (2002).** Computer-aided diagnosis of masses with full-field digital mammography. *Acad Radiol*, 9(1):4-12.
- Liberman L, Abramson AF, Squires FB et al. (1998).** The Breast Imaging Reporting and Data System: positive predictive value of mammographic features and final assessment categories. *AJR Am J Roentgenol*, 171:35-40.
- Obenauer S, Sohns C, Werner C et al. (2006).** Computer-aided detection in full-field digital mammography: detection in dependence of the BI-RADS categories. *Breast J* 12(1):16-19.
- O'Shaughnessy K, Castellino R, Muller S et al. (2001).** Computer-aided detection (CAD) on 90 biopsy-proven breast cancer cases acquired on a full-field digital mammography (FFDM) system. Abstract presented at the annual meeting of the Radiological Society of North America, Chicago, IL.
- Pisano ED. (2000).** Current status of full-field digital mammography. *Radiology*, 214(1):26-8.
- Pisano ED, Yaffe MJ. (2005).** Digital mammography. *Radiology*, 254(2):553-62.
- Pisano ED, Cole EB, Major S et al. (2000).** Radiologists' preferences for digital mammographic display. *Radiology*, 216(3):820-30.
- Pisano ED, Zong S, Hemminger BM et al. (1998).** Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *J Digital Imaging*, 11(4):193-200.
- Pisano ED, Gatsonis CA, Yaffe MJ et al. (2005a).** American College of Radiology Imaging Network Digital Mammographic Imaging Screening Trial: objectives and methodology. *Radiology*, 256(12):404-12.

Pisano ED, Gatsonis C, Hendrick E et al. (2005b). Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med*, 353(17):1775-85.

Roehrig J. (2005). The manufacturer's perspective. *Br J Radiol*, 78:S41-S45.

Taylor P, Given-Wilson RM. (2005). Evaluation of computer-aided detection (CAD) devices. *Br J Radiol*, 78:S26-S50.

Technology Evaluation Center (TEC). (2002). *Computer-Aided Detection (CAD) in Mammography*. Vol. 17, No. 17. Chicago, IL: Blue Cross and Blue Shield Association.

Technology Evaluation Center (TEC). (2006). *Full-Field Digital Mammography*. Vol. 20, No. 16. Chicago, IL: Blue Cross and Blue Shield Association.

Wei J, Sahiner B, Hadjiiski LM, Chan HP et al. (2005). Computer-aided detection of breast masses on full field digital mammograms. *Med Phys*, 32(9):2827-38.

Yaffe M, Jong R, Wolfman J et al. (2005). Performance of CAD for prescreening with digital mammography. Abstract presented at the 2005 annual meeting of the Radiological Society of North America, Chicago, IL.



**Technology
Evaluation
Center**

**Blue Cross and
Blue Shield Association**
225 North Michigan Avenue
Chicago, Illinois 60601-7680
www.bcbs.com/tec