

Computer-Aided Detection (CAD) in Mammography



Assessment
Program
Volume 17, No. 17
December 2002

Executive Summary

The objective of this Assessment is to evaluate the clinical effectiveness of using computer-aided detection (CAD) as an adjunct to mammography. Mammography is used for breast cancer screening and diagnosis to detect and characterize breast masses and calcifications that may be malignant. Conventional mammography uses film-screen technology and achieves approximately 85% sensitivity in detecting cancer though results are operator-dependent and may vary with reader expertise. There is considerable interest in finding techniques to improve sensitivity and reduce variability among readers.

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 if 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). This Assessment will focus exclusively on the use of commercially available computer analysis systems to aid or assist in detection of potential abnormalities.

Currently available CAD systems are intended to be used only after the radiologist has completed an evaluation of the images without CAD prompts and has made an initial decision whether any abnormal areas require recall of the patient for further work-up. If a radiologist identifies an abnormal area of concern on a mammogram during initial reading and that area does not get marked by CAD, the radiologist is still advised to interpret the mammogram as positive and to recall the patient for further work-up. This is because CAD is not 100% sensitive in marking all cancers, particularly larger masses.

Therefore, when used as intended, CAD would be expected to increase the number of mammograms interpreted as positive to the extent that it points out abnormalities previously overlooked by the radiologist on unaided reading. It is possible that a CAD system could be used differently than it is intended to be used. In such a setting, radiologists using CAD before carefully reviewing the films without CAD prompts might be distracted by the CAD prompts and might miss some cancers they would otherwise have detected. In this way, the use of CAD might lead to a lower true-positive rate compared with radiologists reading without CAD.

The outcomes of primary interest in this Assessment are intermediate outcomes including the effect of adding CAD on 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



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.

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 mammography improves these intermediate outcomes.

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-assisted detection (CAD) devices after initial radiographic interpretation as a quality adjunct to single-reader mammography meets the Blue Cross and Blue Shield Association Technology Evaluation Center (TEC) criteria.

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

There are three manufacturers that have received U.S. Food and Drug Administration (FDA) approval to market computer-aided detection (CAD) systems that take film-screen mammograms and digitize the images for computer analysis.

The first device to receive FDA approval, the ImageCheckerM1000® by R2 Technology, Inc. (Los Altos, CA), was approved by premarket application (PMA) approval (P970058) on June 26, 1998. The initial product labeling was for use on routine screening mammograms, but on May 29, 2001, approval was granted for the expansion of the “Indications for Use” to cover diagnostic as well as screening mammograms. The ImageCheckerM1000® is:

“intended to identify and mark regions of interest on routine screening and diagnostic mammograms to bring them to the attention of the radiologist after the initial reading has been completed. Thus, the system assists the radiologist in minimizing observational oversights by identifying areas on the original mammogram that may warrant a second review. (<http://www.fda.gov/cdrh/pma/pmamay01.html>; accessed July 15, 2002)

Multiple supplemental applications and approvals have also been processed relating to modifications of the ImageChecker hardware and software. A “New Efficacy Claim” was approved in PMA supplement 7 (version 2.2 software). The change is from:

“For every 100,000 cancers currently detected by screening mammography, the use of the ImageChecker could result in early detection of an additional 30,500 breast cancers.”

to:

“Use of the ImageChecker could result in earlier detection of up to 25.4% (95% CI, 19.4%–27.4%) of the cancers currently detected with screening mammography in those women who had a prior screening mammogram 9-24 months earlier.” (<http://www.fda.gov/cdrh/pma/pmafeb02.html>; accessed July 15, 2002)

Two additional CAD systems have also been FDA approved. Second Look™ by CADx Medical Systems, Inc. (Northborough, MA) was approved by PMA (P010034) on January 31, 2002, and the MammoReader™ by Intelligent Systems Software, Inc. (Clearwater, FL) was approved by PMA (P010038) on January 15, 2002.

Each of these devices is:

“intended to identify and mark regions of interest on standard mammographic views to bring them to the attention of the radiologist after the initial reading has been completed. Thus, the system assists the radiologist in minimizing observational oversights by identifying areas on the original mammogram that may warrant a second review.” (<http://www.fda.gov/cdrh/pma/pmajan02.html>; accessed July 15, 2002)

In addition, General Electric Medical Systems (Milwaukee, WI) received FDA approval on April 12, 2002 to modify their Senographe 2000D® Full Field Digital Mammography System (P990066/S007) to permit integration and use of the ImageChecker® M1000-DM Computer Aided Detection (CAD) system manufactured by R2 Technology, Inc. (<http://www.fda.gov/cdrh/pma/pmaapr02.html>; accessed August 21, 2002)

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

Indication #1: CAD in Patients Having Film-Screen Mammography. The outcomes of primary interest in this Assessment are intermediate outcomes related to the diagnostic results of mammography, true-positive rate (related measures include cancer detection rate or sensitivity) and false-positive rate (related measures include recall rate, biopsy rate, or specificity). While recall rate and biopsy rate are not actually measures of the false-positive rate per se, an increase in recall rate or biopsy rate represents a combination of the increase in false-positives and any increase in true-positives. In the screening setting, the proportion of false-positives reflected in the recall or biopsy rate is generally higher than the proportion of true-positives based on the very low prevalence of cancer. Thus, the terms “true-positive rate” and “false-positive rate” will be used to establish a conceptual framework within which the various indicators or measures of true-positives and false-positives reported in the studies can be synthesized and analyzed. Similar to diagnostic sensitivity and specificity, true-positive rate and false-positive rate are interrelated outcome measures, and it is important to consider both measures in evaluating the results of an individual study and in comparing results across studies. A better quality study would provide a complete assessment of both of these related and relevant outcomes.

This Assessment looked for studies that included study populations that were relevant to the proposed clinical use, and better quality studies would be those that included unselected consecutive cases in the routine screening or diagnostic setting. Studies that evaluate only a selected subset of cases, such as a subset of cancer cases missed during prospective screening, provide only indirect evidence of the potential effect of CAD in an unselected setting.

With regard to radiologist interpretation setting, radiologists may apply different diagnostic 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.

The systematic review of the literature yielded 11 studies that used computer-aided detection in distinct patient cohorts and met eligibility criteria. All of these studies used CAD in film-screen mammography. Seven of these studies (8 reports) used the ImageChecker® system (n=12,860; n=104; n=280; n=106; n=286; n=110; one study: n=14,817 CAD, n=23,682 no CAD; n=120). Two eligible studies were identified for the Second Look™ system and both are derived from the FDA PMA data (n=3,946; n=177). Similarly, two FDA PMA data studies are available for the MammoReader™ (n=300; n=327).

The best available evidence to determine the effect of CAD on true-positive rate (e.g., cancer detection rate or sensitivity) and false-positive rate (e.g., as measured by recall or biopsy rate or specificity) in the screening mammography setting is provided by one study, and all other studies were rated of lower quality. The highest quality study employs a well-designed prospective protocol using CAD in a representative manner in the actual clinical screening setting for which it is intended. The results of this study suggest that adding CAD provides a clinically significant incremental improvement in cancer detection rate (19.5%) with an increase in recall and biopsy rate that is similar in proportion to the number of recalls and biopsies per cancer detected by radiologists reading alone without CAD.

Three other studies comparing recall rates without and with CAD provide additional substantiation that radiologists are able to discern the many false-positive marks prompted by CAD from those that require further work-up and that the use of CAD is unlikely to cause a very large increase in recall rate.

An additional 4 studies (5 reports) conducted retrospectively in samples of cancers that had been missed on mammography provide some supportive evidence of the potential effectiveness of CAD. By virtue of their design, these studies do not provide direct evidence of the effect that CAD would

have on cancer detection rates; however, they do provide indirect evidence of the potential effect that CAD might have had in improving earlier detection of these missed cancers had CAD been available. Furthermore, this population of cancers missed on mammography is precisely the type of cases where one would hope to have an impact by use of CAD. These studies estimate that the use of CAD might lead to detecting 23-45% of cancers that are missed at mammography.

Three other retrospective studies using selected samples of abnormal and normal mammograms compare sensitivity and specificity of radiologists reading alone versus reading with CAD. However, these studies suffer from methodological weaknesses including potential spectrum bias in case selection and lack of representative radiologist interpretation environment due to the enriched case-mix and the retrospective study setting. While the results of these studies do not suggest that the addition of CAD results in a significant increase in sensitivity, it seems reasonable that the weaknesses in these retrospective studies may account for the apparent differences in the observed effect of adding CAD when results are compared with the higher-quality study.

In summary, the available evidence is considered sufficient 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 film-screen mammography.

Indication #2: CAD in Patients Having Full-Field Digital Mammography. Full-field digital mammography (FFDM) is being studied as a potential alternative to film-screen mammography, and regulatory approval has been granted for at least one CAD system and one digital mammography system to be integrated. No published clinical effectiveness studies were identified using a commercially available CAD system applied directly to raw data derived from digital mammography.

A recent TEC Assessment (Vol. 17, No. 7; 2002) reviewed the available evidence on the clinical effectiveness of full-field digital mammography. This analysis found insufficient evidence to determine that full-field digital mammography is at least as good as film-screen mammography. Furthermore, while not statistically significant, some of the available studies using FFDM in the screening setting showed concerning trends toward lower sensitivity for FFDM in detecting cancer as compared to film-screen mammography.

Given the uncertainty as to the diagnostic performance of FFDM relative to film-screen mammography, it is important to examine empirical evidence of the clinical effectiveness of CAD as applied in full-field digital mammography. Indeed, there are conceptual similarities between the application of CAD to a digitized film-screen mammogram and to a directly acquired digital mammogram, and it might seem reasonable to generalize some of the evidence using CAD in film-screen mammography to digital mammography. However, there are some important differences that will need to be evaluated with clinical studies to determine the clinical effectiveness of applying CAD directly to digital mammography. Unlike with digitized film-screen mammograms, there is more data in a raw digital mammogram than can be displayed 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.

In summary, the available evidence is considered to be 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.

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

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

Indication #1: CAD in Patients Having Film-Screen Mammography. CAD is intended as an adjunct to the radiologist's interpretation of a mammogram. Thus, the primary relevant comparison for this Assessment is whether the benefit of adding CAD (i.e., increase in true-positive rate) outweighs the harm of adding CAD (i.e., increase in false-positive rate). In some settings, CAD

may be considered as an alternative to using a second independent radiologist interpretation (i.e., double-reading), however, this is not a common practice in the U.S.

In the highest quality study, 12,860 eligible screening mammograms were prospectively interpreted by 2 experienced community practice mammographers. Mammograms were analyzed by the ImageChecker M1000® v. 2.0 CAD system. Before reviewing the CAD prompts, the radiologist recorded an initial interpretation and a decision whether or not to recall the patient for additional evaluation. Then the CAD analysis was provided and the radiologist re-evaluated any areas prompted by the CAD system and recorded a second interpretation and decision. Clinical management was guided by the final reading including cases recommended for further evaluation based on CAD findings. The protocol did not permit the radiologist to reverse an initial decision to recall the patient if CAD did not place a mark in the area of interest. Thus, the use of CAD in this study was purely as an aid to detection and CAD could only increase the cancer detection and recall rate.

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. This included an average of 1.2 microcalcification marks and 1.6 mass or architectural distortion marks. The vast majority of CAD marks (97.4%) were dismissed by the radiologist but an additional 156 cases were recommended for recall. Thus, the end result of using CAD information on recall rate was an 18.8% relative increase from 6.5% to 7.7% of subjects.

The number of women actually recommended to have biopsy as a result of the radiologist interpretation alone was 107 and an additional 21 women were recommended to have a biopsy based on additional information provided by CAD. The positive biopsy yield was similar in both groups: 38% (41 of 107) and 38% (8 of 21), respectively.

In comparing the trade-off between increased recall or biopsy rate and increased detection, an additional 156 women were recalled using CAD and an additional 21 women were biopsied in order to detect an additional 8 cancers, yielding 19.5 recalls and 2.6 biopsies for each additional case of cancer detected. It may be useful to compare these benchmarks to the results based on radiologists reading without CAD. In that setting, 830 recalls and 107 biopsies were recommended to detect 41 cancers, yielding 20.2 recalls and 2.6 biopsies per case of cancer detected.

Because this study did not employ an independent reference standard to ascertain the presence of any false-negative cases, sensitivity and specificity are not reported. However, if one makes an assumption about the level of sensitivity at which these study radiologists were reading screening mammograms without CAD, then sensitivity and specificity rates can be calculated from the data.

If one assumes that radiologists reading with CAD detected all cancers in the screening population (100% sensitivity with CAD), then radiologists reading alone in this study had a sensitivity of 84%, corresponding to a 16% absolute increase in sensitivity with the use of CAD. If one assumes instead that radiologists were reading at 79%, 69.5%, or 59% sensitivity levels without CAD, then the corresponding sensitivity levels achieved by adding CAD would be 94%, 83%, or 71%, respectively, for an absolute increase in sensitivity of 15%, 13.5%, or 12%, respectively, associated with using CAD. The relative increase in sensitivity was 19.5% at each of these levels of sensitivity. In all these hypothetical scenarios, the effect of CAD on specificity was minimal, with specificity falling from 94% to 93% by adding CAD.

Another group of studies including 4 retrospective studies (5 reports) evaluate the ability of CAD to mark missed cancers and provide consistent evidence that CAD has the ability to mark a significant majority (62.7–77%) of cancers that were prospectively missed in clinical practice. However, not all prospectively missed cancers would meet a radiologist's criteria to be recommended for further work-up, so further analysis to determine the potential benefit of CAD is necessary. Based on blinded retrospective review by study radiologists, it was estimated that 23–45% of the actionable cancers (i.e., those that warranted further work-up) would have potentially been detected had CAD been used to point these lesions out to the radiologist.

The results of 3 additional retrospective studies comparing radiologist sensitivity and specificity without and with CAD in case samples artificially enriched with cancer cases do not suggest that the addition of CAD results in a significant increase in sensitivity. However, it seems reasonable that the weaknesses in these retrospective studies may account for the apparent differences in the observed effect when results are compared with the highest quality study.

Even though double-reading is not a prevalent practice in the U.S., it may serve as a useful comparison benchmark for CAD to consider that adding a second radiologist reader has been estimated to increase sensitivity in detecting cancer by 7–15%. Thus, the improvements in sensitivity associated with adding CAD (12–16%) calculated from the highest quality study in this Assessment are in the same range as those reported for adding a second radiologist reader. No full-text published studies were identified directly comparing CAD with double-reading.

In summary, the available evidence suggests that the use of CAD after initial radiographic interpretation as a quality adjunct to single-reader mammography in patients having film-screen mammography improves net health outcomes compared with single-reader radiologist interpretation by increasing 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).

Indication #2: CAD in Patients Having Full-Field Digital Mammography. There is insufficient evidence to permit conclusions on the effect on health outcomes of using CAD after initial radiographic interpretation as a quality adjunct to single-reader mammography in patients having full-field digital mammography.

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

Indication #1: CAD in Patients Having Film-Screen Mammography. The outcomes reported in the studies using CAD as an adjunct to the radiologist's interpretation in patients having film-screen mammography are derived from 11 studies using 3 commercially available CAD systems in various settings. The studies were conducted in a range of participating centers including community and referral centers and generally included radiologists who were experienced in mammography.

It is important to recognize that currently available CAD systems are intended to be used only after the radiologist has completed an evaluation of the images without CAD prompts and has made an initial decision whether any abnormal areas require recall of the patient for further work-up. Furthermore, currently available CAD systems should not be used to change a radiologist's reading from positive to negative based on the absence of a CAD mark in an area of concern to the radiologist. It is possible that a CAD system could be used differently than it is intended to be used. In such a setting, radiologists using CAD before carefully reviewing the films without CAD prompts might be distracted by the CAD prompts and might miss some cancers they would otherwise have detected. In this way, the use of CAD might lead to a lower true-positive rate compared with radiologists reading without CAD.

Only one of the 11 studies reported data that would suggest that CAD information was used to change a radiologist's diagnosis from positive to negative, and that study was conducted in Sweden. Based on the overall body of evidence including all the studies conducted in the U.S.,

it is likely that the outcomes observed in the investigational settings will be attainable outside the investigational setting with similarly experienced practitioners using CAD systems according to their intended use.

Indication #2: CAD in Patients Having Full-Field Digital Mammography. Whether the use of CAD as an adjunct to the radiologist's interpretation improves health outcomes in patients having full-field digital mammography has not been demonstrated in the investigational setting.

Therefore, based on the above, use of computer-assisted detection (CAD) after initial radiographic interpretation as a quality adjunct to single-reader mammography meets the TEC criteria for patients having film-screen mammography. Use of CAD in patients without a prior independent interpretation of a film-screen mammogram does not meet the TEC criteria. Finally, use of CAD devices in patients having full-field digital mammography does not meet the TEC criteria.

Contents

Assessment Objective	8	Review of Evidence	16
Background	8	Summary of Application of the Technology Evaluation Criteria	27
Methods	14	References	33
Formulation of the Assessment	15		

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

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*; David M. Eddy, M.D., Ph.D.—Scientific Advisor, *Senior Advisor for Health Policy and Management, Kaiser Permanente, Southern California*. ■ **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, BlueCross BlueShield of the Rochester Area*; Helen Darling, M.A., *President, Washington 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*; Alan M. Garber, M.D., Ph.D., *Professor of Medicine, Economics, and Health Research and Policy, Stanford University*; Steven N. Goodman, M.D., M.H.S., Ph.D., *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*; Michael A.W. Hattwick, M.D., *Woodburn Internal Medicine Associates, Ltd. American College of Physicians Appointee*; I. Craig Henderson, M.D., *Adjunct Professor of Medicine, University of California, San Francisco*; Albert R. Jonsen, Ph.D., *Professor Emeritus of Ethics in Medicine and former Chair, Department of Medical History and Ethics, University of Washington School of Medicine*; 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.*; Stephen G. Pauker, M.D., M.A.C.P., F.A.C.C., *Sara Murray Jordan Professor of Medicine, Tufts University School of Medicine; and Vice-Chairman for Clinical Affairs and Associate Physician-in-Chief, Department of Medicine, New England Medical Center*; William R. Phillips, M.D., M.P.H., *Clinical Professor of Family Medicine, University of Washington—American Academy of Family Physicians' Appointee*; Thomas J. Ryan, M.D., M.A.C.P., M.A.C.C., *Senior Consultant in Cardiology, the University Hospital, Boston University Medical Center; and Professor of Medicine, Boston University School of Medicine*; Earl P. Steinberg, M.D., M.P.P., *President, Resolution Health, Inc.*; Paul J. Wallace, M.D., *Executive Director, Care Management Institute, Kaiser Permanente*; 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

The objective of this Assessment is to evaluate the clinical effectiveness of using computer-aided detection (CAD) as an adjunct to mammography. Mammography is used for breast cancer screening and diagnosis to detect and characterize breast masses and calcifications that may be malignant. Conventional mammography uses film-screen technology and achieves approximately 85% sensitivity in detecting cancer though results are operator-dependent and may vary with reader expertise. There is considerable interest in finding techniques to improve sensitivity and reduce variability among readers.

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* if 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). This Assessment will focus exclusively on the use of commercially available computer analysis systems to aid or assist in detection of potential abnormalities.

Currently available CAD systems are intended to be used only after the radiologist has completed an evaluation of the images without CAD prompts and has made an initial decision whether any abnormal areas require recall of the patient for further work-up. If a radiologist identifies an abnormal area of concern on a mammogram during initial reading and that area does not get marked by CAD, the radiologist is still advised to interpret the mammogram as positive and to recall the patient for further work-up. This is because CAD is not 100% sensitive in marking all cancers, particularly larger masses.

Therefore, when used as intended, CAD would be expected to increase the number of mammograms interpreted as positive to the extent that it points out abnormalities previously overlooked by the radiologist on unaided reading. It is possible that a CAD system could be used differently than it is intended to be used. In such a setting, radiologists using CAD before carefully reviewing the films without CAD prompts might

be distracted by the CAD prompts and might miss some cancers they would otherwise have detected. In this way, the use of CAD might lead to a lower true-positive rate compared with radiologists reading without CAD.

The outcomes of primary interest in this Assessment are intermediate outcomes including the effect of adding CAD on 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 mammography improves these intermediate outcomes.

Background

Mammography

Film-screen mammography has been successful in improving detection of cancer, particularly non-palpable breast masses and calcifications that may be malignant. There has been some recent controversy over the benefit of mammography screening and the available evidence relating mammography screening with mortality may not be definitive. Nonetheless, a recent Institute of Medicine Report on Mammography (Committee on the Early Detection of Breast Cancer 2001) suggests that the reduction in mortality from breast cancer observed in recent years may be due, at least in part, to earlier detection through mammography screening.

Conventional mammography uses film-screen technology and achieves approximately 85% sensitivity in detecting cancer. However, considerable variability has been observed in a national study reporting sensitivity for some radiologists as low as 47% (Beam et al. 1996). Furthermore, Elmore et al. (1994) found substantial diagnostic disagreements between radiologists in 19% of cases selected for the study. The median level of interobserver agreement was 78% for diagnostic interpretation and 85% for biopsy recommendation. There is considerable interest in techniques to improve sensitivity and reduce variability among readers.

Reasons for false-negative mammograms include errors of perception (i.e., failure to perceive an abnormality), errors in interpretation (i.e., failure to interpret an abnormality correctly), technical errors (e.g., lesion not included on the film due to positioning), or limitations of the mammography imaging technique itself such that the lesion is not visible. Perceptual errors may be increased by radiologist fatigue, distraction, or other infringements on the reading environment.

Computer-Aided Detection (CAD) Systems

Computer analysis of mammographic images can either be used as an aid to *detect* potential abnormalities by marking suspicious areas for review by the radiologist or used as an aid for feature analysis to help *diagnose* whether a particular abnormality is likely to be malignant or not. The abbreviation “CAD” can therefore be used to refer to either computer-aided (or “-assisted”) detection or diagnosis. The currently available commercial CAD systems are designed and intended to aid in *detection* of potential abnormalities (i.e., to point to areas that might have been overlooked) and are not intended to influence a radiologist’s interpretation or *diagnosis*.

In some studies, the addition of CAD analysis has been reported to improve the accuracy of diagnostic interpretation compared to radiologists reading alone with increased area under the receiver-operating-characteristic (ROC) curve. However, none of these studies uses CAD algorithms that are commercially available. These studies suggest that CAD may be helpful in selecting which abnormalities should undergo biopsy (Leichter et al. 2000; Chan et al. 1999; Jiang et al. 1999; Baker et al. 1995). Jiang et al. (2001) noted that CAD reduced variability among radiologists’ interpretations of microcalcifications on mammography. However, one study by Zheng et al. (2001) explored the effect of different CAD cuing environments on radiologist performance. This study found that while CAD systems with high sensitivity and specificity can significantly improve radiologists’ performance, CAD systems with poor sensitivity and specificity can potentially distract radiologists and significantly worsen radiologists’ performance in both cued and noncued areas.

This Assessment will focus exclusively on the use of commercially available computer analysis systems to aid or assist in *detection* of potential abnormalities.

Performance of CAD in Marking Proven Cancers

Multiple studies have applied CAD algorithms to mammography cases with known cancers in order to determine how sensitive the CAD algorithm is in correctly marking known cancerous lesions. Each proprietary system uses one type of mark for microcalcifications and another type of mark for mass lesions. The algorithms used to detect microcalcifications or masses do not have one inherent or fixed level of sensitivity and specificity. Rather, the manufacturers developing CAD systems make choices how to adjust or tune the analysis software in an effort to achieve a desirable balance between sensitivity and specificity for detecting various types of cancer lesions.

In these studies, the CAD system is generally credited with correctly marking a lesion if a CAD mark is placed on at least one view (i.e., either the craniocaudal or mediolateral view of the breast). Table 1 summarizes the sensitivity reported for the ImageChecker[®], Second Look[™], and MammoReader[™] CAD systems for detecting cancerous lesions overall, those appearing as a mass, and those appearing with microcalcifications.

Overall, these results show that the commercially available CAD systems that have been evaluated mark approximately 80–90% of cancerous lesions and that these CAD systems are better in detection of malignant microcalcifications (91–100% marked) than they are in detection of malignant masses (67–89% marked). Because the sensitivity of CAD is greater for detection of microcalcifications than masses, the overall sensitivity reported across a sample of mammograms will depend on the proportion of cancer cases in that sample that appear as masses versus microcalcifications. It should also be noted that CAD algorithms for detection of masses may not be uniformly sensitive for masses of differing size, and CAD may be better for detecting smaller masses (e.g., 10–30 mm) than for larger masses (e.g., >40 mm) (Malich et al. 2001b [abstract]). Current CAD systems analyze each image separately, but one recent report of a non-commercial system explored the value of fusing two-view information in order to improve sensitivity in detecting mass lesions (Paquerault et al. 2002).

Table 2 shows one study (Lechner et al. 2002) that compared the ability of the ImageChecker[®] and Second Look[™] CAD systems to mark cancerous lesions in the same set of 126 proven

Table 1. Sensitivity of CAD Systems in Marking Cancerous Lesions

Study	Overall	Masses	Microcalcifications
ImageChecker®			
Castellino 2000, v. 2.2	84% (906/1,083)	75% (506/677)	99% (400/406)
Freer 2001, v. 2.0		67%	100%
Vyborny 2000, v.2.0		86% (322/375) of clearly spiculated masses 79% when a looser definition of spiculation was used.	
Burhenne 2000, v.1.2	84% (906/1,083)	75% (506/677)	99% (400/406)
FDA PMA Data, v. 1.2	83% (903/1,083)	75% (507/679)	98% (396/404)
Thurfjell 1998, v. 1.2	73%		
Nakahara 1998, v.1.2	86% (56/65)	79% (34/43)	100% (22/22)
Brem 2001, v. 1.0			98%
MammoReader™			
FDA PMA Data (n=465 cases)	89% (415/465)	87% (259/296)	91% (154/169)
FDA PMA Data (n=327 cases)	83% (189/228)	79% (129/163)	92% (66/72)
Second Look™			
Malich 2001a	90% (135/150)	88.7% (110/124)	98.2% (55/56)
FDA PMA ROSE-1D	85% (791/930)		

Table 2. Comparison of Sensitivity of ImageChecker M1000 and Second Look CAD Systems

Study	Overall n=126	Masses n=69	Microcalcifications n=47	Mixed mass/calc n=10
Lechner 2002				
ImageChecker®	90.5%	84%	100%	90%
Second Look™	88.1%	82.6%	93.6%	100%

cancer lesions. No statistically significant differences between the two systems were found.

In order to achieve this level of sensitivity, the CAD algorithms also place some false marks in areas that do not represent malignancy. Table 3 summarizes findings of several studies regarding the false marking rate for each of the three CAD systems. For the ImageChecker®, the false mark rate has improved with newer generations of the algorithm software, and on average, looking at all 3 systems, CAD places about one false mark per mammogram image.

Technical Issues in CAD

CAD analysis has most frequently been applied by digitizing conventional film-screen mammograms but it can also be applied directly to digitally acquired mammograms. The direct application of CAD software to digital mammography images offers an efficiency because the intermediate step of having to scan the film image into a laser scanner (digitization) is eliminated; however, challenges exist in designing appropriate interfaces between raw digital data from various digital mammography vendors and CAD technology (Worrell et al. 2001). The potential for permitting interaction between CAD findings and image display formats with digital mammography raises the potential for synergy between the two systems. Qian et al. (1999 and 2001) noted that adaptive modeling with CAD combined with digital mammography might yield greater diagnostic performance compared with nonadaptive CAD methods for detection of masses.

The ImageChecker® has recently received U.S. Food and Drug Administration (FDA) approval to combine its CAD technology directly with the digital mammography system manufactured by General Electric (GE Medical Systems 2002). CADx Medical Systems (Quebec, Canada) also announced its intent to develop and market its Second Look™ CAD product in conjunction with Hologic, Inc.'s LORAD® digital mammography system (CADx Medical Systems 2000). As yet, no clinical effectiveness data using either of these systems have been published in the literature though studies using non-commercially available CAD programs with raw digital mammograms have been published (Li et al. 2002; Nishikawa et al. 1995; Nawano et al. 1999).

Nawano et al. (1999) used a non-commercially available CAD program to analyze 4,148 digital mammograms containing 267 breast cancer

cases. In this dataset, CAD was reported to have 89.9% sensitivity in detecting cancerous masses with 1.35 false-positive marks per image and 92.8% sensitivity in detecting malignant microcalcifications with 0.40 false-positive marks per image. In a separate ROC analysis using 5 observers reading 344 mammograms in 86 women without and with CAD, the average area under the ROC curve was significantly higher for readers using CAD ($A_z=0.92$) compared with readers not using CAD ($A_z=0.84$; $p<0.022$).

Ideally, CAD algorithms should be reproducible in yielding high sensitivity for detecting malignancy and few false-positive marks. Malich et al. (2000) evaluated the reproducibility of CAD markings using the ImageChecker M1000® (v. 3.1, R2 Technology, Los Altos, CA, USA) and found in 100 cases with histologically proven malignancy that only 18% of cases were marked the same when scanned through the CAD system 3 separate times. However, FDA PMA data submitted for the ImageChecker® system used 25 “well-characterized” cases “where the primary lesion was correctly marked in both views.” In this study, the ImageChecker® yielded correct markings in 1499/1500 runs (99.9% precision). Results were similar in FDA PMA data for the Second Look™ CAD system whereby 25 screen-detected cancer cases were correctly marked in 745 of 750 runs (99% reproducibility) and for the MammoReader™ CAD system where 60 cases were scanned with 93.3% reproducibility.

FDA PMA data for the ImageChecker M1000® system documented only 7 days of downtime for 5 devices operated over a period of 31 months (FDA PMA ImageChecker®). Similarly, FDA PMA data for the Second Look™ system estimated 18 days of downtime for 7 systems operated over 59 months (FDA PMA Second Look™).

The impact on efficiency and workflow of using CAD based on digitized film-screen mammograms has been explored in several studies. Baum et al. (2001) presented abstract data suggesting that radiologist reading time was increased by 20% or less in the majority of cases. Shile et al. (2001) published similar results showing that radiologist reading time increased 17 seconds by using CAD compared with an average of 1 minute, 16 seconds, reading without CAD (22% increase). Also, additional technologist processing time associated with the use of CAD was 1 minute, 21 seconds, on average.

Table 3. Studies that Report Number of False CAD Marks Placed on Images

Study	Average Number of False Marks per Image	Comments
ImageChecker®		
Freer 2001, v. 2.0 (n=5,204 cases)	0.7 total 0.3 MC 0.4 mass	Of 14,214 marks, 368 (2.6%) were considered actionable by the radiologist and 13,846 (97.4%) were dismissed.
Birdwell 2001, v. 2.0 (n=110 cases)	0.73 total	CAD placed an average of 4.3 marks per case of which 1.4 marks were on the cancer.
Vyborny 2000, v. 2.0 (n=677 cases)	0.24 mass	
Burhenne 2000, v. 1.2	1.0 Normal/benign cases	
FDA PMA Data, v. 1.2	1.8 all n=1083 0.9 normals n=100 1.9 MC n=404 1.8 mass n=679	
Thurfjell 1998, v. 1.2 (n=120 cases)	0.54 Normals 0.08 MC 0.46 Mass	
Nakahara 1998, v. 1.2 (n=65 cases)	0.58 MC 0.20 mass	
Brem 2001, v. 1.0	1.2 total	0.8 for benign cases 0.5 for normal cases
MammoReader™		
FDA PMA Data (n=265 cases)	0.83 normals 0.12 MC 0.71 mass	
Second Look™		
Malich 2001a (n=150 cases)	1.25 total 0.28 MC 0.97 mass	

The issue of whether CAD may be more useful as an aid in detection when used by less-experienced radiologists has been explored. One abstract by Balleyguier et al. (2001) compared sensitivity and specificity without and with CAD for a senior radiologist with 8 years of experience and for a radiology resident with a 5-month training program in mammography. This study used the Second Look™ CAD system and found that sensitivity for the senior radiologist increased from 76.9% to 84.6% with use of CAD, while sensitivity for the resident increased from 61.9% to 84.6% with use of CAD. Specificity did not change for either reader when using CAD.

Double-Reading in Mammography

The use of a second radiologist to review a mammogram after the first radiologist has reviewed the mammogram is generally referred to as “double-reading.” This practice has been proposed as a means to improve the sensitivity of mammographic screening and improvements in cancer detection of 7–15% have been reported with double-reading (Kopans 2000; Thurfjell et al. 1994). Kopans explains:

Double-reading may be used merely to reduce perceptual errors by having two readers review the cases. This form of double reading may detect additional cancers, but it also leads to an increase in false-positive recalls. Double reading may also be considered as double interpretation where the second reader also reviews the decisions made by the first reader and may reverse a suggested call-back as being unwarranted. This latter double interpretation is potentially a way to reduce false-positive interpretations, although it could lead to an increase in false-negatives if the second reader cancels a recall on a patient who later proves to have cancer.

However, despite the potential improvement in sensitivity afforded by double-reading, this practice is not widely used in the U.S.; double-reading is more common outside the U.S. (Castellino 2002). Suggested reasons for the low rate of double-reading in the U.S. include logistics, shortages of radiologists, and reimbursement (Kopans 2000; Thurfjell et al. 1994; Mitka 2001).

FDA Status. There are three manufacturers that have received FDA approval to market

computer-aided detection (CAD) systems that take film-screen mammograms and digitize the images for computer analysis.

The first device to receive FDA approval, the ImageCheckerM1000® by R2 Technology, Inc. (Los Altos, CA), was approved by PMA (P970058) on June 26, 1998. The initial product labeling was for use on routine screening mammograms, but on May 29, 2001, approval was granted for the expansion of the “Indications for Use” to cover diagnostic as well as screening mammograms. The ImageCheckerM1000® is:

“intended to identify and mark regions of interest on routine screening and diagnostic mammograms to bring them to the attention of the radiologist after the initial reading has been completed. Thus, the system assists the radiologist in minimizing observational oversights by identifying areas on the original mammogram that may warrant a second review. (<http://www.fda.gov/cdrh/pma/pmamay01.html>; accessed July 15, 2002)

Multiple supplemental applications and approvals have also been processed relating to modifications of the ImageChecker hardware and software. A “New Efficacy Claim” was approved in PMA supplement 7 (version 2.2 software). The change is from:

“For every 100,000 cancers currently detected by screening mammography, the use of the ImageChecker could result in early detection of an additional 30,500 breast cancers.”

to:

“Use of the ImageChecker could result in earlier detection of up to 23.4% (95% CI, 19.4%–27.4%) of the cancers currently detected with screening mammography in those women who had a prior screening mammogram 9–24 months earlier.” (<http://www.fda.gov/cdrh/pma/pmafeb02.html>; accessed July 15, 2002)

Two additional CAD systems have also been FDA approved. Second Look™ by CADx Medical Systems, Inc. (Northborough, MA) was approved by PMA (P010034) on January 31, 2002, and the MammoReader™ by Intelligent Systems Software, Inc. (Clearwater, FL) was approved by PMA (P010038) on January 15, 2002.

Each of these devices is:

“intended to identify and mark regions of interest on standard mammographic views to bring them to the attention of the radiologist after the initial reading has been completed. Thus, the system assists the radiologist in minimizing observational oversights by identifying areas on the original mammogram that may warrant a second review.” (<http://www.fda.gov/cdrh/pma/pmajan02.html>; accessed July 15, 2002)

In addition, General Electric Medical Systems (Milwaukee, WI) received FDA approval on April 12, 2002 to modify their Senographe 2000D® Full Field Digital Mammography System (P990066/S007) to permit integration and use of the ImageChecker® M1000-DM Computer Aided Detection (CAD) system manufactured by R2 Technology, Inc. (<http://www.fda.gov/cdrh/pma/pmaapr02.html>; accessed August 21, 2002)

Methods

Search Methods

Studies of the use of CAD were identified through a computerized online search of the MEDLINE database (via PubMed) through November 2002 using the terms “mammography” OR “mammographic” OR “mammogram” combined with “radiographic image interpretation, computer-assisted” OR “diagnosis, computer-assisted” OR “CAD” OR “computer-assisted” OR “computer-aided.” Results were limited to English-language publications in human subjects. To identify more recent studies, the MEDLINE search was supplemented by searches of *Current Contents*, by manual searches of the most recent issues of the pertinent journals, and by reading the reference lists of the most recently published reports.

In addition, the American College of Radiology and each of the three manufacturers of commercially available CAD systems were contacted to provide additional studies.

Study Selection

Studies were systematically included in the review of evidence if the study:

- was published in English as a full text journal article in a peer-reviewed journal, and
- included an analysis comparing the outcomes of radiologist interpretation of mammograms using a commercially available CAD system to the outcomes obtained using single- or double-reader radiologist interpretation.

Studies simply reporting the ability of CAD to mark known cancerous lesions were not systematically included in the review of evidence but were summarized in the background. Studies reported only in abstract form were not systematically included in the review of evidence, but selected abstracts that supported existing published studies were included as supplemental information.

Study Quality Assessment

In order to evaluate studies of the clinical effectiveness of CAD as an adjunct to a radiologist’s interpretation of mammograms, three quality criteria were considered:

- Is the study population representative of the population for intended use?
- Are the study radiologists interpreting mammograms in a representative setting?
- Does the study provide a complete assessment of relevant outcomes (both true-positive rate and false-positive rate)?

This Assessment looked for studies that included study populations that were relevant to the proposed clinical use, and better quality studies would be those that included unselected consecutive cases in the routine screening or diagnostic setting. Studies that evaluate only a selected subset of cases, such as a subset of cancer cases missed during prospective screening, provide only indirect evidence of the potential effect of CAD in an unselected setting.

With regard to radiologist interpretation setting, radiologists may apply different diagnostic 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.

The outcomes of primary interest in this Assessment are intermediate outcomes related to the diagnostic results of mammography, true-positive rate (related measures include

cancer detection rate or sensitivity) and false-positive rate (related measures include recall rate, biopsy rate, or specificity). While recall rate and biopsy rate are not actually measures of the false-positive rate per se, an increase in recall rate or biopsy rate represents a combination of the increase in false-positives and any increase in true-positives. In the screening setting, the proportion of false-positives reflected in the recall or biopsy rate is generally higher than the proportion of true-positives based on the very low prevalence of cancer. Thus, terms “true-positive rate” and “false-positive rate” will be used to establish a conceptual framework within which the various indicators or measures of true-positives and false-positives reported in the studies can be synthesized and analyzed.

Similar to diagnostic sensitivity and specificity, true-positive rate and false-positive rate are interrelated outcome measures, and it is important to consider both measures in evaluating the results of an individual study and in comparing results across studies. The better quality studies would provide assessments of both true-positive rate and false-positive rate in the same study population (i.e., cancer detection rate and recall or biopsy rate). One would want to compare increases in cancer detection rate with any increases in recall or biopsy rates, and studies that report only recall rates without any information on cancer detection rates or studies that provide cancer detection rate without information on recall rate are of lesser value.

Studies were grouped into Level 1 if all 3 of the quality criteria were met or were grouped into Level 2 if any one of the 3 criteria was not met.

Medical Advisory Panel Review

This Assessment was reviewed by the Blue Cross and Blue Shield Association Medical Advisory Panel (MAP) on October 10, 2002. In order to maintain the timeliness of the scientific information in this Assessment, literature searches 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 table(s) and text where appropriate. There were no studies that would change the conclusions of this Assessment.

Formulation of the Assessment

Patient Indications

1. Patients having film-screen mammography
2. Patients having full-field digital mammography

Technologies to be Compared

Computer-assisted detection (CAD) is intended as an adjunct to the radiologist’s interpretation of a mammogram. In some settings, CAD may be considered as an alternative to using a second independent radiologist interpretation (double-reading).

Health Outcomes

The outcomes of primary interest in this Assessment are intermediate outcomes including the incremental effect of using CAD on true-positive rate (related measures include cancer detection rates or sensitivity) and the incremental effect of using CAD on false-positive rate (related measures include recall rate, biopsy rate or specificity). The relationship between these intermediate outcomes and health outcomes has been addressed by studies linking the use of mammography screening with improvements in health outcomes such as mortality. While there has been some recent controversy over the benefit of mammography screening, it seems likely that mammography screening is associated with a reduction in mortality.

Benefits. Improved cancer detection or improved sensitivity (i.e., reduced false-negatives) with the use of CAD would result in earlier detection of breast cancer, which is presumed to translate into reduced disease-related morbidity or mortality.

Harms. Additional false-positive results prompted by the CAD system may result in additional work-up and/or biopsy which have associated anxiety, radiation exposure, and morbidity associated with biopsy.

If used as directed, CAD should not result in a decrease in cancer detection rate (or sensitivity) compared to radiologists reading alone since the lack of a CAD mark on an area of concern should not be used to dissuade a radiologist from working up a finding. However, if the radiologist does not perform a complete review of the films and moves too quickly to looking at

the CAD marks, CAD markings have the potential to distract the radiologist's attention away from other actionable findings and potentially lead to lower sensitivity, with resultant delay in diagnosis and potential for increased disease-related morbidity or mortality.

Specific Assessment Question

Does the available evidence demonstrate that the use of computer-assisted detection (CAD) improves health outcomes when used as an adjunct to single-reader mammography?

Review of Evidence

Overview

The systematic review of the literature yielded 11 studies in distinct patient cohorts that were eligible for inclusion. All of these studies used CAD on screen-film mammography and thus the results apply to the first patient indication for this Assessment. Seven of these studies (8 reports) used the ImageChecker® system (Freer et al. 2001, n=12,860; Garvican et al. 2001, n=104; Moberg et al. 2001, n=280; Brem et al. 2001, n=106; Burhenne et al. 2000^{1,2} Sensitivity Study, n=286; Birdwell et al. 2001², n=110; Burhenne et al. 2000¹ Recall Rate Study, n=14,817 CAD, n=23,682 no CAD; Thurffjell et al. 1998, n=120). Two eligible studies were identified for the Second Look™ system and both are derived from the FDA PMA data (ROSE-2, n=3,946; ROSE-1M, n=177). Similarly, two FDA PMA data studies are available for the MammoReader™ (Specificity Study, n=300; Sensitivity Study, n=327). One unpublished abstract by Cupples (2001) is also described in the review of evidence because it reports estimates of both true-positive rate and false-positive rate using the ImageChecker® system. Table 4 provides an overview of each eligible CAD study including a description of study methods and ratings according to the 3 quality criteria.

Does the available evidence demonstrate that the use of computer-assisted detection (CAD) improves health outcomes when used as an adjunct to single-reader mammography?

Indication #1. Patients having Film-Screen Mammography. The review of evidence will address the evidence describing the effect of CAD on true-positive rate and the evidence describing the effect of CAD on false-positive rate. It is preferred to have both outcomes evaluated within the same study so that the relative weight of benefits and harms can be accurately compared within studies and across studies. However, only one study (Freer et al. 2001) includes both true-positive and false-positive effects in the same report, and all other studies failed to meet this quality criterion. Freer et al. (2001) also meets the other two quality criteria by virtue of being conducted in a representative, consecutive, screening population by readers working prospectively in the usual clinical setting. Three additional studies met the quality criterion for study population, and 6 studies met the criterion for reader setting, but none of the other studies met all 3 criteria. Thus, Freer et al. is the only study that is listed as Level 1 and all other studies are grouped in Level 2.

Freer et al. (2001; n=12,860) is the only published study that provides direct evidence on the effect of CAD on cancer detection rate in a clinically relevant screening setting and also reports effect of CAD on recall and biopsy rates (see Table 5). Another study has been conducted that reports cancer detection rates, recall rates, and biopsy rates in a regional screening center before and after CAD installation, but this study is only available as a presentation abstract (Cupples 2001, n=9,102 without CAD, n=12,217 with CAD)⁵.

In the Freer et al. study, two experienced community practice mammographers prospectively interpreted 12,860 eligible screening mammograms. Mammograms were analyzed by the ImageChecker M1000® v. 2.0 CAD system. Before reviewing the CAD prompts, the radiologist recorded an initial interpretation and a decision whether or not to recall the patient for additional evaluation. Then the CAD analysis was provided, and the radiologist re-evaluated any areas prompted by the CAD system and recorded a second interpretation

¹ The Burhenne et al. (2000) article reports both a retrospective sensitivity study and a prospective recall rate study.

² The Burhenne et al. (2000) sensitivity study was also reported in the FDA PMA Data, and the same mammography case material was re-analyzed using a later software version of the ImageChecker and reported in Birdwell et al. (2001)

⁵ In the presentation made by Dr. Cupples at the 2001 RSNA (personal communication from R2 technology, Inc.), results obtained before and after CAD installation included: cancer detection rate improved from 3.67 to 4.32 cancers per 1,000 screens (17.7% increase); recall rate increased from 7.34% to 7.91% (7.8% increase); and the biopsy rate decreased from 1.59% to 1.49% with an increase in positive predictive value from 26.9% to 29.5% (9.7% increase).

Table 4. Overview of CAD Studies: Quality Criteria Ratings and Study Description

Study Author / Year Country, Software release	Study Population	Reader Setting	Complete Outcomes		Study Description
			TP	FP	
ImageChecker®					
Freer 2001 U.S., Version 2.0	X	X	X	X	Prospective, single-center study of 12,860 consecutive screening mammograms at single center in community practice setting. CAD results displayed after initial radiologist interpretation and CAD could only be used to increase recall decisions not to reverse initial radiologist decision to recall patient Experienced mammographers
Burhenne 2000 U.S., Version 1.2 Recall Rate Study	X	X		X	Prospective, multicenter, comparison of recall rates in 14,817 screenings mammograms after CAD and 23, 682 screenings mammograms before CAD 14 radiologists (MQSA qualified) from 5 centers varying practices
Burhenne 2000 U.S., Version 1.2 Sensitivity Study and FDA PMA Protocol 2		?	X		Retrospective, multicenter study of 286 patients with cancers that were retrospectively considered visible (unblinded review) on prior mammogram (9-24 months earlier) drawn from 493 priors available on 1,083 current mammograms with proven malignancy Radiologists were specialists in breast imaging and made blinded assessment of whether lesions were actionable or not.
Birdwell 2001 ¹ U.S., Version 2.0		?	X		Retrospective, multicenter study of 110 patients (115 lesions) with cancer missed on prior mammograms, but lesions considered to be visible in retrospect by at least 3 of 5 radiologists (blinded review) – 71% invasive and 29% DCIS, 2 ambiguous cases excluded Radiologists were specialists in breast imaging
Garvican 2001 U.K., Version 2.0		?	X		Retrospective, single-center study of 104 cancer cases who had previously reported normal screening mammograms 3 readers: 1 expert, 1 new radiologist, and 1 experienced “second reader” who had not been formally trained in radiology
Moberg 2001 Sweden, Version 1.2		?	X		Retrospective, single-center study of enriched mixed set of 280 films 59 women with “interval tumors” and prior films available out of 60 who had a breast cancer diagnosed within 2 yr. after a negative screening Mammograms were mixed with 221 screening mammograms including 211 healthy women (9 yr. f/u), 5 confirmed malignancies, 5 confirmed benign lesions (9 yr. f/u) 3 radiologists with 5-10 yr. experience in mammography

¹ The Burhenne et al. (2000) sensitivity study was also reported in the FDA PMA Data, and the same mammography case material was re-analyzed using a later software version of the ImageChecker and reported in Birdwell et al. (2001)

Table 4. Overview of CAD Studies: Quality Criteria Ratings and Study Description (cont'd)

Study Author / Year Country, Software release	Study Population	Reader Setting	Complete Outcomes		Study Description
			TP	FP	
ImageChecker® (cont'd)					
Brem 2001 U.S., Version 1.0		?	X ²		Retrospective, single-center study of 106 consecutive patients ³ included: 24 normals; 40 benign MC; 42 malignant MC (47 malignant masses and 35 benign masses were mixed in but not included in the analysis) Reference Standard: Histology for lesions and 1-yr. f/u for normals. 60% of malignant MC were DCIS, 40% invasive 5 highly experienced mammographers
Thurfjell 1998 Sweden, Version 1.2		?	X		Retrospective, single-center study of 120 cases including 74 cases of breast cancer and 46 randomly selected normals. Expert radiologist did not read with CAD. Sensitivity was calculated only in 51 cancers "that had been detected by at least one screener ... in the earlier study" The screener radiologist had 5 yr. experience in mammography and 2 yr. mass screening. The clinical radiologist had 7 yr. experience in mammography
Second Look™					
FDA PMA Data ROSE-1M		?	X		Retrospective, multicenter study of 177 missed cancer cases within 242 retrospectively visible cancer cases drawn from 377 screening mammograms originally interpreted as normal or benign within 24 months prior to screening mammogram with cancer diagnosis.
FDA PMA Data ROSE-2	X	X		X	Prospective, multicenter study of 3,946 consecutive screening mammograms read before and after CAD. Recommendations for recall and further work-up were compared. 10 experienced mammographers at 5 institutions
MammoReader™					
FDA PMA Data Sensitivity Study					Retrospective, multicenter study of 327 cases of cancer with prior negative screening mammogram 4 to 27 months earlier (average 14 mos.) 6 experienced radiologists with MQSA certification and average of 17 yr. mammography experience
FDA PMA Specificity Study	X	X		X	Prospective, multicenter study of 300 consecutive routine screening exams. Cases were randomly assigned so each of 10 radiologists read half the cases with CAD and half without CAD 10 radiologist were MQSA certified with average of 10 yr. mammography experience

² Microcalcifications only³ Patients were reported to be consecutively selected but criteria for sample selection were not specified

Table 5. Studies to Determine Effect of CAD on Cancer Detection Rate or Recall or Biopsy Rate

Study	Cancer Detection: Radiologist Alone	Cancer Detection: Radiologist + CAD	Recall and Biopsy Rate: Radiologist Alone	Recall Rate: Radiologist + CAD	Comments
ImageChecker®					
Freer 2001 v. 2.0	41/12,860 = 0.32% [.23, .43]	49/12,860 = 0.38% [.28, .50]	Recall 830/12,860 6.5% [6.0, 6.9]	Recall 986/12,860 7.7% [7.2, 8.1] Relative increase of 18.8%	Recall rate was based on actionable findings
	96% of masses 68% of microcalc cancers were detected by radiologist alone	Relative increase of 19.5% [9.0-42.2%] Five of 8 cancers detected by CAD were Stage 0 and remaining 3 were Stage 1	Biopsy 107	Of additional 156 cases recalled 49% were microcalcification and 51% were mass/architectural distortion Biopsy 128	27 women excluded (8 with signs of cancer, 11 during time when CAD nonoperational, 18 unable to be analyzed by CAD usually due to small amount of tissue
Burhenne 2000 Recall Rate Study FDA PMA Protocol 1 v. 1.2			Recall: 8.3% 1961/23,682 historical review over period at least 4 months	Recall: 7.6% 1126/14,817 p=n.s. prospectively collected after CAD installation over period at least 4 months	
MammoReader™					
FDA PMA Data Specificity study			Recall: 15.1% [13.3 to 17.0%] 226/1,500	Recall: 18.1% [16.1 to 20.1%] 271/1,500, p=0.03	
Second Look™					
FDA PMA Data ROSE-2			Recall: 16.6% [15.5 to 17.8%] 657/3,946	Recall: 17.2% [16.0 to 18.4%] 677/3,946, p=n.s.	
Data in brackets are 95% confidence intervals MC = microcalcifications					

and decision. Clinical management was guided by the final reading including cases recommended for further evaluation based on CAD findings. The protocol did not permit the radiologist to reverse an initial decision to recall the patient if CAD did not place a mark in the area of interest. Thus, the use of CAD in this study was purely as an aid to detection and CAD could only increase the cancer detection and recall rate.

Radiologists alone detected 41 cancers for a detection rate of 0.32% (41/12,860). 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. This included an average of 1.2 microcalcification marks and 1.6 mass or architectural distortion marks. The vast majority of CAD marks (97.4%) were dismissed by the radiologist but an additional 156 cases were recommended for recall. Thus, the end result of using CAD information on recall rate was an 18.8% relative increase from 6.5% to 7.7% of subjects.

The number of women actually recommended to have biopsy as a result of the radiologist interpretation alone was 107 and an additional 21 women were recommended to have a biopsy based on additional information provided by CAD. The positive biopsy yield was similar in both groups: 38% (41/107) and 38% (8/21), respectively.

In comparing the trade-off between increased recall rate and biopsy and increased detection, an additional 156 women were recalled using CAD and an additional 21 women were biopsied in order to detect an additional 8 cancers, yielding 19.5 recalls and 2.6 biopsies for each additional case of cancer detected. It may be useful to compare these benchmarks

to the results based on radiologists reading without CAD, and in that setting 850 recalls and 107 biopsies were recommended to detect 41 cancers, yielding 20.2 recalls and 2.6 biopsies per case of cancer detected.

Because the Freer study did not employ an independent reference standard to ascertain the presence of any false-negative cases, sensitivity and specificity are not reported. However, if one makes an assumption about the level of sensitivity at which these study radiologists were reading screening mammograms without CAD, then sensitivity and specificity rates can be calculated from the Freer et al. (2001) data. Table 6 summarizes several hypothetical calculations to illustrate the potential effect of CAD on sensitivity and specificity. As a benchmark, sensitivity in screening mammography is cited to be about 80-85% in the literature, but considerably lower levels of performance have also been documented in the literature (Beam et al. 1996).

Recall rates without CAD and with CAD are also reported in two other prospective observational studies (Burhenne et al. 2000; Second Look ROSE-2) and one prospective randomized study (MammoReader[®] specificity study). These studies are all rated in Level 2 quality because they fail to also report outcomes on true-positives.

The Second Look ROSE-2 study was conducted in a moderately large representative screening sample (n=3,946) and found that twenty additional recalls were prompted by CAD, a nonsignificant increase from 16.6% to 17.2%.

Both the MammoReader[™] specificity study and the Burhenne et al. (2000) study compare recall rates in different samples of patients, which is a significant limitation of these studies because of the inherent variability in case-mix and recall rates between different case samples. Burhenne compares recall rates in a multicenter prospective sample of screening patients before CAD was installed (n=23,682) and after CAD was installed (n=14,817). This study found no significant difference in recall rate, 8.3% without CAD (n=23,682) and 7.6% with CAD (n=14,817).

The MammoReader[™] FDA PMA data reports a prospective randomized comparison study of 300 consecutive screening cases where each of 10 radiologists randomly read half the cases without CAD and half the cases with CAD. This study reported a small, statistically significant

Table 6. Hypothetical Calculations of Sensitivity and Specificity (Freer et al. 2001)

Assumptions for Baseline Radiologist Sensitivity	Interpretation	TP	FN	FP ¹	TN ²	Sensitivity (%)	Specificity (%)	Relative Increase in Sensitivity with CAD (%)	Absolute Increase in Sensitivity with CAD (%)
84%	RAD alone	41	8	789	12,022	84	94	19.5	16
	With CAD	49	0	937	11,874	100	93		
79%	RAD alone	41	11	789	12,019	79	94	19.5	15
	With CAD	49	3	937	11,871	94	93		
69.5%	RAD alone	41	18	789	12,012	69.5	94	19.5	13.5
	With CAD	49	10	937	11,864	83	93		
59%	RAD alone	41	28	789	12,002	59	94	19.5	12
	With CAD	49	20	937	11,854	71	93		

¹ Calculated as the number of recalls minus the number of true positives

² Calculated by subtracting (TP+FN+FP) from 12860, the total number of cases.

TP true-positive
 FN false-negative
 FP false-positive
 TN true-negative
 RAD radiologist
 CAD computer-aided detection

difference between observed recall rates, 15.1% versus 18.1% ($p=0.03$). While the use of randomization should help minimize the differences between groups, the study sample size is relatively small, and the FDA report does not provide details on the adequacy of randomization or characteristics of cases in the two groups.

An additional 4 studies (5 reports) conducted retrospectively in samples of cancers that had been missed on mammography provide some supportive evidence of the potential effectiveness of CAD. By virtue of their design, these studies do not provide direct evidence of the effect that CAD would have on cancer detection rates; however, they do provide indirect evidence of the potential effect that CAD might have had in improving earlier detection of these missed cancers had CAD been available. Furthermore, this population of cancers missed on mammography is precisely the type of cases where one would hope to have an impact by use of CAD. These studies estimate that the use of CAD might lead to detecting 23–45% of cancers that are missed at mammography.

Two studies (3 reports⁴) using the ImageChecker[®] system (Burhenne et al. 2000; Birdwell et al. 2001; Garvican et al. 2001), one study using the MammoReader[™] system (FDA PMA Sensitivity Study), and one study using the Second Look[™] system (FDA PMA ROSE-2 study) retrospectively explored the ability of CAD to mark cancers that were initially missed in clinical practice but were retrospectively determined to be visible on prior mammograms (Table 7). These studies address potential effects of CAD on true-positive rate. However, all four of these studies (5 reports) fail to report any information on the effect of CAD on false-positive rate and are, therefore, grouped in Level 2. All but one of these studies (MammoReader[™] FDA PMA Sensitivity Study) had radiologists review the prior films blinded to the ultimate location of the cancer.

In the study on the ImageChecker[®] which was part of the FDA PMA data and reported

in Burhenne et al. (2000); and Birdwell et al. (2001), the original mammography dataset was 1,083 consecutive screen-detected cancers and only 427 of these cases had a prior mammogram available. These 427 current and prior mammograms were reviewed in an unblinded fashion and 286 cases had retrospectively visible lesions corresponding with the cancer even though all prior mammograms had been prospectively read as negative in the clinical setting. These 286 prior cases were retrospectively reviewed by 5 blinded radiologists to determine if lesions were present and whether the lesions should have further work-up (i.e., were actionable lesions)⁵. A majority (at least 3 of 5) recommended further work-up in 110 cases (115 lesions).

Overall, CAD correctly marked 171/286 (59.8%) of visible prior cancers and 89/115 (77%) of actionable cancers. Based on the proportion of radiologists that determined each visible lesion to be actionable, it was estimated that the use of CAD would have resulted in detection of 31% (89 of 286) of the visible lesions on prior exams and that these lesions would have been recommended for further work-up if CAD had been used.

Garvican et al. (2001) also showed that CAD correctly prompted 5 of 11 missed cancers (45%) that radiologists agreed required work-up and 10 of 18 more subtle missed cancers, though only half of these cases would have been recommended for work-up by the radiologist.

In the FDA PMA sensitivity study for the MammoReader, three radiologists independently reviewed current and prior mammograms in 327 consecutive cancer cases for which prior films could be located. In an unblinded fashion, radiologists determined whether a lesion correlating with the current cancer was visible on prior exam. If a lesion was considered visible, then the same radiologist was asked whether the lesion would have warranted work-up had it been brought to the attention of the radiologist in a standard screening environment. This

⁴ The mammography case material was the same for the Burhenne study and the Birdwell study, but Birdwell applied a later version of the CAD software in their updated report. These two studies will be discussed together as one study since updated findings are not significantly different.

⁵ “[Five independent] Radiologists were instructed that there was a mixture of positive and negative cases, but were not advised of the prevalence of each type. The patient’s age and information about any film markers was provided. Radiologists were asked to operate at their usual clinical threshold when reviewing the cases... Priors were also excluded from review by the Panel Radiologists if the case were unilateral (that is the woman had a mastectomy) or if there had been previous surgery. Such cases would have required significant explanation of the patient’s history to the Panel Radiologist which would have created a reading bias.” (ImageChecker FDA PMA SSE).

Table 7. Studies to Determine Effect of CAD in Marking Cancers That Were Originally Missed on Mammography

Study	Sensitivity of RAD+CAD	Comments
ImageChecker		
FDA PMA and Burhenne 2000 v. 1.2 (n=286 prior cases with visible lesion)	Overall, CAD correctly marked 171/286 (59.8%) of visible prior cancers and 89/115 (77%) of actionable cancers. Subgroups: 91/112 = 81% [74 to 83%] for 23% of priors that at least 3/5 read actionable: 63/74 = 85% [77 to 88%] for 15% of priors that at least 4/5 read actionable: 33/36 = 92% [83 to 96%] for 7% of priors that 5/5 read actionable: Estimated that CAD could have correctly detected 89 or 31% of the 286 visible lesions on prior exams and that these lesions would have been recommended for further work-up if CAD had been used.	Percent of microcalcification cases marked by CAD among cases read as "actionable" by 3/5, 4/5, or 5/5 radiologists were 95%, 96%, or 100%, respectively. Percent of mass cases marked by CAD among cases read as "actionable" by 3/5, 4/5, or 5/5 radiologists were 74%, 79%, or 87%, respectively.
Birdwell 2001 v. 2.0	CAD marked 88/115 (77%) of total cases Masses: 36/54 (67%) Microcalcs: 30/35 (86%) Mass/MC: 20/24 (83%) Other: 2/2 (100%)	CAD performance was similar in nondense and dense breasts.
Garvican 2001 v. 2.0	In 11 missed cancers, CAD correctly prompted 5 (45%) and all readers agreed with the prompt In 18 cancers with minimal signs, CAD correctly prompted 10 (56%) but prompts were not accepted by at least one reader in 5 (28%) 75 cancers not visible in retrospect and CAD did not prompt these areas	
MammoReader		
FDA PMA Data Sensitivity Study	CAD correctly marked 75 of 109 missed cancers (69%) Estimated that CAD could have correctly detected 75 or 23% [18.5 to 27.9%] of the cancers in 327 prior exams and that these lesions would have been recommended for further work-up if CAD had been used.	
Second Look		
FDA PMA Data ROSE-1M	CAD correctly marked 111 of 177 missed cancers (62.7%) Estimated that CAD could have correctly detected at least 80.3 or 26.2% [21.9 to 30.7%] of the 306 retrospectively visible lesions ¹ and that these lesions would have been recommended for further work-up if CAD had been used.	

¹ The FDA analysis made an arbitrary but conservative assumption that all 64 cases not initially recommended for further work-up by the blinded readers actually had retrospectively visible lesions. These 64 cases were not actually reviewed by the unblinded readers to determine lesion visibility. Using this assumption, the maximum number of retrospectively visible lesions was 306 (242+64).

study estimated that 109 retrospectively visible cancers were actionable⁶. CAD correctly marked 75 of 109 actionable cancers (69%). It was estimated that CAD could have correctly detected 23% [18.5 to 27.9%] or 75 of the cancers on 327 prior exams and that these lesions would have been recommended for further work-up if CAD had been used.

In the FDA PMA ROSE-2 study using the Second Look™ CAD system, a blinded retrospective review was conducted by 3 radiologists on the prior mammograms of 377 current cancer cases which had previously been called negative on prior mammography. This review yielded 313 cases that were recommended for work-up by at least 1 radiologist (64 were not recommended for work-up by any radiologist). Retrospective review of these 313 cases with knowledge of current cancer location yielded 242 had retrospectively visible lesions. Of these 242 lesions, 177 were considered actionable for further work-up by at least 1 radiologist.

CAD marked 111 of these 177 lesions (62.7%) and based on the number of blinded readers who had recommended work-up it was estimated that 80.3 of these marked lesions would have been worked-up if the clinical reader had been pointed to this area. To estimate the potential improvement in true positives associated with use of CAD, it was assumed that 80.3 or 26.2% (95% CI: 21.9 to 30.7%) of the 306 retrospectively visible lesions⁷ would have been recommended for further work-up if CAD had been used. The conservative estimates used in this calculation make this a lower bound for the potential improvement in true positives (FDA PMA data ROSE-1M).

In summary, these 4 studies (5 reports) provide consistent evidence that CAD has the ability to mark a significant majority (62.7–77%) of cancers that were prospectively missed in clinical practice. Since not all prospectively missed cancers would meet a radiologist's criteria to be recommended for further work-up, further adjustments of the potential benefit of CAD are necessary. Based on blinded retrospective review by study radiologists, it was estimated

that 23–45% of actionable cancers (i.e., those that warranted further work-up) would have potentially been detected had CAD been used to point these lesions out to the radiologist.

Another group of three retrospective studies, all using the ImageChecker® system, evaluated the sensitivity and specificity of radiologists interpreting mammograms alone compared with having CAD information (Thurfjell et al. 1998 [n=120], Brem et al. 2001 [n=106], and Moberg et al. 2001 [n=280]) (see Table 8). All of these studies used enriched case sets including many more cancer cases than would be encountered in the usual clinical screening or diagnostic setting. Furthermore, the selection of cancer cases tended toward cancers that had already been detected by radiologists on mammography. In general, since these cases have already been detected by radiologists, the use of CAD may be less likely to demonstrate a significant improvement in cancer detection rate or sensitivity. All 3 of these studies employed blinded assessment by radiologists. However, the retrospective setting in which radiologists interpreted these cases along with the unusually high proportion of cancer cases included in these studies would generally not be considered representative reading settings. Thus, all 3 studies fail to meet the study population and reading setting quality criteria and are grouped into Level 2.

Thurfjell et al. (1998) noted small to moderate improvements in sensitivity with CAD depending on the reader (80% vs. 84% and 67% vs. 75%) and a small decrease in specificity for one reader (83% vs. 80%) and no change for the other reader (100% vs. 100%), comparing radiologist alone vs. using CAD respectively. However, the calculation of sensitivity was not calculated on the entire 74 total cancer cases. Rather, the analysis was restricted to the 51 cases that had been detected by at least one screener during a previous study. This case selection might bias sensitivity results in favor of the radiologists and leave less room for improvement achievable by addition of CAD results.

⁶ The number of actionable cancers was calculated by multiplying the proportion of readers recommending work-up by the number of cases (i.e., if 60 cases were rated actionable by 1 of 3 radiologists, then that would count as 20 actionable cancers).

⁷ The FDA analysis made an arbitrary but conservative assumption that all 64 cases not initially recommended for further work-up by the blinded readers actually had retrospectively visible lesions. These 64 cases were not actually reviewed by the unblinded readers to determine lesion visibility. Using this assumption, the maximum number of retrospectively visible lesions was 306 (242+64).

Table 8. CAD Studies Comparing Sensitivity and Specificity of Radiologists Reading without and with CAD

Study	Sensitivity RAD Alone	Sensitivity RAD + CAD	Specificity RAD Alone	Specificity RAD + CAD	Comments
ImageChecker®					
Thurfjell 1998 v. 1.2	Expert: 86% 44/51 Screeners: 80% 41/51 Clinical: 67% 34/51	N/A Screeners: 84% 43/51 Clinical: 75% 38/51	Expert: 80% Screeners: 83% 38/46 Clinical: 100% 46/46	N/A Screeners: 80% 37/46 Clinical: 100% 46/46	Expert is a radiologist with 30 yr. mammography experience, incl. 15 yr. mass screening. Screener is a radiologist with 5 yr. mammography experience, incl. 2 yr. mass screening. Clinical is a radiologist with 7 yr. experience in mammography but no experience in mass screening. Sensitivity calculated from 51 cases detected by at least 1 radiologist in earlier study. CAD sensitivity alone was 73% 37/51
Brem 2001 v. 1.0	Five radiologists: 1: 88 2: 98 3: 81 4: 88 5: 93 Avg±SD 89.6±6.3	Five radiologists: 1: 90 2: 98 3: 88 4: 90 5: 93 Avg±SD 91.8±3.9, p=n.s.	Five radiologists: 1: 50 2: 55 3: 61 4: 42 5: 33 Avg±SD 48.2±11.	Five radiologists: 1: 48 2: 53 3: 61 4: 41 5: 33 Avg±SD 47.2±10.8, p=n.s.	Only microcalcifications (MC) included in this analysis and masses were excluded.
Moberg 2001 v. 1.2 (n=59 cases and 211 controls)	Interval cancers Three radiologists: 1: 29 2: 20 3: 19 Avg 23	Interval cancers Three radiologists: 1: 27 2: 17 3: 22 Avg 22	Incorrect location selected in interval tumor cases Three radiologists: 1: 24 2: 8 3: 10 Healthy women 1: 73 2: 82 3: 89 Avg 81	Incorrect location selected in interval tumor cases Three radiologists: 1: 15 2: 2 3: 8 Healthy women 1: 78 2: 90 3: 92 Avg 86.7	The improvement in specificity observed with use of CAD results from using the CAD information outside the labeled instructions. Similarly, the reduction in sensitivity for readers 1 and 2 suggests improper use of CAD

For detection of malignant microcalcifications only, Brem et al. (2001) noted a small, nonsignificant improvement in sensitivity (89.6% vs. 91.8%, $p=n.s.$) and a small, nonsignificant decrement in specificity (48.2% vs. 47.2%, $p=n.s.$). Study radiologists were provided a 1-hour training session including normal cases, benign microcalcifications, and malignant microcalcifications prior to undertaking blinded interpretation of study cases, which may have favorably influenced their performance compared to general practice.

Moberg et al. (2001) found a slight but not significant decrease in sensitivity and increase in specificity associated with the use of CAD, but these results suggest that CAD was not used in this Swedish study according to the FDA-labeled instructions since radiologists changed some initially positive readings to negative readings when CAD information was taken into account. Thus, the results of this study do not seem applicable to the expected results of using CAD based on current recommended practice in the U.S.

In summary, these three studies provide some comparisons of radiologist sensitivity and specificity without CAD and with CAD and suggest that on average, in these types of case sets, adding CAD may not have a significant effect on sensitivity and specificity. However, the type of case-mix used in these studies and the artificial study setting may have introduced biases that limited the potential for CAD to significantly improve cancer detection or sensitivity rates raising concerns over the validity and applicability of these study findings.

There are no studies published in the medical literature using a commercially available CAD system on diagnostic, as opposed to screening, film-screen mammograms. Diagnostic mammograms are generally studies on patients who have either a symptomatic breast finding prompting mammographic evaluation or an abnormality on a prior mammogram that warrants further diagnostic evaluation. However, Data using the ImageChecker™ in diagnostic mammograms was submitted by R2 Technology, Inc. to the FDA in support of supplement 10 and was approved on May 29, 2001. These data are currently being prepared for publication in the peer-reviewed medical literature and are confidential at this time (personal communication R2 Technology, Inc.).

Summary of Evidence for Indication #1

The best available evidence to determine the effect of CAD on true-positive rate (i.e., cancer detection rate) and false-positive rate (e.g., recall or biopsy rate) in the screening mammography setting is provided by the Freer et al. (2001) study. This study employs a well-designed prospective protocol using CAD in a representative manner in the actual clinical screening setting for which it is intended. The results of this study suggest that adding CAD provides a clinically significant incremental improvement in cancer detection rate with an increase in recall and biopsy rate that is similar in proportion to the number of recalls and biopsies per cancer detected by radiologists reading alone without CAD.

The 3 additional Level 2 studies, comparing recall rates without and with CAD, provide additional substantiation that radiologists are able to discern the many false-positive marks prompted by CAD from those that require further work-up and that the use of CAD is unlikely to cause a very large increase in recall rate.

The Level 2 studies addressing the effect of CAD on true-positive rates provide some supportive evidence that CAD might have a significant effect in reducing the number of cancers that are missed at mammography. Four studies (5 reports) estimate that the use of CAD might lead to detecting 23-45% of cancers that are missed at mammography. Three other studies compare sensitivity and specificity of radiologists reading alone versus reading with CAD. However, these studies suffer from methodological weaknesses including spectrum bias in case selection and lack of representative radiologist interpretation environment due to the enriched case-mix and the retrospective study setting. While the results of these studies do not suggest that the addition of CAD results in a significant increase in sensitivity, it seems reasonable that the weaknesses in these retrospective studies may account for the apparent differences in the observed effect when results are compared with the Freer et al. study.

Even though double-reading is not a prevalent alternative practice in the U.S., it may serve as a useful comparison benchmark for CAD to consider that adding a second radiologist reader has been estimated to increase sensitivity in detecting cancer by 7-15% (Kopans 2000; Thurjell et al. 1994). Thus, the improvements

in sensitivity calculated from the Freer study associated with adding CAD (12–16%) are in the same range as those reported for adding a second radiologist reader. No published studies were identified that compared using CAD with double-reading.

Indication #2: Patients having Full-Field Digital Mammography

Full-field digital mammography (FFDM) is being studied as a potential alternative to film-screen mammography, and regulatory approval has been granted for at least one CAD system and one digital mammography system to be integrated. No published clinical effectiveness studies were identified using a commercially available CAD system applied directly to raw data derived from digital mammography.

The Technology Evaluation Center published a TEC Assessment in July 2002 reviewing the available evidence on the clinical effectiveness of full-field digital mammography. This analysis found insufficient evidence to determine that full-field digital mammography is at least as good as film-screen mammography. Furthermore, while not statistically significant, some of the available studies using FFDM in the screening setting showed concerning trends toward lower sensitivity for FFDM in detecting cancer as compared to film-screen mammography.

Given the uncertainty as to the diagnostic performance of FFDM relative to film-screen mammography, it is important to examine empirical evidence of the clinical effectiveness of CAD as applied in full-field digital mammography. Indeed, there are conceptual similarities between the application of CAD to a digitized film-screen mammogram and to a directly acquired digital mammogram, and it might seem reasonable to generalize some of the evidence using CAD in film-screen mammography to digital mammography. However, there are some important differences that will need to be evaluated with clinical studies to determine the clinical effectiveness of applying CAD directly to digital mammography. Unlike with digitized film-screen mammograms, there is more data in a raw digital mammogram than can be displayed 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.

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-assisted detection (CAD) devices after initial radiographic interpretation as a quality adjunct to single-reader mammography meets the Blue Cross and Blue Shield Association Technology Evaluation Center (TEC) criteria.

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

There are three manufacturers that have received FDA approval to market computer-aided detection (CAD) systems that take film-screen mammograms and digitize the images for computer analysis.

The first device to receive FDA approval, the ImageCheckerM1000® by R2 Technology, Inc. (Los Altos, CA), was approved by PMA (P970058) on June 26, 1998. The initial product labeling was for use on routine screening mammograms, but on May 29, 2001, approval was granted for the expansion of the “Indications for Use” to cover diagnostic as well as screening mammograms. The ImageCheckerM1000® is:

“intended to identify and mark regions of interest on routine screening and diagnostic mammograms to bring them to the attention of the radiologist after the initial reading has been completed. Thus, the system assists the radiologist in minimizing observational oversights by identifying areas on the original mammogram that may warrant a second review. (<http://www.fda.gov/cdrh/pma/pmamay01.html>; accessed July 15, 2002)

Multiple supplemental applications and approvals have also been processed relating to modifications of the ImageChecker hardware and software. A “New Efficacy Claim” was approved in PMA supplement 7 (version 2.2 software). The change is from:

“For every 100,000 cancers currently detected by screening mammography, the

use of the ImageChecker could result in early detection of an additional 30,500 breast cancers.”

to:

“Use of the ImageChecker could result in earlier detection of up to 25.4% (95% CI, 19.4%–27.4%) of the cancers currently detected with screening mammography in those women who had a prior screening mammogram 9-24 months earlier.” (<http://www.fda.gov/cdrh/pma/pmafeb02.html>; accessed July 15, 2002)

Two additional CAD systems have also been FDA approved. Second Look™ by CADx Medical Systems, Inc. (Northborough, MA) was approved by PMA (P010034) on January 31, 2002, and the MammoReader™ by Intelligent Systems Software, Inc. (Clearwater, FL) was approved by PMA (P010038) on January 15, 2002.

Each of these devices is:

“intended to identify and mark regions of interest on standard mammographic views to bring them to the attention of the radiologist after the initial reading has been completed. Thus, the system assists the radiologist in minimizing observational oversights by identifying areas on the original mammogram that may warrant a second review.” (<http://www.fda.gov/cdrh/pma/pmajan02.html>; accessed July 15, 2002) (<http://www.fda.gov/cdrh/pma/pmajan02.html>; accessed July 15, 2002)

In addition, General Electric Medical Systems (Milwaukee, WI) received FDA approval on April 12, 2002 to modify their Senographe 2000D® Full Field Digital Mammography System (P990066/S007) to permit integration and use of the ImageChecker® M1000-DM Computer Aided Detection (CAD) system manufactured by R2 Technology, Inc. (<http://www.fda.gov/cdrh/pma/pmaapr02.html>; accessed August 21, 2002)

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

Indication #1: CAD in Patients Having Film-Screen Mammography. The outcomes of primary interest in this Assessment are intermediate outcomes related to the diagnostic

results of mammography, true-positive rate (related measures include cancer detection rate or sensitivity) and false-positive rate (related measures include recall rate, biopsy rate, or specificity). While recall rate and biopsy rate are not actually measures of the false-positive rate per se, an increase in recall rate or biopsy rate represents a combination of the increase in false-positives and any increase in true-positives. In the screening setting, the proportion of false-positives reflected in the recall or biopsy rate is generally higher than the proportion of true-positives based on the very low prevalence of cancer. Thus, the terms “true-positive rate” and “false-positive rate” will be used to establish a conceptual framework within which the various indicators or measures of true-positives and false-positives reported in the studies can be synthesized and analyzed. Similar to diagnostic sensitivity and specificity, true-positive rate and false-positive rate are interrelated outcome measures, and it is important to consider both measures in evaluating the results of an individual study and in comparing results across studies. A better quality study would provide a complete assessment of both of these related and relevant outcomes.

This Assessment looked for studies that included study populations that were relevant to the proposed clinical use, and better quality studies would be those that included unselected consecutive cases in the routine screening or diagnostic setting. Studies that evaluate only a selected subset of cases, such as a subset of cancer cases missed during prospective screening, provide only indirect evidence of the potential effect of CAD in an unselected setting.

With regard to radiologist interpretation setting, radiologists may apply different diagnostic 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.

The systematic review of the literature yielded 11 studies that used computer-aided detection in distinct patient cohorts and met eligibility criteria. All of these studies used CAD in film-screen mammography. Seven of these studies (8 reports) used the ImageChecker® system (n=12,860; n=104; n=280; n=106; n=286; n=110; one study: n=14,817 CAD, n=25,682 no CAD;

n=120). Two eligible studies were identified for the Second Look™ system and both are derived from the FDA PMA data (n=3,946; n=177). Similarly, two FDA PMA data studies are available for the MammoReader™ (n=300; n=327).

The best available evidence to determine the effect of CAD on true-positive rate (e.g., cancer detection rate or sensitivity) and false-positive rate (e.g., as measured by recall or biopsy rate or specificity) in the screening mammography setting is provided by one study, and all other studies were rated of lower quality. The highest quality study employs a well-designed prospective protocol using CAD in a representative manner in the actual clinical screening setting for which it is intended. The results of this study suggest that adding CAD provides a clinically significant incremental improvement in cancer detection rate (19.5%) with an increase in recall and biopsy rate that is similar in proportion to the number of recalls and biopsies per cancer detected by radiologists reading alone without CAD.

Three other studies comparing recall rates without and with CAD provide additional substantiation that radiologists are able to discern the many false-positive marks prompted by CAD from those that require further work-up and that the use of CAD is unlikely to cause a very large increase in recall rate.

An additional four studies (five reports) conducted retrospectively in samples of cancers that had been missed on mammography provide some supportive evidence of the potential effectiveness of CAD. By virtue of their design, these studies do not provide direct evidence of the effect that CAD would have on cancer detection rates; however, they do provide indirect evidence of the potential effect that CAD might have had in improving earlier detection of these missed cancers had CAD been available. Furthermore, this population of cancers missed on mammography is precisely the type of cases where one would hope to have an impact by use of CAD. These studies estimate that the use of CAD might lead to detecting 23–45% of cancers that are missed at mammography.

Three other retrospective studies using selected samples of abnormal and normal mammograms compare sensitivity and specificity of radiologists reading alone versus reading with CAD. However, these studies suffer from methodological weaknesses including potential spectrum

bias in case selection and lack of representative radiologist interpretation environment due to the enriched case-mix and the retrospective study setting. While the results of these studies do not suggest that the addition of CAD results in a significant increase in sensitivity, it seems reasonable that the weaknesses in these retrospective studies may account for the apparent differences in the observed effect of adding CAD when results are compared with the Freer study.

In summary, the available evidence is considered sufficient 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 film-screen mammography.

Indication #2: CAD in Patients Having Full-Field Digital Mammography. Full-field digital mammography (FFDM) is being studied as a potential alternative to film-screen mammography, and regulatory approval has been granted for at least one CAD system and one digital mammography system to be integrated. No published clinical effectiveness studies were identified using a commercially available CAD system applied directly to raw data derived from digital mammography.

The Technology Evaluation Center published a TEC Assessment in July 2002 reviewing the available evidence on the clinical effectiveness of full-field digital mammography. This analysis found insufficient evidence to determine that full-field digital mammography is at least as good as film-screen mammography. Furthermore, while not statistically significant, some of the available studies using FFDM in the screening setting showed concerning trends toward lower sensitivity for FFDM in detecting cancer as compared to film-screen mammography.

Given the uncertainty as to the diagnostic performance of FFDM relative to film-screen mammography, it is important to examine empirical evidence of the clinical effectiveness of CAD as applied in full-field digital mammography. Indeed, there are conceptual similarities between the application of CAD to a digitized film-screen mammogram and to a directly acquired digital mammogram, and it might seem reasonable to generalize some of the evidence using CAD in film-screen mammography to digital mammography. However, there are some important differences that will need to be evaluated with clinical studies to

determine the clinical effectiveness of applying CAD directly to digital mammography. Unlike with digitized film-screen mammograms, there is more data in a raw digital mammogram than can be displayed 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.

In summary, the available evidence is considered to be 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.

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

Indication #1: CAD in Patients Having Film-Screen Mammography. CAD is intended as an adjunct to the radiologist's interpretation of a mammogram. Thus, the primary relevant comparison for this Assessment is whether the benefit of adding CAD (i.e., increase in true-positive rate) outweighs the harm of adding CAD (i.e., increase in false-positive rate). In some settings, CAD may be considered as an alternative to using a second independent radiologist interpretation (double-reading), however, this is not a common practice in the U.S.

In the highest quality study, 12,860 eligible screening mammograms were prospectively interpreted by two experienced community practice mammographers. Mammograms were analyzed by the ImageChecker M1000® v. 2.0 CAD system. Before reviewing the CAD prompts, the radiologist recorded an initial interpretation and a decision whether or not to recall the patient for additional evaluation. Then the CAD analysis was provided and the radiologist re-evaluated any areas prompted by the CAD system and recorded a second interpretation and decision. Clinical management was guided by the final reading including cases recommended for further evaluation based on CAD findings. The protocol did not permit the radiologist to reverse an initial decision to

recall the patient if CAD did not place a mark in the area of interest. Thus, the use of CAD in this study was purely as an aid to detection and CAD could only increase the cancer detection and recall rate.

Radiologists alone detected 41 cancers for a detection rate of 0.32% (41/12,860). 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. This included an average of 1.2 microcalcification marks and 1.6 mass or architectural distortion marks. The vast majority of CAD marks (97.4%) were dismissed by the radiologist but an additional 156 cases were recommended for recall. Thus, the end result of using CAD information on recall rate was an 18.8% relative increase from 6.5% to 7.7% of subjects.

The number of women actually recommended to have biopsy as a result of the radiologist interpretation alone was 107 and an additional 21 women were recommended to have a biopsy based on additional information provided by CAD. The positive biopsy yield was similar in both groups 38% (41/107) and 38% (8/21).

In comparing the trade-off between increased recall or biopsy rate and increased detection, an additional 156 women were recalled using CAD and an additional 21 women were biopsied in order to detect an additional 8 cancers, yielding 19.5 recalls and 2.6 biopsies for each additional case of cancer detected. It may be useful to compare these benchmarks to the results based on radiologists reading without CAD, and in that setting 830 recalls and 107 biopsies were recommended to detect 41 cancers, yielding 20.2 recalls and 2.6 biopsies per case of cancer detected.

Because this study did not employ an independent reference standard to ascertain the presence of any false-negative cases, sensitivity and specificity are not reported. However, if one makes an assumption about the level of sensitivity at which these study radiologists were reading screening mammograms without CAD, then sensitivity and specificity rates can be calculated from the data.

If one assumes that radiologists reading with CAD detected all cancers in the screening population (100% sensitivity with CAD), then radiologists reading alone in this study had a sensitivity of 84%, corresponding to a 16% absolute increase in sensitivity with the use of CAD. If one assumes instead that radiologists were reading at 79%, 69.5%, or 59% sensitivity levels without CAD, then the corresponding sensitivity levels achieved by adding CAD would be 94%, 83%, or 71%, respectively, for an absolute increase in sensitivity of 15%, 13.5% or 12% associated with using CAD. The relative increase in sensitivity was 19.5% at each of these levels of sensitivity. In all these hypothetical scenarios, the effect of CAD on specificity was minimal, with specificity falling from 94% to 93% by adding CAD.

Another group of studies including four retrospective studies (five reports) evaluate the ability of CAD to mark missed cancers and provide consistent evidence that CAD has the ability to mark a significant majority (62.7–77%) of cancers that were prospectively missed in clinical practice. However, not all prospectively missed cancers would meet a radiologist's criteria to be recommended for further work-up, so further analysis to determine the potential benefit of CAD are necessary. Based on blinded retrospective review by study radiologists, it was estimated that 23–45% of the actionable cancers (i.e., those that warranted further work-up) would have potentially been detected had CAD been used to point these lesions out to the radiologist.

The results of three additional retrospective studies comparing radiologist sensitivity and specificity without and with CAD in case samples artificially enriched with cancer cases do not suggest that the addition of CAD results in a significant increase in sensitivity. However, it seems reasonable that the weaknesses in these retrospective studies may account for the apparent differences in the observed effect when results are compared with the highest quality study.

Even though double-reading is not a prevalent practice in the U.S., it may serve as a useful comparison benchmark for CAD to consider that adding a second radiologist reader has been estimated to increase sensitivity in detecting cancer by 7–15%. Thus, the improvements in sensitivity associated with adding CAD (12–16%) calculated from the highest quality study in this Assessment are in the same range as those reported for adding a second radiologist reader. No full-text published studies were identified directly comparing CAD with double-reading.

In summary, the available evidence suggests that the use of CAD after initial radiographic interpretation as a quality adjunct to single-reader mammography in patients having film-screen mammography improves net health outcomes compared with single-reader radiologist interpretation by increasing 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).

Indication #2: CAD in Patients Having Full-Field Digital Mammography. There is insufficient evidence to permit conclusions on the effect on health outcomes of using CAD after initial radiographic interpretation as a quality adjunct to single-reader mammography in patients having full-field digital mammography.

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

Indication #1: CAD in Patients Having Film-Screen Mammography. The outcomes reported in the studies using CAD as an adjunct to the radiologist's interpretation in patients having film-screen mammography are derived from 11 studies using three commercially available CAD systems in various settings. The studies were conducted in a range of participating centers including community and referral centers and generally included radiologists who were experienced in mammography.

It is important to recognize that currently available CAD systems are intended to be used *only after* the radiologist has completed an evaluation of the images *without* CAD prompts and has made an initial decision whether any abnormal areas require recall of the patient for further work-up. Furthermore, currently available CAD systems should not be used to change

a radiologist's reading from positive to negative based on the absence of a CAD mark in an area of concern to the radiologist. It is possible that a CAD system could be used differently than it is intended to be used. In such a setting, radiologists using CAD before carefully reviewing the films without CAD prompts might be distracted by the CAD prompts and might miss some cancers they would otherwise have detected. In this way, the use of CAD might lead to a lower true-positive rate compared with radiologists reading without CAD.

Only one of the 11 studies reported data that would suggest that CAD information was used to change a radiologist's diagnosis from positive to negative, and that study was conducted in Sweden. Based on the overall body of evidence including all the studies conducted in the U.S., it is reasonable to expect that the outcomes observed in the investigational settings will be attainable outside the investigational setting with similarly experienced practitioners using CAD systems according to their intended use.

Indication #2: CAD in Patients Having Full-Field Digital Mammography. Whether the use of CAD as an adjunct to the radiologist's interpretation improves health outcomes in patients having full-field digital mammography has not been demonstrated in the investigational setting.

Therefore, based on the above, use of computer-assisted detection (CAD) after initial radiographic interpretation as a quality adjunct to single-reader mammography meets the Blue Cross and Blue Shield Technology Evaluation Center (TEC) criteria for patients having film-screen mammography. Use of CAD devices after initial radiographic interpretation as a quality adjunct to single-reader mammography for patients having full-field digital mammography 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

- Baker JA, Kornguth PJ, Lo JY et al. (1995).** Breast cancer: prediction with artificial neural network based on BI-RADS standardized lexicon. *Radiology*, 196(3):817-22.
- Balleyguier C, Malan S, Taourel P et al. (2001).** Computer-assisted Diagnosis (CAD) in Mammography: does it help the junior and the senior radiologist? Scientific Program of the Radiological Society of North America 87th Scientific Assembly and Annual Meeting, Abstract #1192, p. 520.
- Baum JK, Raza SR, Brem RF. (2001).** Prospective evaluation of computer-aided detection (CAD) in screening mammography-CAD impact on mammographic interpretation time and work-up rate. Scientific Program of the Radiological Society of North America 87th Scientific Assembly and Annual Meeting, Abstract #1195, p. 521.
- Beam CA, Layde PM, Sullivan DC. (1996).** Variability in the interpretation of screening mammograms by US radiologists – Findings from a national sample. *Arch Intern Med*, 156(2):209-15.
- Birdwell RL, Ikeda DM, O’Shaughnessy KF et al. (2001).** Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. *Radiology*, 219(1):192-202.
- Brem RF, Schoonjans JM. (2001).** Radiologist detection of microcalcifications with and without computer-aided detection: a comparative study. *Clin Radiol*, 56(2):150-4.
- Burhenne LJW, Wood SA, D’Orsi CJ et al. (2000).** Potential contribution of computer-aided detection to the sensitivity of screening mammography. *Radiology*, 215(2):554-62.
- CADx Medical Systems. (2000).** “CADx Signs Letter of Intent with LORAD for Distribution of Second Look Digital CAD Software.” CADx Medical Systems press release. Laval, Canada; November 23, 2000. Available at <http://www.cadxmed.com/press.asp?id=12>. Last accessed June 28, 2002.
- Castellino RA. (2002).** Computer-aided detection in oncologic imaging: screening mammography as a case study. *Cancer J*, 8(2):95-9.
- Castellino RA. (2000).** Computer-aided detection (CAD) in screening mammography. *Cancer Imaging*, 1(1):25-7.
- Chan HP, Sahiner B, Helvie MA et al. (1999).** Improvement of radiologists’ characterization of mammographic masses by using computer-aided diagnosis: an ROC study. *Radiology*, 212(3):817-27.
- Committee on the Early Detection of Breast Cancer, National Cancer Policy Board, Institute of Medicine. (2001).** Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer. National Academy Press; Washington, D.C. Also available online at: <http://search.nap.edu/books/0309072852/html/>.
- Cupples TE. (2001).** Impact of computer-aided detection (CAD) in a regional screening mammography program. Scientific Program of the Radiological Society of North America 87th Scientific Assembly and Annual Meeting, Abstract #1191, p. 520.
- Elmore JG, Wells CK, Lee CH et al. (1994).** Variability in radiologists’ interpretations of mammograms. *N Engl J Med*, 331(22):1495-9.
- Freer TW, Ulissey MJ. (2001).** Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. *Radiology*, 220(3):781-6.
- Garvican L, Field S. (2001).** A pilot evaluation of the R2 image checker system and users’ response in the detection of interval breast cancers on previous screening films. *Clin Radiol*, 56(10):855-7.
- GE Medical Systems. (2002).** “GE Medical Systems and R2 Technology Announce Computer Aided Detection for Digital Mammography Receives FDA Clearance.” GE Medical Systems press release. Waukesha, WI and Sunnyvale, CA; April 18, 2002. Available at www.gemedicalsystems.com/company/pressroom/releases/pr_release_6520.html. Last accessed on May 8, 2002.
- Jiang Y, Nishikawa RM, Schmidt RA et al. (2001).** Potential of computer-aided diagnosis to reduce variability in radiologists’ interpretations of mammograms depicting microcalcifications. *Radiology*, 220(3):787-94.
- Jiang Y, Nishikawa RM, Schmidt RA et al. (1999).** Improving breast cancer diagnosis with computer-aided diagnosis. *Acad Radiol*, 6(1):22-35.
- Kopans DB. (2000).** Double reading. *Radiol Clin North Am*, 38(4):719-24.
- Lechner M, Nelson M, Elvecrog E. (2002).** Comparison of two commercially available computer-aided detection (CAD) systems. *Appl Radiol*, 31(4):31-35.
- Leichter L, Fields S, Nirel R et al. (2000).** Improved mammographic interpretation of masses using computer-aided diagnosis. *Eur Radiol*, 10(2):377-85.

- Li L, Clark RA, Thomas JA. (2002). Computer-aided diagnosis of masses with full-field digital mammography. *Acad Radiol*, 9(1):4-12.
- Malich A, Azhari T, Bohm T et al. (2000). Reproducibility – an important factor determining the quality of computer aided detection (CAD) systems. *Eur J Radiol*, 36(5):170-4.
- Malich A, Marx C, Facius M et al. (2001a). Tumour detection rate of a new commercially available computer-aided detection system. *Eur Radiol*, 11(12):2454-9.
- Malich A, Marx C, Facius M et al. (2001b). Relation of histopathology and size to tumor detectability of a CAD-system in breast cancer detection. Scientific Program of the Radiological Society of North America 87th Scientific Assembly and Annual Meeting, abstract #996, p471.
- Mitka M. (2001). Some radiologists want more money up front. *JAMA*, 285(5):281-2.
- Moberg K, Bjurstam N, Wilczek B et al. (2001). Computed assisted detection of interval breast cancers. *Eur J Radiol*, 39(2):104-10.
- Nawano S, Murakami K, Moriyama N et al. (1999). Computer-aided diagnosis in full digital mammography. *Invest Radiol*, 34(4):510-6.
- Nakahara H, Namba K, Fukami A et al. (1998). Computer-aided diagnosis (CAD) for mammography: preliminary results. *Breast Cancer*; 25; 5(4):401-5.
- Nishikawa RM, Giger ML, Doi K et al. (1995). Computer-aided detection of clustered microcalcifications on digital mammograms. *Med Biol Eng Comput*, 33(2):174-8.
- Paquerault S, Petrick N, Chan HP et al. (2002). Improvement of computerized mass detection on mammograms: fusion of two-view information. *Med Phys*, 29(2):258-47.
- Qian W, Li L, Clarke L et al. (1999). Digital mammography: comparison of adaptive and nonadaptive CAD methods for mass detection. *Acad Radiol*, 6(8):471-80.
- Qian W, Sun X, Song D et al. (2001). Digital mammography: wavelet transform and Kalman-filtering neural network in mass segmentation and detection. *Acad Radiol*, 8(11):1074-82.
- Shile PE. (2001). Changes in workload with the use of computer-aided detection in mammography. *Appl Radiol*, 30(1):33-4.
- Thurfjell EL, Lernevall A, Taube AAS. (1994). Benefit of independent double reading in a population-based mammography screening program. *Radiology*; 191(1):241-4.
- Thurfjell E, Thurfjell MG, Egge E et al. (1998). Sensitivity and specificity of computer-assisted breast cancer detection in mammography screening. *Acta Radiol*, 39(4):384-8.
- Vyborny CJ, Doi T, O'Shaughnessy KF et al. (2000). Breast cancer: importance of spiculation in computer-aided detection. *Radiology*; 215(3):703-7.
- Worrell S, Mitchell R, Collins M et al. (2001). Plug-N-Play CAD. Advance for Imaging and Radiation Therapy Professionals, March 5.
- Zheng B, Ganott MA, Britton CA et al. (2001). Soft-copy mammographic readings with different computer-assisted detection cuing environments: preliminary findings. *Radiology*; 221(3):633-40.



**Technology
Evaluation
Center**

**Blue Cross and
Blue Shield Association**
225 North Michigan Avenue
Chicago, Illinois 60601-7680
888.852.4521
www.bcbs.com