



Australia and New Zealand Horizon Scanning Network

ANZHSN

AN INITIATIVE OF THE NATIONAL, STATE AND
TERRITORY GOVERNMENTS OF AUSTRALIA
AND THE GOVERNMENT OF NEW ZEALAND

National Horizon Scanning Unit

Horizon scanning report

Computer Aided Detection Systems in Mammography

July 2004



© Commonwealth of Australia 2005

ISBN 0 642 82618 8

ISSN

Publications Approval Number: 3605

This work is copyright. You may download, display, print and reproduce this material in unaltered form only (retaining this notice) for your personal, non-commercial use or use within your organisation. Apart from any use as permitted under the Copyright Act 1968, all other rights are reserved. Requests and inquiries concerning reproduction and rights should be addressed to Commonwealth Copyright Administration, Attorney General's Department, Robert Garran Offices, National Circuit, Canberra ACT 2600 or posted at <http://www.ag.gov.au/cca>

Electronic copies can be obtained from <http://www.horizonscanning.gov.au>

Enquiries about the content of the report should be directed to:

HealthPACT Secretariat
Department of Health and Ageing
MDP 106
GPO Box 9848
Canberra ACT 2606
AUSTRALIA

DISCLAIMER: This report is based on information available at the time of research and cannot be expected to cover any developments arising from subsequent improvements to health technologies. This report is based on a limited literature search and is not a definitive statement on the safety, effectiveness or cost-effectiveness of the health technology covered.

The Commonwealth does not guarantee the accuracy, currency or completeness of the information in this report. This report is not intended to be used as medical advice and it is not intended to be used to diagnose, treat, cure or prevent any disease, nor should it be used for therapeutic purposes or as a substitute for a health professional's advice. The Commonwealth does not accept any liability for any injury, loss or damage incurred by use of or reliance on the information.

The production of this *Horizon scanning report* was overseen by the Health Policy Advisory Committee on Technology (HealthPACT), a sub-committee of the Medical Services Advisory Committee (MSAC). HealthPACT comprises representatives from health departments in all states and territories, the Australia and New Zealand governments; MSAC and ASERNIP-S. The Australian Health Ministers' Advisory Council (AHMAC) supports HealthPACT through funding.

This *Horizon scanning report* was prepared by Dr Petra Bywood, Ms Skye Newton, Ms Tracy Merlin, Dr Annette Braunack-Mayer and Professor Janet Hiller from the National Horizon Scanning Unit, Adelaide Health Technology Assessment, Department of Public Health, Mail Drop 511, University of Adelaide, Adelaide, South Australia, 5005.

Table of Contents

Introduction	1
Background	1
Description of the technology.....	1
<i>The procedure</i>	2
<i>Intended purpose</i>	4
<i>Clinical need and burden of disease</i>	5
<i>Stage of development</i>	5
Treatment Alternatives	6
Existing comparators.....	6
Clinical Outcomes	7
Diagnostic accuracy.....	7
<i>CAD vs double or single reader mammography</i>	8
<i>CAD vs single reader mammography</i>	11
<i>Reproducibility of CAD</i>	15
<i>Sensitivity of CAD systems in detecting cancers</i>	16
Effectiveness.....	17
Safety.....	18
Potential Cost Impact	21
Ethical and Social Considerations	21
Training and Accreditation	23
Limitations of the Assessment	23
Sources of Further Information	28
Conclusions	28
HealthPACT Advisory	30
Appendix A	31
References	34

Tables

Table 1	Age-specific mortality rates for breast cancer in Australia	5
Table 2.	Mammogram examinations in Australia for the year 2001	7
Table 3.	CAD vs double reader mammography	10
Table 4.	CAD with single reader vs single reader mammography	14
Table 5.	Reproducibility of CAD systems.....	16
Table 6.	Sensitivity of CAD systems in detecting cancers.....	17
Table 7.	CAD-assisted radiologist vs single radiologist alone	18
Table 8	Safety: CAD-assisted radiologist vs radiologist alone	20
Table 9.	Literature sources utilised in assessment.....	25
Table 10.	Levels of evidence for assessing diagnosis	27
Table 11.	Designations of levels of evidence	27

Figures

Figure 1.	ImageChecker [®] CAD system	2
-----------	--	---

Introduction

The National Horizon Scanning Unit, Department of Public Health, University of Adelaide, on behalf of the Medical Services Advisory Committee (MSAC), has undertaken an Horizon Scanning Report to provide advice to the Health Policy Advisory Committee on Technology (Health PACT) on the introduction and use of computer-aided detection (CAD) for breast cancer screening.

This Horizon Scanning Report is intended for the use of health planners and policy makers. It provides an assessment of the current state of development of CAD for breast cancer screening, its present use, the potential future application of the technology, and its likely impact on the Australian health care system. This Horizon Scanning Report is a preliminary statement of the safety, effectiveness, cost-effectiveness and ethical considerations associated with CAD for breast cancer screening.

Background

Description of the technology

There are two relatively distinct categories of abnormalities that are characteristic of breast cancer: microcalcifications and masses.

Microcalcifications are small deposits of calcium that accumulate in breast tissue and appear white on a mammogram. Approximately 20-25% of breast cancers exhibit clusters of microcalcifications (3-5 microcalcifications/1cm²) in a mammogram (Roque & Andre 2002). Masses, which include architectural distortions with radiating lines, appear grey on a mammogram and are classified by their size, contour, density and shape – round, star-shaped (spiculated), or irregular.

Routine mammographic screening has been shown to reduce breast cancer-related deaths by 40-63% (Tabar et al 2001). Sensitivity of screening mammography to detect cancers ranges from 70-90%, meaning that 10-30% of breast cancers are missed at screening (Brem et al 2003). Efforts to enhance the sensitivity of mammography screening have included improvements in mammographic film and equipment, development of full-field digital mammography techniques, and intensive training to sharpen the interpretive skills of radiologists. Computer-aided (or assisted) detection (CAD) is a health service improvement tool designed to enhance the performance of radiologists by drawing attention to potential abnormalities on the mammogram that may have been overlooked by the reader.

The procedure

The first efforts to develop a CAD system for detecting breast cancer in mammograms were made by Winsberg (Malich et al 2001) approximately 37 years ago to automate the tedious task of examining large numbers of mammograms (Norris 2002). However, although some features of abnormality are readily detected, as their radiographic appearance is unambiguously different from normal mammographic structures (e.g. microcalcifications), others – particularly masses, asymmetries and architectural distortions – share many features of normal background tissue and appear grey on a mammogram. At present, mammographic screening cannot be fully automated as it requires the trained skills of a radiologist/mammogram reader to interpret suspicious areas on the images. Radiologists are instructed to examine a conventional x-ray film mammogram in the usual manner and, if abnormal patterns are discovered, to decide whether further action is required (e.g. additional views, biopsy).

Current CAD technology functions as an automated second expert reader, alerting the radiologist to potential trouble spots.

Currently, there are three commercially available CAD systems with Food and Drug Administration (FDA) approval for use in breast cancer screening. The R2 ImageChecker[®] (Figure 1, R2 Technology), Second Look[™] (CADx Medical Systems), and the MammoReader[™] (Intelligent Systems Software) systems process conventional x-ray film mammograms. The ImageChecker[®] technology has also been combined with Senographe digital mammography technology, which is used to a lesser degree, to process original full-field digital mammograms.



Figure 1. ImageChecker[®] CAD system (R2 Technologies, Inc.)

The principle of CAD is similar for each system, with three stages of breast cancer screening (Doi et al 1999; Roque & Andre 2002):

- 1) *Image processing* – A digital mammogram image is acquired directly or, more usually, by converting an original x-ray film to a lower-resolution

digital image using a computerised processor unit. The digitisation process uses techniques, such as Fourier analysis¹ for contrast enhancement to accentuate structures depicted on the mammogram, and edge detection to define the borders of enhanced structures. For example, the processing unit of the ImageChecker[®] CAD system links mammograms electronically by a bar code to the display unit, an autoviewer that comprises a pair of monitors displaying low spatial resolution digital images (Freer & Ullissey 2001).

- 2) 2) *Feature extraction and quantitation* – The image analysis software in the CAD system matches potential features of interest on the digital image, including size, shape, and contrast, to morphological features that are characteristic of malignancies. A number of data sets containing mammograms with a variety of representative breast cancers are available for this purpose (Roque & Andre 2002).
- 3) 3) *Discrimination of data* - CAD technology uses a range of mathematical and computational techniques, such as rule-based algorithms, discriminant analysis², decision trees, or artificial neural networks³ to discriminate between normal anatomical structures and abnormal features. For example, the detection algorithm of ImageChecker[®] recognises bright spots on the digital image as potential clusters of microcalcifications and marks these areas with a solid triangle (▲). For detecting masses, ImageChecker[®] uses an annulus - concentric circles of six and 32 mm – that passes over the mammogram and marks areas of density that have lines radiating from a common point and marks these regions with an asterisk (*). Based on a predetermined threshold of sensitivity, the image analysis software provides a binary outcome (prompted or not prompted) for each suspicious region. The common threshold of maximum false positive rates are set at 0.5 per image for masses and 0.2 per image for microcalcifications (Zheng et al 2002).

The principal differences between the CAD systems lie in the particular marking system used to indicate suspicious areas on the digitised image and in the output of results. While ImageChecker[®] uses triangles and asterisks to represent microcalcifications and densities, respectively; Second Look[™] uses rectangles and ellipses; and MammoReader[™] uses closed outlines and crosshairs. In addition, the marks used in the Second Look[™] and MammoReader[™] systems may be sized to denote the approximate size or extent of a potential lesion or microcalcification cluster, whereas ImageChecker[®] marks are placed in the centre of the lesion. Output of CAD results is generated electronically in ImageChecker[®] and MammoReader[™], whereas Second Look[™] produces a paper printout (Mammagraph[™]).

The recommended protocol for implementing CAD is that computer consultation occurs only *after* the radiologist has examined the original mammogram and made a decision. Thus, the unprompted cancer detection performance is maintained and the overall sensitivity of screening can only

¹ Fourier analysis: a mathematical technique for detecting the form of objects by processing spatial frequencies

² Discriminant analysis is a technique for classifying a set of observations into predefined classes

³ Artificial neural networks are computer models capable of “learning” relationships between input and output data and analysing patterns

increase. That is, after the CAD system has alerted the radiologist to potential suspicious regions, the radiologist re-examines the area on the original mammogram and decides whether to take further action. However, if an area previously identified by the radiologist fails to be detected by CAD, the decision to recall the patient should not be cancelled and the radiologist should act on their initial finding.

Intended purpose

Mammogram reading is a demanding task that involves a painstaking visual search for barely discernible signs of abnormalities that occur only rarely. Some of the inherent difficulties in reading mammograms that lead to reader variability include the following:

- 1) The complex and dense structure of breast tissue may obscure cancers, particularly small subareolar or deep retroglandular lesions;
- 2) Signs of malignancy may be very subtle, particularly for early stage masses and architectural distortions;
- 3) Radiologist characteristics, such as experience in mammogram reading and interpretive skills, may play a role; and
- 4) Limitations of human perception and reader fatigue may lead to missed cancers. Due to the low incidence of breast cancer, radiologists must read approximately 4000 images (2 views per breast) in 1000 women to detect 2-3 cancers (Beam et al 1996; Norris 2002).

Some studies suggest that up to 25% of cancers are not detected by screening mammography (Taft & Taylor 2001). The missed cancers fall into three categories:

- *Undetected* – not visible on the mammogram. Reducing undetected cancers in this category requires improvement in the sensitivity of equipment, such as enhanced capture and resolution of images.
- *Not interpreted* – visible but not detected by the radiologist. Retrospective analysis of mammograms has shown that 30-70% of confirmed cancers were visible on prior screens (Birdwell et al 2001). This is where CAD may play a role.
- *Misinterpreted* – detected but not actioned by radiologist. Reduction of misinterpreted abnormalities requires improvement in the training or re-training of radiologists and specialised mammogram readers.

Double reading, which is the routine procedure for BreastScreen Australia, is thought to overcome some of these problems and the accuracy of mammogram interpretation is shown to increase by up to 15% when two independent radiologists evaluate the mammograms (Malich et al 2001). However, this strategy is time and resource-intensive and is not necessarily used for screening or diagnostic purposes within private practices.

The advantage of computers is that they provide not only high throughput without fatigue or distraction, but also consistent performance - particularly when well-defined parameters are established. CAD is another way to reduce detection errors by prompting the radiologist to re-examine the mammogram in the location indicated by CAD marks and to decide if further work-up is

necessary. For example, according to the FDA recommendations, the ImageChecker[®] CAD system is “intended to identify and mark regions of interest on routine screening⁴ and diagnostic mammograms⁵ to bring them to the attention of the radiologist after the initial reading has been completed” (Blue Cross and Blue Shield Association 2002).

There are no contraindications for the use of CAD if the recommended protocol is followed.

Clinical need and burden of disease

In Australia, breast cancer is the most common cause of non-skin cancer-related mortality in females. There were 11,314 new cases of breast cancer and 2,521 deaths from breast cancer reported in Australia in 2000. The lifetime risk of developing breast cancer before the age of 75 years is approximately one in eleven Australian women (AIHW and AACR 2003; NHMRC 1999a) and the Australian screening incident rate is 4.5 cancers per 10,000. The age-specific mortality rates for breast cancer in Australia are shown in Table 1.

Table 1 Age-specific mortality rates^a for breast cancer in Australia, 1998-2001

Age group ^b													
20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84	85+
0.2	0.7	3.2	8.5	17.9	29.2	42.1	53.7	63.1	66.5	86.9	101.3	129.2	186.7

^a Rate per 100,000 women, ^b No mortality was recorded for women aged <20 years (AIHW and DHA 2003)

In New Zealand, the number of new female breast cancer registrations was 2,306 and the number of registered female breast cancer deaths was 622, for the year 2000. The age standardised incidence and mortality for breast cancer is 89.4 and 21.1 per 100,000 respectively (New Zealand Health Information Service⁶). New Zealand women have an approximate life time risk of one in ten of developing breast cancer. The risk is similar for Maori and non-Maori women (BreastScreen Aotearoa 2003).

The risk factors associated with breast cancer include increasing age, family history of breast cancer, age at first menstruation, and genetic factors (e.g. BRCA1, BRCA2, or Tp53) (NHMRC 1999a).

Stage of development

Although a group at the CSIRO started developing a CAD system for mammography in 1995, larger commercial organisations, such as R2 Technologies, demonstrated significant advances in this area and had greater capacity for commercialisation. Although no full CAD system has been developed in Australia, a number of Australian research groups continue to work on various aspects of CAD algorithms that may provide substantial

⁴ Screening mammogram: routine mammogram in asymptomatic woman, not undergoing evaluation for a specific finding

⁵ Diagnostic mammogram: mammogram obtained for patient with physical symptoms, or to evaluate a specific finding

⁶ New Zealand Health Information Service, Ministry of Health (2004), Wellington, New Zealand

improvements to detection rates and could be incorporated into existing CAD systems (personal communication, Flinders University of South Australia Cooperative Centre for Sensor Signal and Information Processing).

CAD is not in routine clinical use in Australia. Currently the Second Look™ CAD system is being trialled at St Vincent's Hospital for BreastScreen in Victoria (personal communication, BreastScreen St Vincent's).

Internationally, CAD is used in the context of research throughout the United States, Japan and several European countries, including Germany, The Netherlands, and Sweden. It is also in limited use in clinical settings in the United States. To January 2004, it is estimated that more than 1,200 ImageChecker® CAD systems have been shipped worldwide (R2 Technology).

Treatment Alternatives

Existing comparators

Despite recent debate over its usefulness in reducing breast cancer-related mortality (Humphrey et al 2002; Olsen & Gotzsche 2001), in Australia mammographic screening is considered the gold standard for the early detection of breast cancer and is purported to detect non-palpable masses and calcifications ($\leq 5\text{mm}$). During the screening process, a skilled technician positions the patient's breasts between two plates, which firmly compress the breast, flattening and pulling the breast tissue away from the chest wall. A routine mammographic examination takes approximately 20 minutes and includes two views of each breast taken from different angles, one from the side (medio-lateral oblique) and one from the top view (cranio-caudal). This results in a two-dimensional radiographic representation of each breast. Since mammograms do not distinguish between benign and malignant abnormalities, a fully qualified radiologist examines the mammogram for abnormalities, such as masses, calcifications, and architectural distortions. The initial mammogram serves as a baseline reference to enable the radiologist and clinicians to track any changes in the breast that may occur over time. A mammogram of normal breast tissue appears grey in areas of fat and white in areas of dense tissue. Calcifications also appear white on a mammogram and the radiologist must determine whether they deserve further investigation, such as a repeat mammogram, additional magnified x-rays, alternative imaging modalities (ultrasound or Magnetic Resonance Imaging), or a biopsy.

Typically, large calcifications (macrocalcifications), indicated by large white areas on the mammogram, are not associated with cancer, whereas clusters of bright spots suggest small calcifications (microcalcifications) that may indicate breast cell division, a hallmark of cancer. One of the most troublesome areas is architectural distortions - dense areas (with or without radiating lines) that appear grey on a mammogram. Abnormalities are easier to identify in older, post-menopausal women as their breasts have proportionally greater amounts of fat. They are less conspicuous in mammograms of those who take hormone replacement therapy, which results in denser breast tissue (Forrest & Anderson 1999; President and Fellows of Harvard College 2003).

Breast screening was introduced as a national program in Australia in 1991 and is known as BreastScreen Australia. The program aims to provide mammographic screening at two-year intervals for asymptomatic women aged 50-69 years, however women aged 40-49 and over 70 years of age are eligible to attend free of charge (Forrest & Anderson 1999; National Breast Cancer Centre 2002). Diagnostic mammographic examinations, which are conducted as needed on those who report pain, nipple discharge, or self-detected lumps, are available under the MBS (item numbers 59300 and 59303) for individuals not covered by BreastScreen Australia. The total number of mammogram examinations conducted in Australia during 2001 is shown in Table 2.

Table 2. Mammogram examinations, by age, in Australia for the year 2001 (number of females '000s)

Frequency of mammogram	Age (years)						Total
	18-29	30-39	40-49	50-59	60-69	70+	
Annually	9.9	28.8	144.4	227.2	118.0	72.1	600.5
More than 1, up to 2 years apart	2.4	18.4	195.3	510.3	363.7	214.7	1,304.8
More than 2 years apart		6.1	12.8	9.3	6.3	7.9	42.5
Not stated			4.6	13.6	6.7	9.3	34.3
Total	12.4	53.3	357.2	760.5	494.7	304.0	1,982.1

(AIHW and DHA 2003)

In 2000-2001, BreastScreen Australia screened 1,567,544 women as part of their ongoing program. Of these 1,063,479 (68%) were in their target group of 50-69 years of age. The participation rate for all women in Australia for this target group was 56.9 per cent (AIHW and DHA 2003). Current data suggests that screening 10,000 women aged 50-69 years of age, over 10 years, will prevent approximately 18 deaths, compared to preventing seven deaths in 10,000 women aged 40-49 years of age (National Breast Cancer Centre 2002).

The practice of “double-reading”, which is mandatory for BreastScreen Australia, involves two radiologists independently examining and interpreting each mammogram (Forrest & Anderson 1999; President and Fellows of Harvard College 2003). The accuracy of mammographic interpretation is reported to increase with double reading by 5-15% (Brem et al 2003). For example, double reading with arbitration has been shown to detect 18% more small (<10mm) invasive cancers compared to single reader (Blanks et al 1999).

Clinical Outcomes

Diagnostic accuracy

One systematic review and eleven studies were assessed in terms of diagnostic accuracy for this report. Seven studies used the ImageChecker[®], three used the

Second Look™ system, and one study evaluated a non-commercial CAD system.

Sensitivity and specificity are two parameters of diagnostic accuracy in breast cancer screening. Sensitivity relates to how well the mammogram interpretation performed in the detection of ALL potential abnormalities (no missed cancers - false negatives). For the CAD system, this entails the proportion of cancers detected and marked by the system. Specificity reflects the correct identification of normal breast tissue (no unnecessary biopsies – false positives). For CAD, false positives are the number of non-cancer regions that are marked as potentially suspicious. True positives are lesions, which are identified by screening mammogram, that prove to be malignant, whereas false negatives are lesions deemed to be negative that later prove to be malignant. Sensitivity is expressed as a ratio of true positive diagnoses to the total number of women with proven cancer in the screened population, whereas specificity is the rate of true negative diagnoses in women without cancer in the screened population.

CAD (with single or double reader) vs double or single reader mammography

In four diagnostic case-control studies (level 4 diagnostic evidence), radiologists reviewed mammograms in a dataset containing biopsy-proven cancers and normal (negative 2 years later) mammograms (Table 3). Comparisons were made between independent double reading and single radiologist reading, with or without CAD prompts. Three of these studies focussed on interval cancers and masses which, compared to microcalcifications, are generally more difficult to detect - both by radiologists and by a CAD system. One study, which used a National Proficiency dataset of mammograms for training readers, contained no “difficult” cancers (Ciatto et al 2003a).

One of these studies included a blinded review of mammograms seeded with prior negative screening mammograms of interval cancers (Ciatto et al 2003b). CAD-assisted reading significantly increased the sensitivity for detecting cancers compared to a single radiologist’s first reading of mammograms ($p < 0.01$) and was equivalent to simulated double reading ($p = 0.07$). Simulated independent double reading was contrived by combining pairs of single radiologist’s readings, whereby a case was deemed positive if either one or the other reader called it positive. Taylor et al (2004) found equivalent sensitivity for radiologists prompted by CAD compared to double reading with arbitration (Taylor et al 2004). The higher sensitivity with independent double reading in the Taylor et al study (80%) compared to the simulated double reading in the Ciatto et al study (46%) (Ciatto et al 2003b) may reflect a variety of factors including a more challenging case-mix in the Ciatto et al study and differences in the experience of the readers. This was evident in another study by Ciatto et al, which used a proficiency test dataset that contained no “difficult” cancers, such as interval cancers or those with subtle signs (Ciatto et al 2003a). Both highly experienced and inexperienced radiologists examined the mammograms. Although cancer detection improved for 6/10 readers, improvement was significant in only one reader who had poor initial performance (64.7 to 82.3%), and there was no statistically significant difference in cancer detection rates between a single radiologist alone, a radiologist with CAD assistance, or a

simulated double reading of the dataset. However, the high sensitivity of a radiologist alone (85.5%) may have created a ‘ceiling effect’ that masked potential improvement with CAD. Recall rates - where a woman was recalled for further diagnostic imaging or testing - were higher with CAD assistance than single reading ($p<0.01$) but were lower than for double reading ($p=0.04$) (Ciatto et al 2003b).

In another blinded review of prior mammograms to evaluate the use of a CAD system for detecting masses, independent double reading of mammograms, and single reading of mammograms, were compared to a simulated reading with CAD⁷ (Karssemeijer et al 2003). Although the simulated reading with CAD had significantly higher sensitivity compared to a single radiologist alone ($p<0.001$), it was significantly less sensitive compared to independent double reading ($p<0.001$). A major limitation of this study was that the simulated reading with CAD did not reflect usual practice. That is, if CAD marked locations that were *not* marked by a radiologist, they were ignored.

These studies have several key limitations. Retrospective interpretation of selected cases in an ‘experimental’ setting is not representative of the usual procedures that are recommended for CAD. The study datasets of mammograms were seeded with a higher prevalence of positive cases than would be seen in the clinical setting (spectrum bias), and radiologists may be prone to a ‘Hawthorne effect’ as their awareness of being tested is likely to increase their alertness. These factors are likely to bias diagnostic accuracy towards higher sensitivity. However, although results are not necessarily predictive of what would happen in current practice, both reading modalities were exposed to the same biases and comparisons may still be valid.

⁷ Simulated reading with CAD: combined findings of radiologists’ results and CAD markings. Only CAD marks that coincided with an area also marked by the radiologist were included and CAD marks for microcalcifications were ignored.

Table 3. CAD (with single or double reader) compared to double reader mammography

Study	Level of evidence	Study design	Population	Outcomes			
					Radiologist alone	Radiologist + CAD p value ^a	Independent double reading (simulated) p value ^b
Ciatto et al 2003b	4	Case-control study Retrospective analysis of prior mammograms 19 radiologists performed blinded review of mammograms CAD system: Second Look™	120 cases with 31 interval cancers and 89 negative control mammograms				
				Sensitivity	32.2%	42.1% p<0.01	46.1% p=0.07
				Recall rate	16.9%	23.9% p<0.01	26.1% p=0.04
Ciatto et al 2003a	4	Case-control study Retrospective analysis of prior mammograms 10 radiologists - 4 highly experienced in mammogram reading and 6 with no experience CAD system: Second Look™	Dataset of 150 original mammograms with 17 screen-detected cancers (National Proficiency Dataset for training mammogram readers)		Radiologist alone	Radiologist + CAD p value ^a	Independent double reading (simulated) p value ^b
				Cancer detection rate <i>Mean of 4 readings</i>	85.8%	90.0% NS 97.0%	– 96.0% NS
				Sensitivity, mean % ± SD	85.5%±10.5	90.0%±7.9	–
				Recall rate <i>Mean of 4 readings</i>	7.9%	11.4% p=0.003 10.7%	– 10.6% NS
Karssemeijer et al 2003*	4	Case-control study 10 radiologists performed blinded review of mammograms ImageChecker®	125 screen-detected cancers; 125 interval cancers; 250 negative controls		Radiologist alone	Radiologist + CAD (simulated) p value ^a	Independent double reading p value ^b
				Sensitivity	39.4%	46.4% p<0.001	49.9% p<0.001
Taylor et al 2004	4	Case-control study Retrospective analysis of prior mammograms 50 experienced film readers CAD system: ImageChecker®	180 cases: 120 negative controls; 40 cancers detected by routine screening; 20 interval cancers		CAD alone	Radiologist + CAD	Independent double reading
				Sensitivity, % [95% CI]	66.7%	78% [76, 80]	80% [76, 83]
				Specificity, % [95% CI]		84% [81, 87]	84% [81, 86]

* Authors have a vested interest in the company distributing the CAD system (e.g. employees or stockholders); ^a Radiologist + CAD vs radiologist alone; ^b Simulated double reading vs radiologist + CAD; Sensitivity = number of cancers detected/number of biopsy-proven cancers; Recall rate = number of patients recommended for additional imaging or biopsy/total number of patients screened; Cancer detection rate = number of cancers detected/number of patients screened; SD = standard deviation; Specificity = number of correctly identified negative mammograms/number of negative mammograms

CAD with single reader (or CAD alone) vs single reader mammography

One systematic review and five studies assessed the diagnostic accuracy of CAD-assisted mammography compared to single reader mammography alone (Table 4).

The highest quality study (level 1b diagnostic evidence) evaluated the performance of a non-commercial CAD system (Helvie et al 2004). Expert radiologists examined mammograms in the usual manner and recorded suspicious masses. Readers were required to select a BI-RADS category (shown in Box 1) for the mammogram before the CAD system displayed a digitised image marking areas of potentially suspicious masses. The radiologist reviewed the original mammogram in the location of the marks and made a second decision and BI-RADS assessment concerning the presence or absence of a potential mass. This process was repeated for microcalcifications.

Box 1. BI-RADS breast cancer assessment categories

Category	Assessment	Description
0	Incomplete	Needs additional imaging evaluation and/or prior mammograms for comparison
1	Negative	Breasts appear normal; no masses, microcalcifications, or architectural distortion present
2	Benign	Presence of non-malignant factors, such as calcified fibroadenomas, oil cysts, or lipomas
3	Probably benign	Presence of factors with <2% risk of malignancy – noncalcified circumscribed solid mass. Initial short-interval follow-up
4	Suspicious abnormality	Probability of malignancy. Biopsy recommended
5	Highly suspicious abnormality	Lesions with high probability ($\geq 95\%$) of malignancy. One-stage surgery without preliminary biopsy
6	Proven malignancy	Biopsy proof of malignancy prior to imaging

(American College of Radiology 2003)

The results (Table 4) indicate that non-palpable breast cancers were detected in eleven of 2389 patients. Radiologists or CAD alone correctly identified ten of the eleven cancers (91% [95% CI: 74, 100]). Radiologists alone missed one malignant microcalcification - a ductal carcinoma in situ (DCIS), while CAD alone missed one malignant mass - an invasive ductal carcinoma. With CAD as the second reader, all malignancies were detected. The use of CAD increased the recall rate from 14.4% to 15.8%. The higher sensitivity associated with CAD assistance resulted in slightly higher recall and biopsy rates, but this was counterbalanced by an improved cancer detection rate. This study was limited by the participation of highly specialised mammography radiologists and patient volunteers, which may not be representative of the typical radiologist or patient population.

Good quality evidence was also provided by a systematic review that evaluated the effectiveness of CAD or CAD assistance on cancer detection and recall rates in eleven studies (Blue Cross and Blue Shield Association 2002). This review evaluated one lower level (level 3b diagnostic evidence), although good quality, cohort study that employed a prospective protocol using CAD in the

recommended manner in a representative clinical screening setting (Freer & Ulissey 2001). The remaining included studies involved retrospective analysis of a selected subset of cancer cases that were missed during routine mammographic screening. Since appropriate proportions of normal cases were not included in the study datasets, these studies provide only indirect evidence of the potential effect of CAD in the usual clinical setting.

In the good quality level 3b diagnostic study (Table 4), almost 13,000 consecutive screening mammograms from patients in a community practice were interpreted initially by one experienced mammogram radiologist (without the assistance of CAD), followed immediately by re-examination of areas marked by CAD (Freer & Ulissey 2001). Initial decisions made without the assistance of CAD were not reversed if CAD failed to mark an area of interest. Thus, the cancer detection rate could not decrease following CAD analysis. Use of CAD resulted in detection of an additional eight cancers, including six ductal carcinomas in situ and two invasive ductal cell carcinomas. Radiologists alone (without CAD) detected 96% of cancer-associated masses and 68% of malignant microcalcifications. In contrast, CAD alone identified all of the malignant microcalcifications and slightly fewer (67%) cancerous masses. Overall, radiologists' cancer detection rate and recall rates increased by 19% with the aid of CAD.

The main limitation of this study was that follow-up data were not provided for patients who were not recalled – i.e. reference standards were not consistently applied to the study population (partial verification bias). Therefore, since outcomes for patients who were not recalled were not verified, the false negative rate is unknown. In addition, only two radiologists participated, making it difficult to generalise to radiologists with variable levels of skill and experience. Not unexpectedly, radiologists' performance on the first reading may be compromised by their anticipation of a second chance, acting as a 'safety net'. Therefore, any improvement in sensitivity following the second reading might not be entirely attributable to prompting by CAD. That is, a second look alone may improve accuracy. Double reading may, therefore, be a more appropriate reference standard.

This study also showed an increase in recall rate with CAD prompting, suggesting that radiologists are reticent to discount CAD marks without further investigation. This may change over time as radiologists learn the strengths and weaknesses of a particular CAD system and use it more optimally. Recall rates may also have been influenced by the lack of blinding. This was apparent in the Helvie et al study. Without CAD assistance, radiologists' recall rate was higher when they were aware that mammograms were part of the study population compared to those in the non-study population without CAD (13.5% vs 10.3%) (Helvie et al 2004). The difference in cancer detection rates between the Freer et al and Helvie et al studies may also reflect the differences in expertise and the volume of mammogram reading completed by the radiologists.

The remaining lower level studies (level 4 diagnostic evidence), which used enriched data sets including more cancers than would be encountered in the clinical setting, showed no significant improvement in cancer detection with CAD prompting (Brem & Schoonjans 2001; Moberg et al 2001). However, selection of cancer cases tended towards cancers that had already been detected

by radiologists and use of CAD would be less likely to demonstrate a significant improvement in cancer detection rate.

Further, in a substudy of case series, 19 cancers that were reported as normal on a previous mammogram and deemed retrospectively visible by two experienced radiologists were examined by two independent radiologists with and without the assistance of CAD (Taft & Taylor 2001). CAD detected two cases that were missed by both readers (both were invasive ductal carcinomas). However, since it is not certain whether radiologists would have requested follow-up if they had been alerted to the location by a CAD system, the validity of these results is unclear.

Several studies included in the systematic review provided substantiation that radiologists can discern the many false positive marks prompted by CAD from those marks requiring further workup, without an associated large increase in recall rates (Blue Cross and Blue Shield Association 2002). These studies provide some evidence that the addition of CAD does not have a significant effect on sensitivity and specificity. However, the case-mix (cancers that were detected readily by radiologists) and artificial setting used in these studies may have introduced biases that limit the potential for CAD to demonstrate improved sensitivity.

Table 4. CAD with single reader (or CAD alone) vs single reader mammography

Study	Level of evidence	Study design	Population	Outcomes			
					Radiologist alone	CAD alone	Radiologist + CAD
Helvie et al 2004	1b	Cross-classification of screening mammograms examined with and without CAD assistance CAD system: non-commercial (M-vision)	2,389 women aged 26-85 years (mean = 50.1±11.2) 13 qualified and experienced radiologists				
				Cancer detection rate	10/238 9 (0.42%)	10/238 9 (0.42%)	11/238 9 (0.46%)
				Sensitivity, % [95% CI]	10/11 91% [74, 100]	10/11 91% [74, 100]	11/11 100%
				Recall rate	4.4%	2.2%	15.8%
				Biopsy rate	1.6%	1.5%	1.7%
				PPV	27%		27.5%
Freer & Ulissey 2001	3b	Cross-classification of screening mammograms examined with and without CAD assistance Reference standards applied to actionable findings only (1026) CAD system: ImageChecker®	12,860 consecutive screening mammograms Two experienced radiologists		Radiologist alone	Radiologist + CAD	
				Cancer detection rate % [95% CI]	41/12860 0.32% [0.23, 0.43]	49/12860 0.38% [0.28, 0.50]	
				Recall rate, % [95% CI]	6.5% [6.0, 6.9]	7.7% [7.2, 8.1]	
				<i>Detection rate of 49 lesions</i>			
					Radiologist-detected	CAD-detected	
				Micro-calcifications	68.2%	100%	
				Masses	96.3%	66.7%	
Brem & Schoonjans 2001*	4	Case-control study Retrospective analysis of prior mammograms 5 radiologists performed blinded review of mammograms CAD system: ImageChecker®	106 cases: 24 normal; 40 benign micro-calcifications; 42 malignant micro-calcifications		Radiologist alone	Radiologist + CAD	
				Sensitivity, % ± SD	89.6% ±6.3	91.8% ±3.9	
				Specificity, % ± SD	48.2% ±11.0	47.2% ±10.8	
				CAD marks 41/42 malignant microcalcifications (98%) 32/40 benign microcalcifications (80%) 11/24 normal cases correctly identified as normal			
Moberg et al 2001	4	Case-control study 3 radiologists reviewed mixed set of mammograms CAD system: ImageChecker®	59 interval cancer cases mixed with 211 healthy controls, 5 screen-detected cancers, 5 benign lesions		Radiologist alone	Radiologist + CAD	
				Interval cancer sensitivity	A. 29% B. 20% C. 19%	A. 27% B. 17% C. 22%	
				Specificity	A. 73% B. 82% C. 89%	A. 78% B. 90% C. 92%	

Taft & Taylor 2001	4	Case series Study 2. Retrospective analysis of prior mammograms of proven cancers CAD system: Second Look™	19 cases of biopsy-proven cancers retrospectively visible on prior mammograms		CAD alone	Radiologist alone	Radiologist + CAD
				Sensitivity	63.2%	A. 42.1% B. 42.1%	68.4%

* Authors have a vested interest in the company distributing the CAD system (e.g. employees or stockholders); Cancer detection rate = number of cancers detected/number of patients screened; Sensitivity = number of cancers detected/number of biopsy-proven cancers; Recall rate = number of patients recommended for additional imaging or biopsy/total number of patients screened; Biopsy rate = number of patients receiving biopsy/number of patients screened; PPV = positive predictive value; Specificity = number of correctly identified negative mammograms/number of negative mammograms; SD = standard deviation

The systematic review (Blue Cross and Blue Shield Association 2002) also examined several retrospective studies conducted on samples of cancers that had been missed during routine mammographic screening. These studies focussed on the *expected* rather than the *actual* benefits of CAD (Birdwell et al 2001; Burhenne et al 2000; Garvican & Field 2001). By virtue of their design, these studies do not provide direct evidence of the impact of CAD on cancer detection rates. Taking only those marks that would meet radiologists' criteria for further workup into consideration, it was estimated that the use of CAD *may* lead to detecting 23-45% of cancers that are missed at mammography.

While the results from lower quality studies suggest that the addition of CAD does not increase sensitivity significantly, it appears that the methodological weaknesses may account for some of the differences in observed effects compared to that demonstrated in the better quality prospective cohort studies (Freer & Ulisse 2001; Helvie et al 2004).

Reproducibility of CAD

In the available systematic review (Blue Cross and Blue Shield Association 2002), one study showed that only 18% of cases were marked in the same location when scanned through the CAD system three separate times (Malich et al 2000), while other studies (FDA data) showed reproducibility rates ranging from 93-99%.

Two other case series examined the reproducibility of CAD prompts in proven cancer cases (Taylor et al 2003; Zheng et al 2003) (Table 5). Correct prompts were generated consistently in over 80% of screen-detected cancers and less than 50% of more subtle cancers (Taylor et al 2003). Both studies showed that prompts for microcalcifications were more reproducible than those for masses.

Table 5. Reproducibility of CAD systems

Study	Level of evidence	Study design	Population	Outcomes		
Taylor et al 2003	4	Case series Retrospective analysis of prior proven cancers CAD system: ImageChecker®	40 screen-detected or interval cancers digitised and analysed 10 times	<i>Correctly prompted by CAD</i>		
					Masses	Microcalcifications
				Screen-detected	86%	81%
				False negative	46%	21%
				Minimal signs	45%	50%
				Previously missed by CAD	19%	30%
				Total	53%	76%
37.5% prompted correctly 10 times 25.0% never prompted by CAD 37.5% prompted correctly 2-9 times						
Zheng et al 2003	4	Case series Retrospective analysis of prior proven cancers CAD system: ImageChecker®	100 cases biopsied: 65 malignant masses, 31 benign masses, 31 malignant microcalcifications, 19 benign microcalcifications	Cancer detection sensitivity Masses: 70% Microcalcifications: 89% <i>False positives:</i> Masses: 44% Microcalcifications: 32% 96% microcalcifications marked 3 times 67-71% masses marked 3 times		

Sensitivity of CAD systems in detecting cancers

CAD algorithms have been applied to mammography cases with known cancers to determine the sensitivity of the CAD system in marking known cancerous lesions. The analysis software is adjusted to balance the levels of sensitivity and specificity in an effort to achieve optimal detection rates. Overall, commercially available CAD systems mark 91-100% of malignant microcalcifications and 67-89% of cancerous masses (Blue Cross and Blue Shield Association 2002).

The sensitivity of the three commercially available systems, which were utilised in the studies examined in this review, is summarised in Table 6.

Table 6. Sensitivity of CAD systems in detecting cancers (Blue Cross and Blue Shield Association 2002)

CAD system	Study	Masses (%)	Microcalcifications (%)	Overall (%)
ImageChecker® v 2.0	Freer & Ulissey 2001	67	100	
	Vyborny et al 2000	86 (322/375) ^a 79 ^b		
	Burhenne et al 2000	75 (506/677)	99 (400/406)	84 (906/1083)
v. 1.2	Thurfjell et al 1998			73
	Nakahara et al 1998	79 (34/43)	100 (22/22)	86 (56/65)
v. 1.0	Brem & Schoonjans 2001		98	
MammoReader™	FDA PMA Data (465 cases)	87 (259/296)	91 (154/169)	89 (415/465)
	FDA PMA Data (327 cases)	79 (129/163)	92 (66/72)	83 (189/228)
Second Look™	Malich et al 2001	89 (110/124)	98 (55/56)	90 (135/150)
	FDA PMA ROSE-1D			85 (791/930)

^a for clearly spiculated masses; ^b for more loosely defined spiculation

Effectiveness

Evaluation of the effectiveness of CAD systems is limited by a shortage of good quality evidence and by the fact that the most important clinical outcome – overall survival – is not yet available in studies with long-term follow-up after screening with CAD. Therefore, the intermediate outcome, cancer detection rate, is reported.

One average quality before-and-after study (level 4 intervention evidence) was conducted in a representative, consecutive screening population in the clinical setting (Gur et al 2004). This study examined the recall and cancer detection rates before and after a CAD system was introduced into a large clinical breast imaging practice. Mammogram interpretations with the assistance of CAD were compared to those acquired during an 18-month period before CAD was installed. Results (Table 7) show that the introduction of CAD had no significant effect on cancer detection rate or recall rate in the practice.

The main limitation in this study is that the use of a historical control group for comparison does not consider secular changes or learning effects that may occur across time. For example, radiologists' interpretation of mammograms without CAD occurred chronologically prior to those with CAD. As a result, continuous efforts to improve radiologists' performance, such as reviews of false negatives and recall rates, are not taken into account. In addition, findings may have been influenced by the characteristics of the women screened. For example, in the latter part of the study (with CAD), more women returned for subsequent routine screen check-ups while the number of initial (or first time) screening examinations declined. Abnormalities in women screened for the first time are more likely to be detected at the initial screening and radiologists are less likely

to recommend recalls at subsequent screens in cases where comparison mammograms are available.

These results contrast with the significant improvement demonstrated in the Freer et al (2001) study (Table 4). To some extent, these results could be explained by the differences in methodology of the two studies. While radiologists reported on the same cases, with and without CAD in the same sitting in the Freer et al (2001) study, Gur et al (2004) compared cancer detection rates with and without CAD in different cases. Radiologists examining the same cases may experience lower levels of vigilance during the initial evaluation with the knowledge that they will get a second look at the same case with CAD assistance. In addition, radiologists in the Gur et al practice were highly experienced specialist mammogram readers who had little room for improvement in sensitivity (ceiling effect). In practices with relatively low recall rates or less experienced radiologists, CAD may have a more substantial effect.

Table 7. CAD-assisted radiologist vs single radiologist alone – effectiveness

Study	Level of evidence	Study design	Population	Outcomes		
Gur et al 2004	IV	Before-and-after study with historical controls CAD system: ImageChecker®	115,571 screening mammograms 24 highly experienced radiologists		Radiologist alone	Radiologist + CAD
				Cancer detection rate	197/56432 0.35%	210/59139 0.36%

Cancer detection rate = number of cancers detected/number of patients screened

Safety

The risk of breast cancer development associated with a standard mammogram is small. The use of ionising radiation limits the age of patients who can undergo a mammogram and the frequency with which mammograms can be used. The radiation dose used for a mammogram will depend on the breast size, thickness and density of the tissue (Warren 2001).

The obvious safety outcome of concern for CAD screening is the number of false positives and negatives reported. False positive findings may result in patients undergoing unnecessary biopsies or surgery. False negatives give false reassurance to patients that they are disease free and therefore may have serious consequences in terms of their future treatment.

Six studies provided data on safety (Table 8). One good quality cohort study reported no false negatives when CAD was included, but at the expense of slight (non-significant) increases in false positives, recall and biopsy rates (Helvie et al 2004). In an evaluation of the effect of CAD on the detection of interval cancers, Ciatto et al (2003) reported false negative rates with CAD that were similar to that seen with simulated double reading.

While recall and biopsy rates are not true measures of the false positive rate, an increase in the rate of recall or biopsy represents a combination of increased false positives and true positives. For example, Freer et al (2001) reported a

small increase in recall rate as a result of CAD prompts. Although radiologists dismissed 97.4% of CAD marks, an additional 156 cases were recommended for recall and 21 additional women were biopsied to detect eight extra cancers. This translates to 19.5 additional recalls and 2.6 additional biopsies for every additional cancer detected by CAD.

One recent lower quality case-control study (Level III-3 evidence) evaluated the rate of unnecessary follow-up procedures recommended by radiologists using the CAD system (Marx et al 2004). Five blinded radiologists assessed prior images of cases that included histologically proven cancers, benign lesions, and normal controls. Results demonstrated no statistically significant difference in sensitivity, specificity, positive or negative predictive values between interpretations by the radiologist alone or with CAD prompts. Authors reported slightly increased rates of short-term follow-up recommendations and lower rates of unnecessary biopsies. However, they fail to state whether these changes were statistically significant.

Table 8 Safety concerns related to CAD-assisted radiologist vs radiologist alone

Study	Level of evidence	Study design	Population	Outcomes																				
Marx et al 2004	III-3	Case control CAD system: Second Look™	185 cases: 36 cancers; 49 benign lesions; 100 negative controls	<p>CAD alone</p> <p>False positive 1.41 marks/image</p> <p>False negative 11.1%</p> <table border="0"> <tr> <td></td> <td style="text-align: center;">Radiologist alone</td> <td style="text-align: center;">Radiologist + CAD</td> </tr> <tr> <td>Short-term recall, % ±SD</td> <td style="text-align: center;">40.4±4.6</td> <td style="text-align: center;">42.6±4.4</td> </tr> <tr> <td>Biopsy rate, %±SD</td> <td style="text-align: center;">38.5±7.8</td> <td style="text-align: center;">30.2±10.1</td> </tr> </table>		Radiologist alone	Radiologist + CAD	Short-term recall, % ±SD	40.4±4.6	42.6±4.4	Biopsy rate, %±SD	38.5±7.8	30.2±10.1											
	Radiologist alone	Radiologist + CAD																						
Short-term recall, % ±SD	40.4±4.6	42.6±4.4																						
Biopsy rate, %±SD	38.5±7.8	30.2±10.1																						
Helvie et al 2004	IV	Case series CAD system: non-commercial system (M-vision)	2,389 consecutively screened women aged 26-85 years	<table border="0"> <tr> <td></td> <td style="text-align: center;">Radiologist alone</td> <td style="text-align: center;">CAD alone</td> <td style="text-align: center;">Radiologist + CAD</td> </tr> <tr> <td>False positive</td> <td style="text-align: center;">14.0%</td> <td style="text-align: center;">14.0%</td> <td style="text-align: center;">15.4%</td> </tr> <tr> <td>False negative</td> <td style="text-align: center;">9.1%</td> <td style="text-align: center;">9.1%</td> <td style="text-align: center;">0%</td> </tr> <tr> <td>Recall rate</td> <td style="text-align: center;">14.4%</td> <td style="text-align: center;">12.2%</td> <td style="text-align: center;">15.8%</td> </tr> <tr> <td>Biopsy rate</td> <td style="text-align: center;">1.6%</td> <td style="text-align: center;">1.5%</td> <td style="text-align: center;">1.7%</td> </tr> </table>		Radiologist alone	CAD alone	Radiologist + CAD	False positive	14.0%	14.0%	15.4%	False negative	9.1%	9.1%	0%	Recall rate	14.4%	12.2%	15.8%	Biopsy rate	1.6%	1.5%	1.7%
	Radiologist alone	CAD alone	Radiologist + CAD																					
False positive	14.0%	14.0%	15.4%																					
False negative	9.1%	9.1%	0%																					
Recall rate	14.4%	12.2%	15.8%																					
Biopsy rate	1.6%	1.5%	1.7%																					
Freer & Ulissey 2001	IV	Case series CAD system: ImageChecker®	12,860 consecutive screening mammograms	<table border="0"> <tr> <td></td> <td style="text-align: center;">Radiologist alone</td> <td style="text-align: center;">Radiologist + CAD</td> </tr> <tr> <td>Recall rate</td> <td style="text-align: center;">6.5% [6.0, 6.9]</td> <td style="text-align: center;">7.7% [7.2, 8.1]</td> </tr> <tr> <td>Recalls per additional cancer detected:</td> <td colspan="2" style="text-align: center;">19.5</td> </tr> <tr> <td>Biopsy rate</td> <td style="text-align: center;">0.8%</td> <td style="text-align: center;">1.0%</td> </tr> <tr> <td>Biopsies per additional cancer detected:</td> <td colspan="2" style="text-align: center;">2.6</td> </tr> </table> <p>CAD: 1.2 microcalcification marks/image; 1.6 mass or architectural distortion marks/image</p>		Radiologist alone	Radiologist + CAD	Recall rate	6.5% [6.0, 6.9]	7.7% [7.2, 8.1]	Recalls per additional cancer detected:	19.5		Biopsy rate	0.8%	1.0%	Biopsies per additional cancer detected:	2.6						
	Radiologist alone	Radiologist + CAD																						
Recall rate	6.5% [6.0, 6.9]	7.7% [7.2, 8.1]																						
Recalls per additional cancer detected:	19.5																							
Biopsy rate	0.8%	1.0%																						
Biopsies per additional cancer detected:	2.6																							
Ciatto et al 2003b	IV	Case series CAD system: Second Look™	120 cases with 31 interval cancers and 89 negative mammograms	<table border="0"> <tr> <td></td> <td style="text-align: center;">Radiologist alone</td> <td style="text-align: center;">Radiologist + CAD</td> <td style="text-align: center;">Double reading (simulated)</td> </tr> <tr> <td>False negative</td> <td style="text-align: center;">50.2%</td> <td style="text-align: center;">62.6%</td> <td style="text-align: center;">64.8%</td> </tr> <tr> <td>Minimal signs</td> <td style="text-align: center;">22.3%</td> <td style="text-align: center;">30.7%</td> <td style="text-align: center;">35.7%</td> </tr> </table>		Radiologist alone	Radiologist + CAD	Double reading (simulated)	False negative	50.2%	62.6%	64.8%	Minimal signs	22.3%	30.7%	35.7%								
	Radiologist alone	Radiologist + CAD	Double reading (simulated)																					
False negative	50.2%	62.6%	64.8%																					
Minimal signs	22.3%	30.7%	35.7%																					
Taft & Taylor 2001	IV	Case series CAD system: Second Look™	19 cases of biopsy-proven cancers retrospectively visible on prior mammograms	<table border="0"> <tr> <td></td> <td style="text-align: center;">Radiologist alone</td> <td style="text-align: center;">CAD alone</td> </tr> <tr> <td>False positive</td> <td style="text-align: center;">R1 47.4% R2 57.9%</td> <td style="text-align: center;">1.6 marks per image</td> </tr> </table> <p>CAD detected 2/19 cases missed by both readers (invasive ductal carcinomas)</p>		Radiologist alone	CAD alone	False positive	R1 47.4% R2 57.9%	1.6 marks per image														
	Radiologist alone	CAD alone																						
False positive	R1 47.4% R2 57.9%	1.6 marks per image																						
Zheng et al 2003	IV	Case series CAD system: ImageChecker®	100 cases biopsied: 65 malignant masses, 31 benign masses, 31 malignant microcalcifications, 19 benign microcalcifications	<table border="0"> <tr> <td></td> <td style="text-align: center;">Masses</td> <td style="text-align: center;">Microcalcifications</td> </tr> <tr> <td>False positive</td> <td style="text-align: center;">44%</td> <td style="text-align: center;">32%</td> </tr> </table> <p>96% microcalcifications marked 3 times 67-71% masses marked 3 times</p>		Masses	Microcalcifications	False positive	44%	32%														
	Masses	Microcalcifications																						
False positive	44%	32%																						

R1 = reader 1; R2 = reader 2; SD = standard deviation

Cost Analysis

There is currently no cost-effectiveness data available for utilising CAD for routine mammography screening for breast cancer in Australia. In the USA, the cost of CAD is funded, in part, through public and private health insurance companies. Medicare (USA) pays approximately \$US18.00 per CAD-assisted diagnostic and screening test (FDA, 2004). This may have influenced the rapid uptake of CAD systems into clinical practice.

The cost of a new CAD system is estimated to range from approximately \$US70,000 to \$US250,000 depending on the options and specifications, such as the volume of cases analysed per day. This cost is in addition to the cost of an existing film x-ray mammography system, or it can be integrated with digital mammography systems.

In Australia, the MBS fees for diagnostic mammography in item numbers 59300 and 59303 are \$82.00 and \$49.45 (MBS, 2004). Screening mammography is provided free of charge to women by BreastScreen Australia and is funded through joint Commonwealth and State/Territory agreements.

Although quality data on the effectiveness of CAD is lacking, CAD uptake in the USA has been rapid. This may be due to financial considerations as the US Medicare system reimbursement (\$15) covers more than the cost of CAD use, effectively paying for the installation of the system within a short period (Norris 2002).

Ethical and Social Considerations

Informed Consent

Women attending breast-screening clinics that utilise CAD systems must be informed of the “newness of the technology”, whether it is part of a research project, and how it differs from the usual practice (double reading). Accurate information about the predictive value of screening with CAD is essential, particularly with respect to the potential increase in false positives. To date, although positive predictive value (27%) of a single radiologist interpretation was equivalent to that with the addition of CAD (27.5%) (Helvie et al 2004), information about the positive and negative predictive values for CAD in the general population (compared to double reading) is currently unavailable. If sensitivity increases with CAD use, specificity and negative predictive values are expected to decrease. This means that women who screen negative for breast cancer may be *falsely* reassured that they are disease-free. Explaining negative and positive predictive values in language that is meaningful to patients can be problematic.

In addition, women may falsely believe that CAD technology provides a safer or more accurate result and become less vigilant about their own health – e.g. they may forgo breast self-examination or routine clinical breast examinations.

That is, since women tend to overestimate the risk reduction of screening anyway, the addition of computer technology may over-inflate their estimate of its effect.

Harms and Benefits

Current BreastScreen mammography programs have a well-established track record for acknowledging and managing women's fears and anxieties associated with a positive or equivocal result. From the available data, there is little evidence that the use of CAD improves the cancer detection rate compared to the double reading procedure that is current practice in Australia. A key area of concern is the high rate of false positive marks generated by the commercially available CAD systems and the potential increase in anxiety, radiation exposure, and morbidity associated with additional work-up and/or biopsy. Although most of the CAD marks (approximately 99%) tend to be dismissed by the reader, radiologists' practices are likely to be strongly influenced by the use of this technology. Computerised diagnostic technology does more than simply log data – it creates social and behavioural change. The nature and direction of this change, however, is not clear. If used correctly, CAD should not produce lower cancer detection rates (sensitivity). However, it has the potential to reduce sensitivity if it is used differently than in the manner intended, as radiologists may learn to rely on CAD and become less vigilant or thorough and less anxious about their decisions. Alternatively, as radiologists learn the limitations of the CAD system, their performance may improve. One study, which was limited by its small sample size, indicated that, despite higher false positive rates and increased numbers of short-term follow-up recommendations, the use of CAD led to a lower rate of unnecessary biopsies (Marx et al 2004).

With respect to the quality of radiologists' performance in interpreting mammograms, future radiologists may become less interested in mammography reading (as a career) if they expect computers to take over the task. This may impact on the requirements and training for the profession.

Failure to detect breast cancer is the leading cause of medical malpractice lawsuits in the USA (Kopans 2004). Therefore, there are potential medico-legal implications associated with retrospective analysis of malignancies that were marked by CAD and not followed up by the radiologist. For example, if CAD marked an area that the radiologist dismissed as negative or benign and a malignant lesion later developed in the same area, the delay in diagnosis may lead to claims of medical malpractice.

Access Issues

This technology is currently not available in large public and private hospitals in Australia. Due to the expense of acquisition of CAD technology, it is possible that they will only be purchased by large tertiary hospitals and would not be made available in rural areas of Australia.

Women may request access to CAD-assisted screening, basing their request on the experience of other women, and they may be more worried if their health service does not provide CAD. A strategy for addressing this issue needs to be in place if CAD-assisted screening is introduced.

Training

The Royal Australian and New Zealand College of Radiologists (RANZCR) conduct a training course for radiologists. For admission into the RANZCR training program candidates must be a graduate of a recognised medical school, be fully registered as a medical practitioner and have completed two full years in an approved hospital as an intern or resident. In order to be recognised as a Specialist in Radiodiagnosis and Fellow of the College (FRANZCR), the trainee must complete Parts I and II of the FRANZCR examinations in radiodiagnosis and complete a minimum of five years practical training positions accredited by the RANZCR. The training program aims to provide experience and training in general radiology, computed tomography (CT), nuclear medicine, ultrasound, MRI, angiography and basic interventional techniques. Currently, in order to be recognised as a specialist in nuclear medicine, a holder of the FRANZCR must complete two years of full time training in Nuclear Medicine in Joint Specialist Advisory Committee (JSAC) approved centres. Advanced training in women's imaging, including breast imaging, neonatal ultrasound, CT and MRI, is available at the Royal Women's Hospital, Queensland.

Currently the Quality and Accreditation Program of the RANZCR requires that radiologists reporting on mammography view a minimum of 480 mammograms each year and that they document 15 hours of continuing professional development in mammography every three years (RANZCR 2003).

Information on specific training for CAD systems is not available. For the studies described in this report, radiologists generally received a demonstration and/or a brief practice with a dataset of mammograms immediately prior to the study period. It is estimated that one hour of training in the use of the CAD system is sufficient for radiologists to familiarise themselves with its operation (personal communication, Flinders University of South Australia). However, it is unknown how much training is needed for radiologists to achieve optimal performance, whether reading time decreases with experience, or the potential learning effects of longer-term use of CAD.

Clinical Guidelines

In Australia, there are currently no clinical practice guidelines for breast cancer screening using CAD.

Limitations of the Assessment

Methodological issues and the relevance or currency of information provided over time are paramount in any assessment carried out in the early life of a technology.

Horizon Scanning forms an integral component of Health Technology Assessment. However, it is a specialised and quite distinct activity conducted for an entirely different purpose. The rapid evolution of technological advances

can in some cases overtake the speed at which trials or other reviews are conducted. In many cases, by the time a study or review has been completed, the technology may have evolved to a higher level leaving the technology under investigation obsolete and replaced.

An Horizon Scanning Report maintains a predictive or speculative focus, often based on low level evidence, and is aimed at informing policy and decision makers. It is not a definitive assessment of the safety, effectiveness, ethical considerations and cost effectiveness of a technology.

In the context of a rapidly evolving technology, an Horizon Scanning Report is a 'state of play' assessment that presents a trade-off between the value of early, uncertain information, versus the value of certain, but late information that may be of limited relevance to policy and decision makers.

This report provides an assessment of the current state of development of CAD systems for detecting breast cancer in mammography, its present and potential use in the Australian public health system, and future implications for the use of this technology.

Search strategy used for the Report

The medical literature (Table 9) was searched utilising the search terms outlined in Box 2 to identify relevant studies and systematic reviews, until June 2004. In addition, major international health assessment databases were searched.

Table 9. Literature sources utilised in assessment

Source	Location
<i>Electronic databases</i>	
AustHealth	University library
Australian Medical Index	University library
Australian Public Affairs Information Service (APAIS) - Health	University library
Cinahl	University library
Cochrane Library – including, Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, the Cochrane Central Register of Controlled Trials (CENTRAL), the Health Technology Assessment Database, the NHS Economic Evaluation Database	University library
Current Contents	University library
Embase	Personal subscription
Pre-Medline and Medline	University library
ProceedingsFirst	University library
PsycInfo	University library
Web of Science – Science Citation Index Expanded	University library
<i>Internet</i>	
Current Controlled Trials metaRegister	http://controlled-trials.com/
Health Technology Assessment international	http://www.htai.org
International Network for Agencies for Health Technology Assessment	http://www.inahta.org/
Medicines and Healthcare products Regulatory Agency (UK). <ul style="list-style-type: none"> • Medical Device Alert Safety warnings • Device evaluations • Diagnostic imaging review • Disability equipment assessments 	http://www.medical-devices.gov.uk/
National Library of Medicine Health Services/Technology Assessment Text	http://text.nlm.nih.gov/
National Library of Medicine Locator Plus database	http://locatorplus.gov
New York Academy of Medicine Grey Literature Report	http://www.nyam.org/library/greylit/index.shtml
Trip database	http://www.tripdatabase.com
U.K. National Research Register	http://www.update-software.com/National/
US Food and Drug Administration, Center for Devices and Radiological Health. <ul style="list-style-type: none"> • Manufacturer and User facility Device Experience (MAUDE). MAUDE data represents reports of adverse events involving medical devices. • PreMarket Approvals database 	http://www.fda.gov/cdrh/databases.html
Websites of Specialty Organisations	Dependent on technology topic area

Box 2. Search terms utilised

Search terms
MeSH Breast neoplasms
Text words Neoplasm*, cancer*, carcinoma*, tumour*, tumor*, breast*, computer-aided detection, computer-assisted detection, computer-aided diagnosis, computer-assisted diagnosis
Limits Human, English, female

Availability and Level of Evidence

One systematic review and eleven full-text studies for assessment of diagnostic accuracy were included in this report. All were classified according to the levels of evidence for assessing diagnosis (Table 10) (Phillips et al 2001). These included one level 1b, one level 3b, and nine level 4 studies.

One study was assessed for effectiveness and six studies (level IV evidence) provided data on safety considerations according to the designated levels of evidence described in Table 11 (NHMRC 1999b).

Table 10. Levels of evidence for assessing diagnosis^a

Level of evidence	Study design
1a	SR (with homogeneity*) of Level 1 diagnostic studies; CDR with 1b studies from different clinical centres
1b	Validating** cohort study with good† reference standards; or CDR tested within one clinical centre
1c	Absolute SpPins and SnNouts††
2a	SR (with homogeneity*) of Level ≥2 diagnostic studies
2b	Exploratory** cohort study with good† reference standards; CDR after derivation, or validated only on split-sample§ or databases
2c	n/a
3a	SR (with homogeneity*) of 3b and better studies
3b	Non-consecutive study; or without consistently applied reference standards
4	Case-control study, poor or non-independent reference standard
5	Expert opinion without explicit critical appraisal, or based on physiology, bench research or “first principles”

^a (Phillips et al 2001). SR = systematic review; CDR = clinical decision rule - these are algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category; RCT = randomised controlled trial; n/a = not applicable. * Homogeneity means a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. Studies displaying worrisome heterogeneity should be tagged with a “-” at the end of their designated level. ** Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the data (e.g. using a regression analysis) to find which factors are ‘significant’. † Good reference standards are independent of the test, and applied blindly or objectively to all patients. Poor reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the ‘test’ is included in the ‘reference’, or where the ‘testing’ affects the ‘reference’) implies a level 4 study. †† An “Absolute SpPin” is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An “Absolute SnNout” is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis. § Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into “derivation” and “validation” samples

Table 11. Designations of levels of evidence^a

Level of evidence	Study design
I	Evidence obtained from a systematic review of all relevant randomised controlled trials
II	Evidence obtained from at least one properly-designed randomised controlled trial
III-1	Evidence obtained from well-designed pseudorandomised controlled trials (alternate allocation or some other method)
III-2	Evidence obtained from comparative studies (including systematic reviews of such studies) with concurrent controls and allocation not randomised, cohort studies, case-control studies, or interrupted time series with a control group
III-3	Evidence obtained from comparative studies with historical control, two or more single arm studies, or interrupted time series without a parallel control group
IV	Evidence obtained from case series, either post-test or pre-test/post-test

^a Modified from (NHMRC 1999b)

Sources of Further Information

- Jennifer Cawson – currently evaluating the Second Look™ CAD system in BreastScreen Victoria
- Murk Bottema – currently researching CAD algorithms at Flinders University of South Australia Cooperative Centre for Sensor Signal and Information Processing

Conclusions

Breast cancer is the most common cause of non-skin cancer-related mortality in Australia and New Zealand and over 2,500 women die of breast cancer each year in Australia. Since there is no effective method to prevent breast cancer, early detection offers the best chance for reducing morbidity and mortality. The current recommendation is that, along with breast self-examination and clinical breast examination, women aged 50 or over undergo mammography screening every two years.

Mammogram reading, however, can be a laborious task – particularly considering the low incidence of the disease and the subtle nature of the early stages of cancer. It is estimated that routine mammographic screening misses approximately 10-30% of breast cancers. The introduction of CAD systems in mammography screening is underpinned by the belief that if CAD shows high prompt sensitivity to mammograms with cancers that were missed at a prior screening, then radiologists' performance will increase, thereby leading to earlier detection of cancer and (presumably) reduced disease-related morbidity and mortality. In practice, however, the problem is more complex.

There was substantial heterogeneity in the studies that were available for assessment for this report. Although double reader mammography is the usual practice for BreastScreen Australia, most of the available research investigated the effect of CAD as an adjunct to single-reader mammography.

All studies using independent double reading were retrospective analyses of mammograms with proven cancers, in which the CAD system was not used in the recommended manner, and the proportion of cancers was substantially higher than would be seen in a typical screening population. Overall, CAD-prompted radiologists showed equivalent or lower cancer detection sensitivity and recall rates compared to double reading mammography (including simulated double reading). The case-mix of datasets used for these studies varied substantially - a severe limitation for comparing the results of studies. Studies that used selected test sets with a high proportion of screen-detected cancers, and without normal controls, or with unrepresentative proportions of normal mammograms, ignore the continuum of normal-abnormal appearance of images seen in the usual screening population. Studies using unrepresentative

proportions of cases and small numbers of radiologists are prone to significant biases, limiting both internal and external validity. In contrast, datasets of unselected consecutive mammograms containing benign lesions, interval cancers, masses, architectural distortions, multiple abnormalities, or lesions with minimal signs on prior mammograms are more challenging for both readers and CAD.

The majority of studies, which compared the performance of a single radiologist alone to that of the CAD-assisted radiologist, also provided mixed results, depending on the study design. The better quality studies, which were conducted in more relevant settings, showed some improvement in radiologists' performance with CAD prompting. This was at the expense of slightly higher recall rates and may reflect radiologists' reticence to discount CAD-marked areas without obtaining additional views. Compared to a single radiologist reader, results indicate that CAD may be more effective in detecting microcalcifications and less effective in detecting masses and architectural distortions.

Since only one average quality study, which was conducted in a representative population, reported no change in cancer detection rates after the introduction of CAD, further good quality studies are necessary to fully evaluate the effectiveness of CAD.

The paradox of CAD is that, despite the high sensitivity, it may not influence the decision-making of readers (according to level IV intervention evidence). This is probably due to the low specificity of the CAD system. Radiologists must rely on their existing knowledge of cancer characterisation and their experience in reading mammograms to make an assessment of the probability of malignancy. Inaccurate prompting may act against the potential benefits of CAD in two aspects: 1) they create a distraction, drawing attention away from truly abnormal areas; and 2) readers, who become swamped by the potentially large number of false positives place less weight on the prompts. In contrast, if readers become reliant on prompts and less conscientious in their first unaided examination of a mammogram, they may miss cancers that are not detected easily by CAD, such as masses and architectural distortions. The interaction between individual mammogram readers and the CAD system is an undetermined variable that may vary according to the experience of the reader. It appears that highly experienced readers are able to readily distinguish and dismiss false positives without a concomitant increase in false negatives. In addition, experienced readers are less likely to show improved performance in cancer detection with CAD assistance, while those with initial poor performance have greater improvement with CAD.

Potentially high false positive prompts and poor detection of subtle abnormalities (false negatives) are important limitations of the CAD technology. CAD prompts are also subject to the interpretive skills of the mammogram readers – limiting its effectiveness.

CAD has the potential to alleviate some problems of high throughput – particularly if the screening population extends to wider age ranges. It also has the potential to enable less qualified/less experienced readers to function at a higher level.

Future improvements in the sensitivity of CAD prompts to masses and architectural distortions, and in the integration of information, such as both views of the breast and comparison with prior images, are likely to have a significant impact on the usefulness of this device. There is a clear need for proper large-scale prospective studies to be conducted with radiologists of variable experience and in representative screening populations. Better quality information on the accuracy of CAD assistance (in particular, specificity and positive or negative predictive values) in comparison to double-reading of mammography is essential to determine the potential benefits of the technology in Australia.

Appendix A

Profiles of the studies included for assessment for the diagnostic accuracy, safety and effectiveness of CAD systems in mammography for breast cancer.

Level of evidence	Study	Location	Study design	Study population	Outcome	Length of follow-up
	Blue Cross and Blue Shield Association 2002	Kaiser Permanente, USA	Systematic review containing level 3b and case-control studies	11 studies: 1 prospective cohort study and 10 retrospective case-control studies	Cancer detection rate Specificity Recall rates Biopsy rates False positive	Not stated
4	Brem & Schoonjans 2001	United States	Case-control study Blinded retrospective analysis of prior mammograms CAD system: Imagechecker®	106 cases: 24 normal; 40 benign microcalcifications; 42 malignant microcalcifications 5 experienced radiologists	Cancer detection rate Specificity	1 year
4	Ciatto et al 2003b	Tuscany, Italy	Case-control study Blinded retrospective analysis of prior mammograms CAD system: Second Look™	120 cases with 31 interval cancers and 89 negative mammograms 19 radiologists	Cancer detection rate Recall rate False positive False negative	1-2 weeks
4	Ciatto et al 2003)	Naples, Italy	Case-control study Blinded retrospective analysis of prior mammograms CAD system: Second Look™	Dataset of 150 original mammograms with 17 screen-detected cancers (National Proficiency Dataset for training mammogram readers) 10 radiologists - 4 highly experienced in mammogram reading and 6 with no experience	Cancer detection rate Recall rate	1-2 weeks
3b	Freer & Ulissey 2001	Plano, United States	Non-consecutive study – ref standards applied to actionable findings only (1026) CAD system: ImageChecker®	12,860 consecutive screening mammograms 2 experienced radiologists	Cancer detection rate	Not stated

IV	Gur et al 2004	Pittsburgh, United States	Before-and-after study with historical controls CAD system: ImageChecker®	115,571 consecutive screening mammograms 24 highly experienced radiologists	Cancer detection rate Recall rate	Not stated
1b	Helvie et al 2004	Michigan, United States	Cross-classification of screening mammograms examined with and without CAD assistance CAD system: non-commercial (M-vision)	2,389 consecutively screened women aged 26-85 years (mean = 50.1±11.2) 13 qualified and experienced radiologists	Cancer detection rate Recall rate Biopsy rate Positive predictive value False positive False negative	1 year
4	Karssemeijer et al 2003	Nationwide population study, The Netherlands	Case-control study Blinded retrospective analysis of prior mammograms CAD system: ImageChecker®	115 cases with masses at the location of the cancer that were visible on prior mammograms 10 radiologists (≥5 years experience)	Sensitivity	3-26 months
4	Marx et al 2004	Jena, Germany	Case-control study Blinded retrospective analysis of prior mammograms CAD system: Second Look™	185 cases: 36 cancers; 49 benign lesions; 100 negative controls 5 radiologists (6 months mammography training)	Sensitivity Specificity Positive predictive value Negative predictive value False positive marks False negative Short-term recall rate Biopsy rate	Not stated
4	Moberg et al 2001	Stockholm, Sweden	Case-control study Retrospective analysis of mixed sets of prior mammograms	59 interval cancer cases mixed with 211 healthy controls, 5 screen-detected cancers, 5 benign lesions 3 radiologists (5-10 years experience)	Interval cancer detection rate Specificity	Not stated
4 3b	Taft & Taylor 2001	Maroondah, Australia	Case series 1. Retrospective CAD analysis of proven cancers 2. Non-consecutive study CAD system: Second Look™	100 random cases of biopsy-proven cancers 19 cases of biopsy-proven cancers retrospectively visible on prior mammograms 2 experienced radiologists	True positive Sensitivity False negative False positive marks False positive	Not stated

4	Taylor et al 2003	London, United Kingdom	Case series Retrospective analysis of prior proven cancers CAD system: ImageChecker®	40 screen-detected or interval cancers digitised and analysed 10 times	Screen-detected cancers False-negative cancers Minimal sign cancers Previously missed by CAD % cancers correctly prompted by CAD Cancer detection rates	Not stated
4	Taylor et al 2004	London, United Kingdom	Case-control study Blinded retrospective analysis of prior mammograms CAD system: ImageChecker®	180 cases (120 negative controls; 40 cancers detected by routine screening; 20 interval cancers) 50 experienced film readers (30 radiologists, 5 breast clinicians, 15 radiographers)	Sensitivity Specificity % correct prompts screen-detected False prompts per case	Not stated
4	Zheng et al 2003	Pittsburgh, United States	Case series Retrospective analysis of prior proven cancers CAD system: ImageChecker®	100 cases biopsied: 65 malignant masses, 31 benign masses, 31 malignant microcalcifications, 19 benign microcalcifications	Sensitivity False positives	3 weeks

References

AIHW and AACR (2003). *Cancer in Australia 2000*, Australian Institute of Health and Welfare (AIHW) and Australasian Association of Cancer Registries (AACR), Canberra.

AIHW and DHA (2003). *BreastScreen Australia monitoring report 2000-2001*, Australian Institute of Health and Welfare (AIHW) and Australian Government Department of Health and Ageing, Canberra.

American College of Radiology (2003). *Breast Imaging Reporting and Data System Atlas: BI-RADS® Atlas*, 4th Edition ACR Publications.

Beam, C. A., Layde, P. M. & Sullivan, D. C. (1996). 'Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample', *Arch Intern Med*, 156 (2), 209-213.

Birdwell, R. L., Ikeda, D. M. 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.

Blanks, R. G., Wallis, M. G. & Given-Wilson, R. M. (1999). 'Observer variability in cancer detection during routine repeat (incident) mammographic screening in a study of two versus one view mammography', *J Med Screen*, 6 (3), 152-158.

Blue Cross and Blue Shield Association (2002). 'Computer-aided detection (CAD) in mammography', *TEC Bull (Online)*, 19 (3), 18-27.

BreastScreen Aotearoa (2003). *What causes breast cancer?* [Internet]. Available from: http://www.healthywomen.org.nz/BSA/bsa_causes.aspx [Accessed 20th May 2004].

Brem, R. F., Baum, J. et al (2003). 'Improvement in sensitivity of screening mammography with computer-aided detection: a multiinstitutional trial', *AJR Am J Roentgenol*, 181 (3), 687-693.

Brem, R. F. & Schoonjans, J. M. (2001). 'Radiologist detection of microcalcifications with and without computer-aided detection: a comparative study', *Clin Radiol*, 56 (2), 150-154.

Burhenne, L. J. W., Wood, S. A. et al (2000). 'Potential contribution of computer-aided detection to the sensitivity of screening mammography', *Radiology*, 215 (2), 554-562.

Ciatto, S., Del Turco, M. R. et al (2003a). 'Comparison of standard reading and computer aided detection (CAD) on a national proficiency test of screening mammography', *Eur J Radiol*, 45 (2), 135-138.

Ciatto, S., Rosselli Del Turco, M. et al (2003b). 'Comparison of standard and double reading and computer-aided detection (CAD) of interval cancers at prior negative screening mammograms: blind review', *Br J Cancer*, 89 (9), 1645-1649.

Doi, K., MacMahon, H. et al (1999). 'Computer-aided diagnosis in radiology: potential and pitfalls', *Eur J Radiol*, 31 (2), 97-109.

- Forrest, A. P. & Anderson, E. D. (1999). 'Breast cancer screening and management', *MJA*, 171 (9), 479-484.
- Freer, T. W. & Ulissey, M. J. (2001). 'Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center', *Radiology*, 220 (3), 781-786.
- 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), 833-837.
- Gur, D., Sumkin, J. H. et al (2004). 'Changes in breast cancer detection and mammography recall rates after the introduction of a computer-aided detection system', *J Natl Cancer Inst*, 96 (3), 185-190.
- Helvie, M. A., Hadjiiski, L. et al (2004). 'Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection: Pilot clinical trial', *Radiology*, 231 (1), 208-214.
- Humphrey, L. L., Helfand, M. et al (2002). 'Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force', *Ann Intern Med*, 137 (5 Part 1), 347-360.
- Karssemeijer, N., Otten, J. D. M. et al (2003). 'Computer-aided detection versus independent double reading of masses on mammograms', *Radiology*, 227 (1), 192-200.
- Kopans, D. B. (2004). 'Mammography screening is saving thousands of lives, but will it survive medical malpractice?' *Radiology*, 230 (1), 20-24.
- Malich, A., Azhari, T. et al (2000). 'Reproducibility- an important factor determining the quality of computer-aided detection (CAD) systems', *Eur J Radiol*, 36 (3), 170-174.
- Malich, A., Marx, C. et al (2001). 'Tumour detection rate of a new commercially available computer-aided detection system', *Eur Radiol*, 11 (12), 2454-2459.
- Marx, C., Malich, A. et al (2004). 'Are unnecessary follow-up procedures induced by computer-aided diagnosis (CAD) in mammography? Comparison of mammographic diagnosis with and without use of CAD', *Eur J Radiol*, 51 (1), 66-72.
- Moberg, K., Bjurstam, N. et al (2001). 'Computed assisted detection of interval breast cancers', *Eur J Radiol*, 39 (2), 104-110.
- Nakahara, H., Namba, K. et al (1998). 'Computer-Aided Diagnosis (CAD) for Mammography: Preliminary Results', *Breast Cancer*, 5 (4), 401-405.
- National Breast Cancer Centre (2002). *Breast imaging: a guide for practice*, National Breast Cancer Centre, Sydney, NSW.
- NHMRC (1999a). *Familial aspects of cancer: a guide to clinical practice*, National Health and Medical Research Council (NHMRC), Canberra, ACT.
- NHMRC (1999b). *A guide to the development, implementation and evaluation of clinical practice guidelines*, National Health and Medical Research Council, Commonwealth of Australia, Canberra, ACT.

- Norris, T. G. (2002). 'Computer-aided detection in mammography', *Radiologic Technology*, 73 (3), 225-242.
- Olsen, O. & Gotzsche, P. C. (2001). 'Screening for breast cancer with mammography', *Cochrane Database Syst Rev*, (4), CD001877.
- Phillips, B., Ball, C. et al (2001). *Oxford Centre for Evidence-Based Medicine Levels of Evidence (May 2001)* [Internet]. Centre for Evidence-Based Medicine. Available from: http://www.cebm.net/levels_of_evidence.asp [Accessed 28th January 2004 2004].
- President and Fellows of Harvard College (2003). 'Update on breast imaging. New imaging techniques have led to advances in breast cancer detection. What does that mean for the annual mammogram?' *Harvard Womens Health Watch*, 10 (10), 4-6.
- RANZCR (2003). *Radiodiagnosis handbook*, Royal Australian and New Zealand College of Radiologists, Sydney, NSW.
- Roque, A. C. & Andre, T. C. (2002). 'Mammography and computerized decision systems: a review', *Ann N Y Acad Sci*, 980, 83-94.
- Tabar, L., Vitak, B. et al (2001). 'Beyond randomized controlled trials: organized mammographic screening substantially reduces breast carcinoma mortality', *Cancer*, 91 (9), 1724-1731.
- Taft, R. & Taylor, A. (2001). 'Potential improvement in breast cancer detection with a novel computer-aided detection system', *Appl Radiol*, 30 (12), 25-28.
- Taylor, C. G., Champness, J. et al (2003). 'Reproducibility of prompts in computer-aided detection (CAD) of breast cancer', *Clin Radiol*, 58 (9), 733-738.
- Taylor, P. M., Champness, J. et al (2004). 'An evaluation of the impact of computer-based prompts on screen readers' interpretation of mammograms', *Br J Radiol*, 77 (913), 21-27.
- Thurfjell, E., Thurfjell, M. G. et al (1998). 'Sensitivity and specificity of computer-assisted breast cancer detection in mammography screening', *Acta Radiol*, 39 (4), 384-388.
- Vyborny, C. J., Doi, T. et al (2000). 'Breast cancer: importance of spiculation in computer-aided detection', *Radiology*, 215 (3), 703-707.
- Warren, R. (2001). 'Screening women at high risk of breast cancer on the basis of evidence', *Eur J Radiol*, 39 (1), 50-59.
- Zheng, B., Hardesty, L. A. et al (2003). 'Mammography with computer-aided detection: Reproducibility assessment - Initial experience', *Radiology*, 228 (1), 58-62.
- Zheng, B., Shah, R. et al (2002). 'Computer-aided detection in mammography: an assessment of performance on current and prior images', *Acad Radiol*, 9 (11), 1245-1250.