Digital mammography for breast cancer screening, surveillance and diagnosis

November 2007

MSAC reference 37

Assessment report

ISBN 1-74186-547-6
Online ISBN 1-74186-548-4
ISSN (Print) 1443-7120

ISSN (Online) 1443-7139
First printed July 2008

Electronic copies of the report can be obtained from the Medical Service Advisory Committee's internet site at http://www.msac.gov.au/

Printed copies of the report can be obtained from:

The Secretary
Medical Services Advisory Committee
Department of Health and Ageing
Mail Drop 106
GPO Box 9848
Canberra ACT 2601

Enquiries about the content of the report should be directed to the above address.

The Medical Services Advisory Committee (MSAC) is an independent committee which has been established to provide advice to the Minister for Health and Ageing on the strength of evidence available on new and existing medical technologies and procedures in terms of their safety, effectiveness and cost-effectiveness. This advice will help to inform government decisions about which medical services should attract funding under Medicare.

MSAC recommendations do not necessarily reflect the views of all individuals who participated in the MSAC evaluation.

This report was prepared for the Medical Services Advisory Committee by Ms Tracy Merlin, Ms Hedyeh Hedayati, Ms Shuhong Wang, Mr Thomas Sullivan, Ms Skye Newton, Ms Liz Buckley, Mr Florian Kreisz, Ms Christina Zimprich, and Professor Janet Hiller from Adelaide Health Technology Assessment, Discipline of Public Health, University of Adelaide. Our thanks to Linda Mundy for identifying TGA-listed units. The report was edited by Matthew Stevens, ScienceScape Editing, Sydney. This recommendation was endorsed by the Minister for Health and Ageing on 11 April 2008. Publication approval number: P3–3487

Contents

Tables

Figures

Executive summary

Rationale for assessment

Film-screen mammography (FM) is currently funded by the joint Australian, state and territory governments' BreastScreen Australia Program as a population-based screening test targeted at women aged 50–69 years. It is also subsidised under the Australian Government's Medicare Benefits Scheme as a diagnostic test for individuals with symptoms or a previous history of breast cancer, or at potentially high risk of breast cancer due to family history.

Government and stakeholders recognise that new digital technology is likely to replace FM owing to marketing decisions by manufacturers and to workforce issues. Inevitable changes in technology availability, including the forced redundancy of conventional mammography technology, are becoming apparent, as is the need for new technology to assist with an Australian shortage of radiologists and radiographers. Thus, digital mammography (DM) for the screening, surveillance and diagnosis of breast cancer was referred for assessment by the Medical Services Advisory Committee (MSAC) by the Australian Government's Department of Health and Ageing. An assessment of the comparative safety, effectiveness and cost-effectiveness of this new technology is a matter of some urgency.

The procedure

DM allows the processes of image acquisition, processing and display to be managed (and optimised) more separately than is possible in conventional FM. In FM, an image of the breast is generated by converting the energy from the x-ray beam (which has passed through the breast) into light via a phosphor screen, which then exposes a film (hard copy) held within the screen. The film is processed chemically and an image is developed. DM also uses x-rays to produce images of the breast tissue, but the x-ray photons encounter a solid-state detector that converts the absorbed energy into electrical signals. These electrical signals are used to produce images of the breast that can be displayed on a computer monitor (soft copy) or be printed on laser film (hard copy). DM overcomes the need for physical co-location of the image acquisition, processing and image display processes associated with FM, and allows the three to be physically and operationally separated. This feature permits additional functionality, including telemammography, which allows the transmission of mammographic images from a remote site to a radiologist for real-time assessment and interpretation.

Medical Services Advisory Committee – role and approach

The MSAC was established by the Australian Government to strengthen the role of evidence in health financing decisions in Australia. MSAC advises the Minister for Health and Ageing on the evidence relating to the safety, effectiveness and cost-effectiveness of new and existing medical technologies and procedures, and under what circumstances public funding should be supported.

A rigorous assessment of evidence is thus the basis of decision-making when funding is sought under Medicare. A team from Adelaide Health Technology Assessment (AHTA), Discipline of Public Health, University of Adelaide, was engaged to conduct a systematic review of literature

on DM for the screening of asymptomatic women, surveillance of women at potentially high risk of breast cancer, and diagnosis of breast cancer in symptomatic women. An advisory panel with expertise in this area then evaluated the evidence and provided advice to the MSAC.

MSAC's assessment of digital mammography

Clinical need

Breast cancer is the most frequently diagnosed cancer among women in Australia, as evidenced by 11 788 new cases in 2003. Australia has the fifth highest rate of breast cancer incidence in the world: one in 11 Australian women will develop breast cancer before reaching 75 years of age. The incidence of breast cancer in women rose from 105.3 cases per 100 000 population in 1993 to 111.8 cases per 100 000 in 2003 (AIHW 2007b).

Breast cancer is the most common cause of cancer-related death in Australian women, accounting for 2641 deaths in 2004 (AIHW 2007a). However, international comparisons of developed countries show that Australia has low mortality rates, indicating good survival rates, through early detection and treatment strategies (NBCC 2006). A national breast cancer screening program was introduced in 1991 to provide free film-screening mammograms biennially for women 50–69 years old without breast cancer symptoms or signs. Asymptomatic women who are in their 40s or 70 years and older are also offered free screening mammograms, although they are not targeted for invitation and recall.

Safety

Three studies have compared radiation exposure levels between DM and FM. They suggest that radiation exposure is very similar between the two. However, phantom studies using more recently developed mammography units suggest that the radiation dose from DM is likely to be lower than that from FM. It can therefore be concluded that DM is at least as safe as FM in regards to radiation dose.

The best-quality evidence on false positive rates suggests that DM may produce slightly more unnecessary further investigations than FM. The statistical significance or clinical relevance of this difference could not be determined from the data provided, and these results were contrasted by a second large screening trial that reported lower false positive rates from DM. In conjunction with the diagnostic accuracy data provided for screening effectiveness, the data suggest that the false positive-rates of both methods are similar. False positive rates are likely to depend on familiarity and experience with interpreting digital mammograms.

Overall, DM appears to be as safe as FM.

Effectiveness

Screening (asymptomatic women)

The good quality evidence indicates that DM is as accurate as FM in screening asymptomatic women. The case for replacing FM is not obvious overall, given the similar accuracy and cancer detection rates of the two methods. DM would, however, appear to be a reasonable alternative to FM on the basis of the population-based effectiveness data alone. In the largest good quality

screening trial to date (Digital Mammography Imaging Screening Trial: DMIST), DM was found to be more accurate at detecting breast cancer in women who are conventionally difficult to image with FM, specifically women aged under 50 years, those who are pre- or perimenopausal, or those with heterogeneously dense or extremely dense breast tissue. As such, DM should replace FM for these women.

The relative impact of DM and FM on interval cancer rates has yet to be properly established.

Given the similar diagnostic accuracy of DM and FM in an asymptomatic population, it is likely that the health benefits from FM versus no screening would apply to a population screened with DM.

If DM were used for screening asymptomatic women aged under 50 years, women who are pre- or perimenopausal, or women with radiographically dense breasts, it is likely that the mortality reduction from screening would be higher in these subgroups than with FM.

Surveillance (women at potentially high risk) and diagnosis (symptomatic women)

For the diagnosis of women at potentially high risk or symptomatic women, it is unclear whether DM is as effective as FM. Although no significant differences in the diagnostic properties of the two methods have been reported in the literature, it is difficult to make any strong conclusions from the available evidence. However, it appears likely that as women at high risk present for surveillance at a younger age than the general population, and thus are more likely to be pre- or perimenopausal with dense breasts, the improved diagnostic accuracy of DM in these subgroups in the DMIST study would be applicable – probably leading to increased benefits over FM, given that these women are starting at a higher baseline risk for breast cancer than the general population.

There is currently no evidence to suggest that DM should be used in addition to FM in either a symptomatic or a surveillance population.

Benefits from breast cancer treatment for symptomatic women are well known and should remain the same if DM replaces or is used as an alternative to FM for diagnosis – assuming that similar diagnostic accuracy between the two methods is confirmed by a larger evidence base in a diagnostic population.

Other relevant considerations

One of the major reasons DM is currently being considered as an option is the shortage of qualified radiologists and radiographers within Australia. This, in conjunction with the increasing demand for mammography, has created an imbalance between supply and demand. DM could reduce labour in image development, image hanging, film processing and archiving. Further, the accessibility of the same digital mammograms from different reading stations through a picture archiving and communication system (PACS) could reduce the need for radiologists to travel between centres to read films, and thus improve their work flexibility and efficiency.

As DM is replacing FM, FM is unlikely to be supported by vendors and developers in the near future. However, given DM is currently not publicly funded any out-of-pocket costs associated with this technological change are likely to be borne by the consumer.

Economic considerations

This assessment determined that DM is at least as safe and effective as FM overall. DM also appears to be associated with improved diagnostic accuracy in women under the age of 50 years, women with radiographically dense breasts, and pre- or perimenopausal women relative to FM. Therefore, we compared the costs of DM and FM in breast cancer screening, diagnosis and surveillance of the overall population. In addition, we analysed the cost-effectiveness for the subgroups in which DM is more accurate than FM in breast cancer screening.

The cost comparison included all cost categories of importance. Given the considerable variation in costs between the two types of DM, computed radiography (CR) and digital radiography (DR), we analysed two diagnostic scenarios: one using CR without a PACS and one using DR with a PACS, as would happen in screening. The outcome in the cost comparison was the incremental cost per examination per year for DM compared with FM .

In the analysis of the cost-effectiveness of screening women younger than 50 years, pre- or perimenopausal women, and women with heterogeneously or extremely dense breasts, the cost-effectiveness of DM was measured as the incremental cost per additional area under the receiver operator characteristic (ROC) curve (AUC), because the difference in the AUC was the primary outcome in the DMIST trial. To enhance the interpretation of the cost-effectiveness data, we also assessed the incremental cost per additional cancer detected as the more clinically relevant outcome.

The cost comparison indicates that DM is $11 more expensive per examination than FM in a **screening** setting, and $36 or $33 more expensive per examination in a **diagnostic** setting when DR or CR is used. In both settings, the average throughput of the mammography system has the most significant impact on the incremental cost per examination per year. Although women at potentially high risk of breast cancer are advised to have annual surveillance (NHMRC 1999), the eligibility of these women for annual screening through the national BreastScreen program depends on the state or territory where they live. The incremental cost per DM examination could thus be between $11 and $36 in a **surveillance** setting.

The cost-effectiveness analysis appears to indicate that the incremental cost per extra cancer detected is around $10 000 in each subgroup examined, suggesting that DM represents good value for money when compared with FM. However, sensitivity analyses demonstrate that the incremental cost per extra cancer detected varies widely as a result of wide confidence intervals of differences in cancer detection rates between DM and FM (due to small numbers in the very large DMIST study). This is particularly the case in the estimates for women with heterogeneously or extremely dense breasts.

Mammography is a frequently performed diagnostic procedure, and the associated financial impact on Medicare is considerable. A change in the cost of the procedure therefore entails extensive financial implications for Medicare. Should DM replace FM, an additional $33 to $36 per examination would be borne by Australian society. This change in technology would represent a total additional cost of $10–$13 million per annum given the number of procedures performed in 2005–06.

Given that the estimated incremental cost per procedure of DM is $11 in a screening setting and that the number of screening examinations was 1.6 million in 2003–2004 (over two years), replacing FM with DM would have a substantial financial impact on Australian society. The annual incremental cost to Australian society would be $9 million for screening 800 000 women. However, with the wide use of DM, the prices of DM machines and associated PACS

will continue to decrease, and thus in the future the actual cost borne by society may not be as great as estimated in this evaluation.

Recommendation

MSAC has considered the safety, effectiveness and cost-effectiveness of digital mammography when compared with conventional film mammography: as a screening test for breast cancer in asymptomatic women aged over 40 years or women at potentially high risk, and in the investigation of women with symptoms of breast cancer.

MSAC finds that digital mammography is as safe and as effective as film mammography. There may be subgroups of patients in whom it is more effective.

Film mammography is being superseded by digital mammography and will lose technical support.

MSAC recommends that public funding for this procedure be supported under the arrangements that currently apply to film mammography.

> – The Minister for Health and Ageing endorsed this recommendation on 11 April 2008.

Glossary

ADC	analogue-to-digital converter
AGD	average glandular dose
AHTA	Adelaide Health Technology Assessment
AJCC	American Joint Committee on Cancer
AOP	automatic optimisation of parameters
ARTG	Australian Register of Therapeutic Goods
area under the curve (AUC)	Calculated as the area under a ROC curve, the AUC provides a numerical description of the accuracy of a diagnostic test. A test with no diagnostic value has an AUC of 0.5, while a perfect test has an AUC of 1
BIRADS	Breast Imaging Reporting and Data System
BRCA1	Breast cancer 1 gene
BRCA2	Breast cancer 2 gene
BSE	breast self-examination
CAD	computer-aided detection; computer-assisted diagnosis
CBE	clinical breast examination
CC	craniocaudal
CCD	charge-coupled device
CNBSS	Canadian National Breast Screening Study
CR	computed radiography
CRD	NHS Centre for Reviews and Dissemination, York, England
CsI	caesium iodide
CsI (Tl)	thallium-activated caesium iodide
CX	chest x-ray
DALY	disability-adjusted life year
DCIS	ductal carcinoma *in situ*
DM	digital mammography
DMIST	Digital Mammography Imaging Screening Trial
DR	digital radiography
ESAK	entrance surface air kerma
false alarm rate	The proportion of false positive tests among people receiving a positive test
false negative	A negative test result when disease status is positive
false negative rate	The proportion of negative tests among people with the disease
false positive	A positive test result when disease status is negative
false positive rate	The proportion of positive tests among people without the disease or condition
false reassurance rate	The proportion of false negative tests among people receiving a negative test
FDA	US Food and Drug Administration
FFDM	full-field digital mammography
FM	film-screen mammography
FNA	fine-needle aspiration biopsy
GaAs	gallium arsenide
He-Ne	helium-neon
heterogeneity	In meta-analysis, the variability in the statistical estimates of studies
HTA	health technology assessment
ICER	incremental cost-effectiveness ratio

interquartile range (IQR)	A measure of dispersion calculated as the difference between the 75th and 25th percentiles of a distribution
ITS	intention to screen
ITT	intention to treat
kerma	kinetic energy released per unit mass
kVp	kilovolt peak
LYG	life years gained
MBS	Medicare Benefits Schedule
meta-analysis	The statistical analysis of a number of individual study results for the purpose of integrating findings
MLO	mediolateral oblique
MM	mammography
MRI	magnetic resonance imaging
MSAC	Medical Services Advisory Committee
NBCC	National Breast [and Ovarian] Cancer Centre
negative predictive value (NPV)	Proportion of patients with negative test results who are correctly diagnosed as not having the disease
NHMRC	National Health and Medical Research Council
NHS	[UK] National Health Service
PACS	picture archiving and communication system
positive predictive value (PPV)	Proportion of patients with positive test results who are correctly diagnosed as having the disease
power	The ability of a statistical test to reject a false null hypothesis
PP	per protocol analysis
QALY	quality-adjusted life year
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
relative risk	A measure of how much a particular risk factor influences the likelihood of an outcome. It is calculated as the incidence of an outcome in the experimental (exposed) group divided by the incidence in the control (non-exposed) group
ROC curve	receiver operator characteristic curve
sensitivity	The proportion of people with a disease who test positive
specificity	The proportion of people without a disease who test negative
STD	standard
TGA	Therapeutic Goods Administration
Tp53	Tumour protein 53 gene

Introduction

The Medical Services Advisory Committee (MSAC) has reviewed the use of digital mammography (DM), which is a screening and diagnostic tool for detecting breast cancer in women. The MSAC evaluates the safety, effectiveness and economics of new and existing health technologies and procedures for which funding is sought under the Medicare Benefits Scheme, while taking into account other issues such as access and equity. The MSAC adopts an evidence-based approach to its assessments, based on reviews of the scientific literature and other information sources, including clinical expertise.

The MSAC's terms of reference and membership are shown in Appendix A. The MSAC is a multidisciplinary expert body comprising members drawn from diagnostic imaging, pathology, surgery, internal medicine, general practice, clinical epidemiology, health economics, consumer health and health administration.

A team from Adelaide Health Technology Assessment (AHTA), in the Discipline of Public Health, School of Population Health and Clinical Practice, University of Adelaide, was engaged to conduct a systematic review of the literature on DM. An advisory panel with expertise in this area then evaluated the evidence and provided advice to the MSAC. The advisory panel members are listed in Appendix B.

This report summarises the current evidence pertaining to the safety, effectiveness and economics of the use of DM for the screening (asymptomatic women), surveillance (women at potentially high risk) and diagnosis (symptomatic women) of breast cancer.

Rationale for assessment

Film-screen mammography (FM) is currently funded by the joint Australian, state and territory governments' BreastScreen Australia Program as a population-based screening test targeted at women aged 50–69 years. It is also subsidised under the Medicare Benefits Scheme as a diagnostic test for individuals with symptoms or a previous history of breast cancer, or those at potentially high risk of breast cancer due to family history.

Government and stakeholders recognise that new digital technology is likely to replace FM owing to technology availability and workforce issues. Inevitable changes in technology availability, including the forced redundancy of conventional mammography technology, are becoming apparent, as is the need for new technology to assist with an Australian shortage of radiologists and radiographers. Thus, DM for the screening, surveillance and diagnosis of breast cancer was referred for assessment by the MSAC by two sections of the Australian Government's Department of Health and Ageing: the Screening Section, Population Health Division (with respect to DM screening of asymptomatic women), and the Diagnostics and Technology Branch (with respect to DM screening of women at high risk and symptomatic women). An assessment of the comparative safety, effectiveness and cost-effectiveness of this new technology is a matter of some urgency.

Background

Digital mammography

DM separates the processes of image acquisition and display, which are currently intertwined in conventional FM. With FM, an image of the breast is generated by converting the energy from the x-ray beam (which has passed through the breast) into light via a phosphor screen, which then exposes a film (hard copy) held within the screen. The film is processed chemically and an image is developed. DM also uses x-rays to produce images of the breast tissue, but the x-ray photons encounter a solid-state detector that converts the absorbed energy into electrical signals (*directly*: radiation → electrical signal; or *indirectly*: radiation → light → electrical signal) (Dershaw 2006). These electrical signals are used to produce images of the breast that can be displayed on a computer monitor (soft copy) or be printed on laser film (hard copy). As DM overcomes the interdependence between image acquisition, processing and image display associated with FM, it allows these three characteristics to be individually manipulated and improved (Pisano & Yaffe 2005).

Digital image quality depends, among other things, on the number of **pixels**, **grey-scale bit resolution** or contrast resolution, and the **signal-to-noise ratio**. A digital image consists of discrete units or picture elements, or pixels. A higher number of pixels within the field of view results in better spatial resolution but perhaps at the expense of a higher radiation dose. The greater the number of grey-scale values or shades of grey (measured in binary digits or bits), the greater the potential contrast resolution. Image noise is caused by random variations in the number of x-ray photons detected per pixel. This causes image degradation. Increasing the dose received by the x-ray detector can reduce image noise but again at the expense of increased radiation dose to the patient. A high-intensity signal with low-intensity noise is required for good digital image quality (James 2004).

Image acquisition and interpretation

Digital receptors have a wider dynamic range (up to 5000:1) than film images (40:1), and thus have the capacity for greater contrast resolution. This can improve image acquisition, especially from women with dense breasts (Dershaw 2006). Images can be interpreted by one or more radiologists, as in conventional mammography, but with an additional option of interpretation through computer-aided diagnosis (CAD) programs if needed (see 'Other relevant considerations' section). CAD involves the use of neural networks and computer algorithms to detect abnormalities in digital or digitised mammograms and to distinguish whether these abnormalities are benign or malignant masses and calcifications (Gur & Sumkin 2006). FM images need to be digitised in order to allow the use of computer-aided detection (also CAD). DM bypasses this process.

Image display and processing

Electronic display has a number of advantages in relation to image storage and processing. Digital images can be stored on a local computer or a centralised picture archiving and communication system (PACS) rather than in film-screen libraries. This storage enables images to be retrieved faster and to be sent electronically (Dershaw 2006; Jong & Yaffe 2005), improving comparison between current and previous mammograms (to identify new breast

abnormalities) and allowing the consultation of radiologists at remote sites (telemammography). The latter practice may be of particular benefit for rural and remote clinics and breast screening (see 'Other relevant considerations' section).

The contrast and resolution of a DM image can be adjusted, images can be magnified and the display can be manipulated to facilitate interpretation of images (Pisano & Yaffe 2005; James 2004). By comparison, limited modifications can be made to film-screen mammograms (only digitisation and spot-view magnification), so if an image is unclear, additional images may need to be taken. This can expose patients to extra radiation and involve additional labour, time and cost (Feig & Yaffe 1998; Hemminger 2003).

Available digital mammography systems

'Digital mammography' in this report relates to full-field (or large-field-of-view) DM (FFDM). This definition applies to any mammographic equipment that produces digital images of a field size comparable to FM, ie, 18 cm × 24 cm or 24 cm × 30 cm (McLean et al 2007). It therefore excludes film-screen digitisation and small-field-of-view digital imaging for biopsy.

Currently there are six types of DM systems available worldwide:

- computed radiography (CR)
- slot-scanning charge-coupled devices (CCD)
- indirect flat-panel detectors
- direct flat-panel detectors
- scanning photon counting systems
- phase-contrast mammographic x-ray units.

The differences between the systems include the type of x-ray detector or transistor used, and each has its advantages and disadvantages. Types of systems available in Australia are listed later in Table 4. The six types are described below.

Computed radiography

CR was the first digital imaging system to be used for mammography. It is a phosphor plate cassette-based digital radiography (DR) system. It uses existing mammographic x-ray equipment and thus can be considered a bridging technology from film-screen systems to flat-panel systems (Feig & Yaffe 1998; McLean et al 2007; James 2004). Both single-plate and multi-plate systems are used for the digitisation of film-screen cassettes, and use will depend on the environment within which the mammography is conducted.

In CR mammography, the x-ray absorber is a phosphor screen with photostimulable luminescence properties. The energy from x-ray absorption causes electrons in the phosphor crystal to be freed and then trapped in a crystal lattice. The number of filled traps in the crystal lattice is proportional to the absorbed x-ray signal. To generate an image, the phosphor crystal plate is scanned line by line with a laser beam (He-Ne or GaAs). The laser stimulates the release of the electrons from the traps, causing the emission of light, which is collected by optics and filtered. The output is then digitised and the signal is sampled to construct a digital image line

by line (Figure 1). The image is subsequently processed and enhanced, forwarded and stored on a radiologist workstation. The plate is then illuminated with a sodium discharge lamp to remove the image and initialise it for reuse (McLean et al 2007; James 2004). The advantages of this system include small detector-element size, retrofitting for use with standard FM units, availability of multiple plate sizes and relatively low costs, including start-up costs (Pisano & Yaffe 2005).

A CR system (eg Fuji Medical Systems) collects stimulated light emissions from both sides of the plate, enhancing efficiency. Pixel sizes range from 43 to 50 μm (McLean et al 2007). A dual-sided reading function allows the stored plate image to be read from both sides simultaneously, resulting in lower image noise (James 2004).

Figure 1 Computed radiography system (dual-sided)

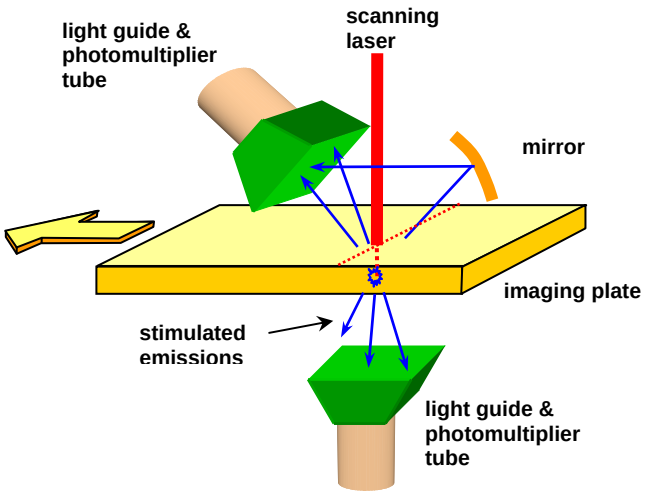
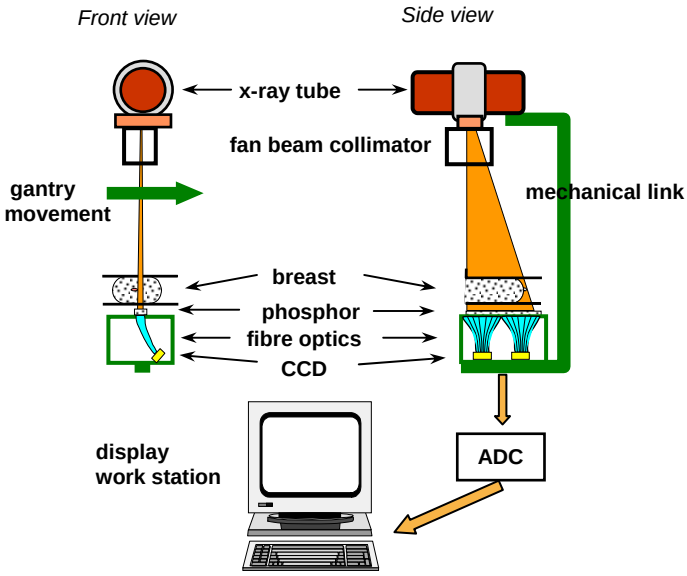


Figure reproduced with permission (McLean et al 2007).

Slot-scanning charge-coupled devices

The slot-scanning phosphor CCD mammography unit was originally used for imaging of stereotactic needle biopsy owing to its small field of view. FFDM devices have since been developed that consist of a conventional phosphor screen (ie, CsI) which is coupled with a CCD via fibre optics (Figure 2). X-rays transmitted through the compressed breast produce light from the phosphor screen. This light is collected by fibre optics and conducted to multiple CCD arrays which convert the light into an electronic signal, which is then digitised by an analogue-to-digital converter (ADC). The image is developed line by line by scanning the fan x-ray beam and the slot detector together across the breast at a constant speed (James 2004; McLean et al 2007). Image acquisition time is 5 seconds, which is longer than with other systems, and higher than the usual load and heat capacity of mammographic x-ray tubes. As a consequence, tungsten x-ray tubes (with better heat dissipation and efficient x-ray production) are preferred over molybdenum x-ray tubes. The detector has a dynamic range of 5000:1. An example of this type of unit was developed by Fischer Medical Imaging and marketed as the SenoScan system (McLean et al 2007; Pisano & Yaffe 2005). One of the advantages of this system is a significant reduction in scatter, which eliminates the need for a grid and may reduce the dose (James 2004).

Figure 2 Slot-scanning phosphor CCD



ADC = analogue-to-digital converter

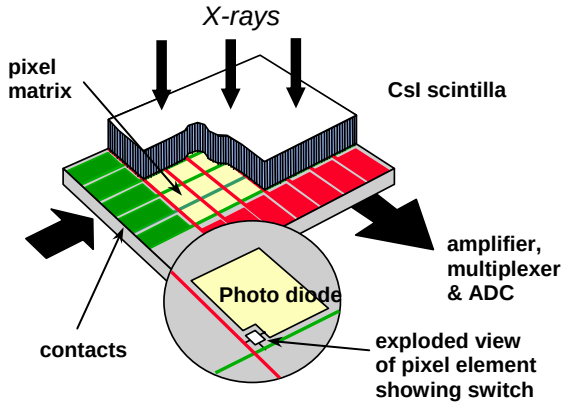
CCD = charge-coupled device

Figure reproduced with permission (McLean et al 2007).

Indirect flat-panel detectors

This indirect system is depicted in Figure 3. X-rays are detected at a layer of thallium (Tl)-activated CsI phosphor (CsI(Tl) phosphor) and are absorbed and converted into light. An amorphous silicon photodiode array then absorbs the light and converts it into an electronic charge signal. The charge at each photodiode is read out at thin-film transistor switches and, with the aid of an ADC, is turned into a digital image. Each photodiode corresponds to a pixel in the image. Pixel sizes are approximately 100 μm, giving favourable spatial resolution. This technology has the advantage of taking rapid image sequences, useful for some applications. Limitations may include the cost of the detector, difficulty in changing format and reducing detector-element size. The Senographe 2000D (General Electric Medical Systems) is an example of an indirect flat-panel detector (McLean et al 2007; Pisano & Yaffe 2005; James 2004).

Figure 3 Indirect flat-panel detector



CsI = caesium iodide

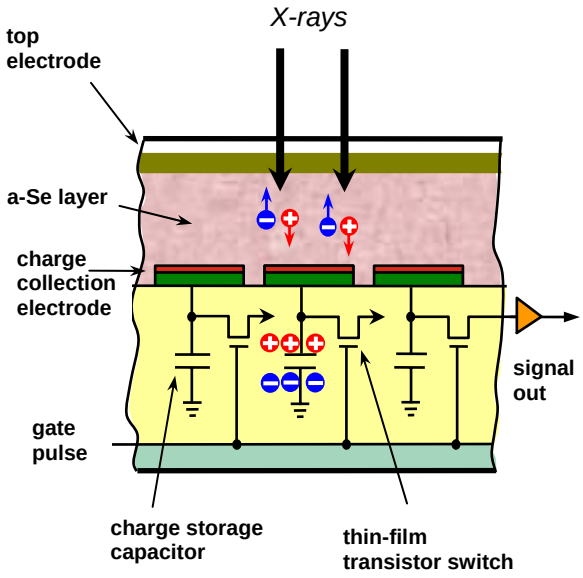
ADC = analogue-to-digital converter

Figure reproduced with permission (McLean et al 2007).

Direct flat-panel detectors

This system is depicted in Figure 4. It differs from the other FFDM systems in that the flat-panel x-ray absorber is composed of amorphous selenium instead of phosphor. This allows for conversion of x-rays to charge pulses without the intermediary of phosphor and consequent light scatter. The potential advantage of this system is its high spatial resolution, from a pixel size of 70 μm. Limitations include the cost of the system and the requirement for 'relaxation' time (the grid must complete an integral number of cycles during an exposure), making screening slow. The technology was developed by Hologic, and marketed examples include the Lorad Selenia Digital Mammography system and the Siemens Novation DR (McLean et al 2007; Pisano & Yaffe 2005; James 2004).

Figure 4 **Direct flat-panel detector**



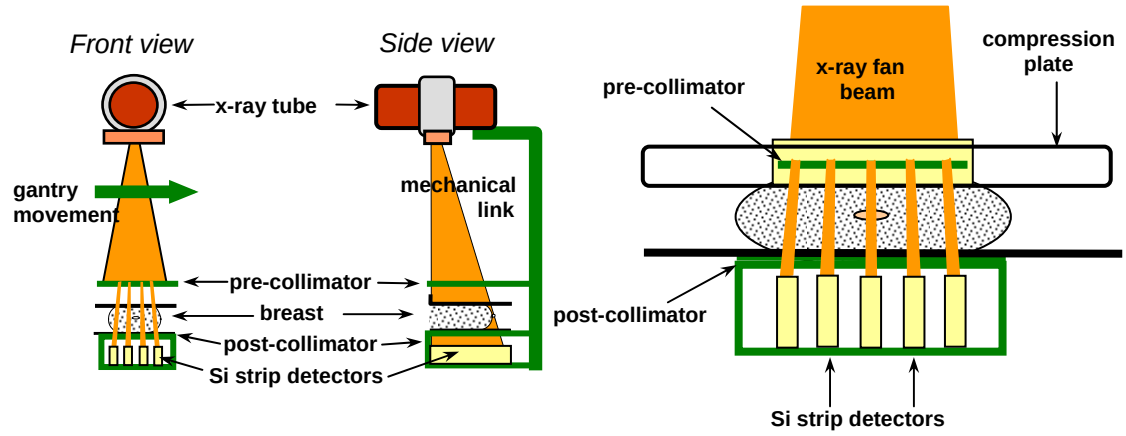
a-Se = amorphous selenium

Figure reproduced with permission (McLean et al 2007).

Scanning photon counting systems

This DM system has some similarities with the slot-scanning CCD system described above. This system, marketed as MicroDose by Sectra, also allows for energy discrimination and the rejection of electronic noise through single photon counting (see Figure 5). X-ray energy is directly converted to charge through the use of a crystal silicon detector. The field of view is 24 cm × 26 cm and the image consists of 4800 × 5200 pixels of size 50 μm. The system provides pre- and post-breast collimation with thin fan beams that move across the breast, thus ensuring that all breast tissue is imaged. Scatter and radiation dosage are reduced (McLean et al 2007).

Figure 5 Scanning photon counting system



Si = silicon

Figure reproduced with permission (McLean et al 2007).

Phase-contrast mammographic x-ray units

Phase-contrast mammography relies on the principle that x-rays refracted at the boundaries of tissues with differing densities will have different refraction indices. When x-rays pass through an object, factors such as the photoelectric effect and scattering reduce the intensity of the x-ray, and at the same time a phase shift occurs, observed as refraction or interference. By converting this phase shift to an image – as changes in x-ray intensity – it is possible to obtain a phase contrast which allows sharper images and an edge effect at reduced radiation exposure (Ingal et al 1998; Konica Minolta 2006). Phase contrast DM uses a 43 cm × 35 cm high-resolution CR plate detector that allows the image to be magnified (McLean et al 2007) and scanned at a very high resolution of 43.75 μm (70 million pixels) (Ingal et al 1998).

Intended purpose

DM can be used as a screening, surveillance and diagnostic test for breast cancer. This report assesses DM as an alternative or replacement for breast cancer screening in asymptomatic women aged 40 years and over; investigation of women with symptoms or a clinical abnormality of the breast suggesting malignancy; and surveillance of women at potentially high risk of breast malignancy due to a previous personal or familial history of breast cancer.

Breast cancer (according to ICD50) may be defined as 'malignant neoplasia, stated or presumed to be primary to the breast'.[1] For the purposes of this review, 'breast cancer' is used to refer to carcinoma arising within the female breast. Although currently there is no universally accepted definition of early breast cancer, the National Health and Medical Research Council's clinical practice guidelines define early breast cancer as 'tumours of not more than five centimetres diameter, with either impalpable or palpable but not fixed lymph nodes and with no evidence of distant metastases' (NHMRC 2001). According to the International Union Against Cancer's staging system, this reflects stages I and II cancer. Advanced breast cancer, on the other hand,

1 WHO International Classification of Diseases (ICD) version 10, 2007.
Source: www.who.int/classifications/apps/icd/icd10online/

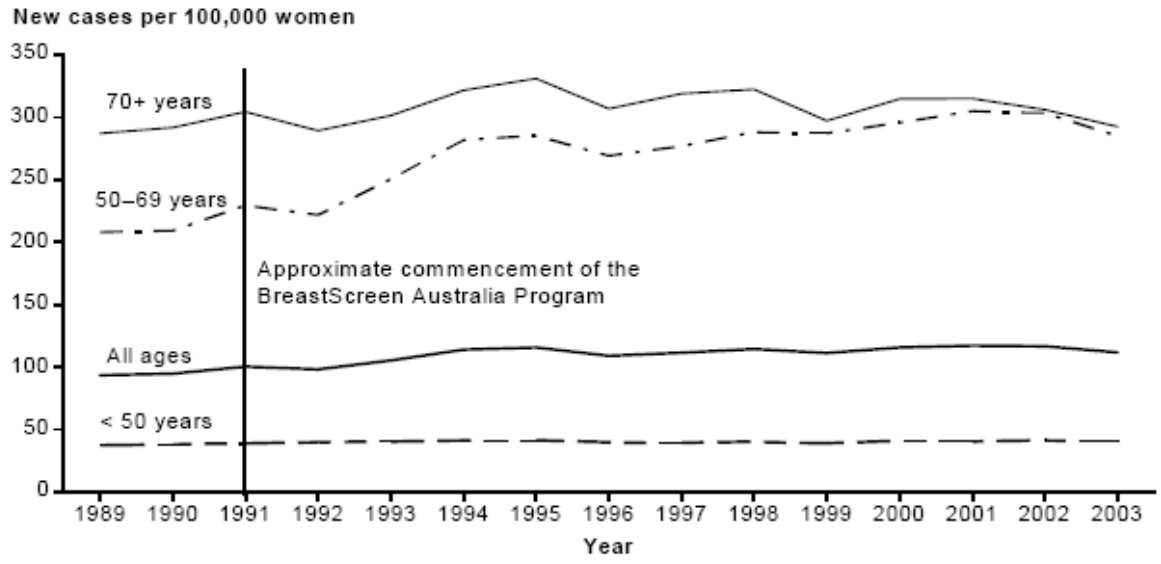
is defined as locally advanced and metastatic. This is otherwise considered as stages III and IV breast cancer. Breast cancer is classified from stages 0 to IV, within which are various subcategories related to tumour type, nodal status and presence of metastasis. This categorisation is done mainly to identify or facilitate the best treatment option (see Appendix C).

Clinical need or burden of disease

Incidence of breast cancer and mortality rates

Breast cancer is the most frequently diagnosed cancer among women in Australia, as evidenced by 11 788 new cases in 2003 (AIHW 2007b). The incidence rose from 105.3 cases per 100 000 population in 1993 to 111.8 cases per 100 000 in 2003 (AIHW 2007b). The increase over this 10-year period reflects both an increasing underlying rate and increased early detection through mammographic screening (Figure 6).

Figure 6 Incidence of breast cancer in women, Australia, 1989–2003



Source: AIHW (2007a)

In the GLOBOCAN database of the International Agency for Research on Cancer, the 2002 Australian age-standardised incidence of breast cancer (83.2 per 100 000) was higher than that of some developed countries, but was on a par with that of Northern Europe, Canada, New Zealand and the United Kingdom (all of which have mammographic screening programs), and lower than that of the United States (AIHW & NBCC 2006). Australia has the fifth highest rate of breast cancer incidence in the world: one in 11 Australian women develop breast cancer before reaching 75 years of age (AIHW 2007b).

Breast cancer is the most common cause of cancer-related death in Australian women, accounting for 2641 deaths in 2004 (AIHW 2007a). However, international comparisons of developed countries show that Australia has low mortality rates, indicating good survival rates, through early detection and treatment strategies (NBCC 2006).

Mortality among Australian women with breast cancer has been declining and relative survival after diagnosis has been increasing. Age-standardised mortality from breast cancer in women of all ages declined from 30.4 per 100 000 women in 1990 to 23.4 per 100 000 in 2004 (AIHW 2007a). The 5-year survival rate of women whose breast cancer was diagnosed in 1998–2002 was 86.6 per cent across all ages, and as high as 90.4 per cent in women aged 60–69 years (AIHW & NBCC 2006). Ten-year relative survival increased by over 15 percentage points, from 57.5 per cent in women diagnosed in 1982–1986 to 73.6 per cent in women diagnosed in 1992–1997 (AIHW & NBCC 2006).

Risk factors

Breast cancer has a number of risk factors, age being the best indicator aside from being female. Over 75 per cent of all new cases occur in women over the age of 50 (AIHW 2006). The age-specific incidence rate ranged from 26.3 cases per 100 000 aged 30–34 years to 299.1 per 100 000 aged 85 years or over in 2002 (AIHW & NBCC 2006). Another known risk factor is genetic predisposition, as outlined in Box 1 (NBCC 2006).

Box 1 Indicators of potentially high risk of breast cancer

Women at potentially high risk of breast cancer (<1% of the female population) are defined as follows:

1. women who are at potentially high risk of ovarian cancer (NBCC, 2006)
2. two 1º or 2º relatives on one side of the family diagnosed with breast or ovarian cancer plus one or more of the following features on the same side of the family:
   - additional relative(s) with breast or ovarian cancer
   - breast cancer diagnosed before the age of 40
   - bilateral breast cancer
   - breast and ovarian cancer in the same woman
   - Ashkenazi Jewish ancestry
   - breast cancer in a male relative (NBCC, 2006)
3. one 1º or 2º relative diagnosed with breast cancer at age 45 or younger plus another 1º or 2º relative on the same side of the family with sarcoma (bone/soft tissue) at age 45 or younger (NBCC, 2006)
4. member of a family in which the presence of a high risk breast cancer gene mutation has been established eg *BRCA1*, *BRCA2*, *Tp53* (NBCC, 2006)
5. personal history of breast cancer
6. pre-malignant conditions: lobal carcinoma *in situ* or atypical ductal hyperplasia

Women in this high-risk population are more likely to develop cancer at a younger age (<50 years) than those in the general population. A number of other factors may play a role in increased breast cancer risk, including (AIHW 1997):

- larger body size
- reproductive factors such as nulliparity, older age at birth of first child, early menarche and late menopause

- long-term use of exogenous oestrogens in hormone replacement therapy
- exposure of breast tissue to ionising radiation (Kricker & Jelfs 1996).

Risk of breast cancer in women with a genetic predisposition

Genetic mutations in *BRCA1*, *BRCA2* or both can increase the risk of developing breast cancer. These account for one to five per cent of all breast cancers. The lifetime risk of developing breast cancer in this population has been estimated to be up to 80 per cent (NHMRC 1999). Recent studies, however, suggest that it may be lower. A population-based study of young Australian women and their families reported that the average age-specific cumulative risk for breast cancer associated with *BRCA1* and *BRCA2* was, on average, nine times the population risk [95%CI 4, 23; $P < 0.001$] (Hopper et al 1999). Penetrance to age 70 years was 40 per cent [95% CI 15%, 65%]. Similarly, in a study of multiple-case breast cancer families recruited from a clinical setting, after adjustment for ascertainment by conditioning on the genotype of the proband and family phenotype, the average cumulative risk of breast cancer for mutations in either *BRCA1* or *BRCA2* was reported as 27 per cent [95% CI 16%, 43%] to age 50 and 64 per cent [95% CI 44%, 83%] to age 70 (Scott et al 2003).

Another genetic mutation responsible for increasing the risk of developing breast cancer occurs in the gene for tumour antigen p53 (*Tp53*). This mutation is associated with 25 times the risk of developing breast cancer by 50 years of age than in the general population (Evans & Lozano 1997). Table 1 shows the frequency of known gene mutations in the general population.

Table 1 Known gene mutations in hereditary breast cancer, frequency and risk

Gene	Mutation frequency	Major sites at risk	Breast cancer risk to age 75 years in mutation carriers
BRCA1	1/1000	breast ovary	40%–80%
BRCA2	1/1000	breast ovary	40%–80%
Tp53	1/10 000	breast bone or soft tissue	50%

Source: (NBCC 2006)

Methodologies for early detection of breast cancer in asymptomatic women

There are a number of approaches to early detection of breast cancer in asymptomatic women. These include breast self-examination (BSE), clinical breast examination (CBE), FM or FFDM with or without magnetic resonance imaging (MRI), and ultrasound. These procedures primarily aim to reduce the rate of mortality from breast cancer.

BSE is a patient-centred, non-invasive approach to detecting lumps in breast tissue. Although women are encouraged to undertake BSE, it is not considered an appropriate screening approach, as it has very low sensitivity and specificity and has not been shown to reduce breast cancer mortality. There is no evidence to support the use of BSE alone as a screening tool. Additionally, the rate of this examination is low: a US survey (Women physician's health study) identified that only 21 per cent of women ($n = 4501$) performed monthly BSE (Frank et al 2000).

CBE is a physical examination carried out by a health professional and may be undertaken at the time of biennial Pap smears in asymptomatic women; annually for women at potentially high risk of breast cancer; or on presentation with symptoms. It is often undertaken in conjunction with mammography.

FM is currently the primary method for the early detection of breast cancer in Australia and is the gold standard in Australia (Table 2). The procedure involves firmly placing the breast between two plates, which compress it, and pull the breast tissue away from the chest wall. Screening mammography involves taking two views of the breast: one from the side (mediolateral oblique) and one from the top (craniocaudal) (Forrest & Anderson 1999). This procedure takes approximately 15 minutes.

A considerable body of literature on the effectiveness of FM has accumulated during the last 30 years. In general, this shows that FM can reduce breast cancer mortality in a screening population by approximately 20 to 30 per cent (Tabar et al 2000; Woolf 2001), although there has been much debate regarding the trade-offs in benefits and risks (Gøtzsche & Nielsen 2006; Wright & Mueller 1995), in particular for women aged 40–49 years (Smart 1994). The high resolution of FM enables the identification of microcalcifications and spiculations, and its high contrast allows visualisation of differences in soft tissue density. It allows imaging of breasts of different size and the simultaneous display of multiple breast images (Feig & Yaffe 1998).

However, conventional mammography is neither highly sensitive nor specific. Sensitivity of screening mammography to detect cancers ranges from 70 to 90 per cent, meaning that 10 to 30 per cent of breast cancers are missed at screening (Brem et al 2003), commonly because of the difficulties in imaging dense breast tissue. Owing to the limited range of soft tissue densities in the breast, FM requires high contrast. However, if high contrast is obtained in medium-density tissue, there will be lower contrast within the thicker, denser fibroglandular tissues, and thus lesions may be obscured (Feig & Yaffe 1998). Efforts to enhance the sensitivity of mammographic screening have included improving film and equipment, intensive training to sharpen the interpretive skills of radiologists, the use of double reading of mammograms (ie, two radiologists), digitisation of mammograms to allow spot magnification, and the assistance of CAD (Hemminger 2003).

Guidelines for breast cancer screening differ internationally, with respect to the target age group to be screened, the suggested intervals between screening, the number of views taken of each breast and the recommended screening methods.

Existing procedures

Screening of asymptomatic women

BreastScreen Australia, introduced in 1991, is a national breast cancer screening program which provides free film-screening mammograms biennially for women without breast cancer symptoms or signs who are 50–69 years old. Asymptomatic women who are in their 40s or who are 70 years and older are also offered free screening mammograms, although they are not targeted. This screening program is operated independently by the various Australian states and territories but within a national policy framework (Table 2).

Table 2 FM for screening in Australian states and territories

State or territory	Mammography for screening (asymptomatic women)	Mammography for surveillance through screening program [a] (women at high risk)
Australian Capital Territory	✓ Fixed units (77% of screens); 1 relocatable unit (23%)	Annual screen; after cancer in one breast, women can return to screening program, but waiting period depends on clearance from physician; continuity of care with own doctor is advised
New South Wales	✓ Fixed units (67% of screens); 18 mobile or relocatable units (33%)	Annual screen; after cancer in one breast, women can return to screening program, but waiting period depends on clearance from physician; continuity of care with own doctor is advised
Northern Territory	✓ Fixed units (67% of screens); 1 relocatable unit (33%)	Annual screen; after cancer in one breast, women can return to screening program after 5-year waiting period [b]; continuity of care with own doctor is advised
Queensland	✓ Fixed units (77.4% of screens); 5 relocatable units (10.9%); 4 mobile units (11.7%)	Annual screen and CBE; after cancer in one breast, women can return to screening program after 5-year waiting period [b]
South Australia	✓ 6 fixed units (72.7% of screens); 3 mobile units (27.3%)	Annual screen; after cancer in one breast, women can return to screening program after 10-year waiting period unless otherwise cleared by physician
Tasmania	✓ 2 fixed units (67% of screens); 1 mobile unit (33%)	Annual screen; after cancer in one breast, women can return to screening program after 5-year waiting period [b]
Victoria	✓ Fixed units (92% of screens); 2 mobile units: 1 digital, 1 analogue (8%)	Previous mastectomy – annual screen; continuity of care with own doctor is also advised Previous breast-conserving surgery – not eligible; continuity of care is advised
Western Australia	✓ Fixed units (73% of screens); 4 mobile units (27%)	Annual screen; after cancer in one breast, women can return to screening program (no waiting period); continuity of care is advised

a: If a woman is at potentially high risk, personal surveillance is at the discretion of an individual and her doctor. Annual surveillance (mammography, CBE) is recommended, and mammography would occur through private radiology and be billed through Medicare if the individual is not of an age or clinical history to be eligible through the screening program, or there is a waiting period (see Figure 8 later). b: Depending on usual access to radiology services; n/a = not available; CBE = clinical breast examination.

Approximately 1.6 million women of all ages across Australia were screened through BreastScreen in the period 2003–2004, including in remote areas (AIHW 2007a). The participation target for breast cancer screening of asymptomatic women aged 50–69 years is 70 per cent. Participation among women in Australia aged 50–69 years increased from 51.4 per cent in 1996–1997 to 57.1 per cent in 2001–2002 and decreased significantly to 55.6 per cent in 2003–2004 (AIHW 2007a). BreastScreen Australia detected 3851 invasive cancers (any size, all screening rounds) in women in 2004 (AIHW 2007a).

Following routine screening, women may be recalled for further investigation at a BreastScreen service if the screening mammogram suggests mammographic abnormalities that cannot be definitively determined as benign. The recall of women may lead to further tests, including assessment mammography and ultrasound (AIHW 2007a).

The aim of BreastScreen Australia is to maximise the number of cancers detected while reducing unnecessary investigations. In South Australia, approximately three per cent of women have a mammographic abnormality detected at screening and are recalled for further assessment. Approximately three-quarters of the women who attend for assessment will have a

benign outcome and will be re-invited for routine screening (BreastScreen South Australia 2005).

The National Accreditation Standards for recall for additional assessment require:

- <10 per cent of women aged 50–69 years who attend for their first (prevalent) screen to be recalled for assessment
- <5 per cent of women aged 50–69 years who attend for their second or subsequent (incident) screen to be recalled for assessment.

Table 3 shows age-standardised recall-to-assessment rates among women aged 40 years and over and 50–69 years in the years 1998, 2002 and 2003.

Table 3 Age-standardised recall-to-assessment rates among women aged 40 years and over and 50–69 years in 1998, 2002 and 2003

	1998	2002	2003
First screening round			
Rate (%) among women aged 50–69 years [95% CI]	7.2 [7.1, 7.4]	8.7 [8.5, 9.0]	9.3 [9.1, 9.6]
Rate (%) among women aged 40 years and over [95% CI]	7.2 [7.0, 7.3]	8.6 [8.4, 8.7]	9.3 [9.1, 9.6]
Subsequent screening rounds			
Rate (%) among women aged 50–69 years [95% CI]	3.9 [3.9, 4.0]	4.0 [4.0, 4.1]	4.0 [4.0, 4.1]
Rate (%) among women aged 40 years and over [95% CI]	3.9 [3.9, 4.0]	4.1 [4.1, 4.2]	4.2 [4.1, 4.2]

Source: (AIHW 2006)

Figure 7 outlines the current approach to screening for breast cancer in asymptomatic Australian women.

Annual surveillance of women at potentially high risk

Figure 8 describes the current protocol for surveillance for breast cancer in Australian women at potentially high risk.

Australian women at potentially high risk of breast cancer or with a previous history of breast cancer are encouraged to undertake a high-surveillance regimen, involving annual CBE and diagnostic mammography, with or without ultrasound, and with or without MRI. MSAC recently completed a report on the use of MRI for surveillance of breast cancer in a high-risk population (MSAC 2007). It recommended interim Australian public funding for breast MRI in the diagnosis of breast cancer in asymptomatic women less than 50 years of age with a high risk of developing breast cancer as part of an organised surveillance program (with the evidence to be reviewed in not less than 3 years).

Since BreastScreen Australia provides screening services to asymptomatic women at average population risk of breast cancer, MRI scanning for women at potentially high risk is not routinely available through the screening program at the time of this report (Table 2). Continuity of care with a selected health professional (general practitioner or physician) who knows the woman's clinical history is advised (NHMRC 2001).

Diagnosis of symptomatic women

Australian women who present to their doctor with breast symptoms such as a palpable lump or nipple discharge will undergo a battery of clinical and imaging tests to determine whether the symptoms are associated with breast cancer (Figure 9).

Originally, diagnostic mammography involved obtaining three views of each breast – craniocaudal (CC), mediolateral oblique (MLO) and lateral. Currently, however, in Australia, two radiologic views (CC and MLO) are used, and additional views are used only when the results obtained from CC and MLO views indicate that they are necessary.

Diagnostic mammography is carried out through private radiology practices and rebated by Medicare. Complementary imaging procedures such as breast ultrasound and MRI can be undertaken. Although breast ultrasound is not routinely used for screening, it has been shown to detect small breast cancers in younger women with dense breast tissue, who are not entirely suitable for mammography (Jackson et al 1996; MSAC 2007). MRI produces images of tissues by using a strong external magnetic field and radiofrequency. This procedure has been used in the diagnosis of breast cancer through the use of multiple cross-sectional images of the breast (side-to-side, top-to-bottom, front-to-back), but its use for diagnostic mammography is currently not rebated.

Palpable lesions are commonly investigated by fine-needle biopsy or core biopsy. Fine-needle biopsy is the sampling of cells from breast tissue for cytological examination. When suction is applied during the sampling, this is referred to as fine-needle aspiration biopsy (FNA). Core biopsy uses a wide-bore needle to obtain a tissue sample (NHMRC 2001). In some practices, core biopsy has replaced FNA because of the greater specificity of core biopsy over FNA. Locating impalpable tumours may require imaging guidance by means of mammography, ultrasound or MRI (NHMRC 2001).

Comparator

The comparator for DM differs depending on whether a woman is asymptomatic or not, and in those who are asymptomatic it depends on their level of risk. For asymptomatic women aged 40 years and over attending a breast screening program, the comparator is FM, consisting of two x-ray views of each breast and follow-up assessment and imaging of those with abnormal mammograms (Figure 7).

The comparator for the use of DM in women at potentially high risk of breast cancer is diagnostic FM, consisting of two (or more) x-ray views of each breast with or without a breast ultrasound scan, and with or without breast MRI (Figure 8).

In women with symptoms or a clinical abnormality of the breast suggesting malignancy, the comparator is diagnostic FM, consisting of two (or more) x-ray views of the affected breast and imaging of the contralateral breast with or without a breast ultrasound scan, and with or without breast MRI (Figure 9).

The **reference standard** for a diagnosis of breast cancer – in women with any of these indications (asymptomatic, potentially high risk, symptomatic) – is histopathology.

Current treatments for breast cancer

A basic clinical flow chart for breast cancer treatment is provided in Figure 10. The choice of type of breast cancer treatment to be initiated depends on the stage of tumour development, the tumour characteristics and the woman's preferences.

Surgery for primary breast cancer aims to completely remove the tumour and any local extension of it. There are two main surgical options for the removal of breast cancer: breast-conserving surgery (otherwise known as lumpectomy) and mastectomy. A **lumpectomy** involves the local excision of the breast tumour around histologically clear margins along with a rim of normal breast tissue on all sides. This is a suitable approach when margins can be defined. On the other hand, a **mastectomy** involves the complete removal of the breast. This is considered when the tumour spreads widely throughout the breast, the margins are not clear, there is involvement of the nipples or overlying skin, or the patient decides not to undergo a lumpectomy (Crawford et al 2000; Hayes et al 1998). Mastectomy may also be undertaken as a prophylactic measure for women at very high risk of breast cancer (Lostumbo et al 2004).

Chemotherapy is the use of drugs to remove cancer cells or slow down their growth. Multi-agent chemotherapy has been shown to decrease the annual risk of recurrence and death in women up to 70 years of age. A systematic review of 47 trials involving 18 000 women with early breast cancer found that some months of adjuvant polychemotherapy (eg, with cyclophosphamide, methotrexate and fluorouracil or an anthracycline-based regimen) typically produces an absolute improvement of about 7 to 11 per cent in 10-year survival for women aged under 50 at presentation, and of about 2 to 3 per cent for those aged 50–69 (unless their prognosis is likely to be extremely good even without such treatment) (Early Breast Cancer Trialists' Collaborative Group 2001a).

Radiotherapy uses radiation, in the form of x-rays or gamma rays, to destroy tumour cells. This is commonly used in conjunction with a lumpectomy as it has been shown to significantly reduce the risk of local recurrence and the need for further surgery. Meta-analysis of 10-year and 20-year results from 40 unconfounded randomised trials of radiotherapy for early breast cancer (20 000 women) found that radiotherapy regimens that are able to produce a two-thirds reduction in local recurrence, but without long-term hazard, would be expected to produce an absolute increase in 20-year survival of about 2 to 4 per cent (except for women at particularly low risk of local recurrence). The authors conclude, however, that the average hazard seen in these trials would likely reduce this 20-year survival benefit in young women and reverse it in older women (Early Breast Cancer Trialists' Collaborative Group 2002).

Hormonal therapy is the use of drugs or hormones which specifically inhibit the growth of hormone-responsive cancer cells. This type of therapy can be used adjuvantly in the treatment protocol, such as in addition to chemotherapy or to surgery and radiotherapy (National Breast Cancer Centre 2004). Hormone drugs include anti-oestrogens such as tamoxifen. Level I evidence suggests that in trials of 1 year, 2 years, and about 5 years of adjuvant tamoxifen, the proportional recurrence reductions produced among 30 000 women (predominantly oestrogen-receptor–positive tumours or with oestrogen receptor status unknown) during about 10 years of follow-up were 21, 29, and 47 per cent, respectively, with a highly significant trend towards greater effect with longer treatment ($P < 0.00001$). The corresponding proportional mortality reductions were 12, 17, and 26 per cent, respectively, and again the trend was significant ($P = 0.003$). The absolute improvement in recurrence was greater during the first 5 years, whereas the improvement in survival grew steadily larger throughout the first 10 years. The proportional mortality reductions were similar for women with node-positive and node-negative disease, but the absolute mortality reductions were greater in node-positive women (Early Breast Cancer

Trialists' Collaborative Group 2001b). In addition, the option of **ovarian ablation** destroys the function of the ovaries by removal, irradiation or hormonal suppression (Early Breast Cancer Trialists' Collaborative Group 1996).

Market status of the technology

Nine types of DM units are included on the Australian Register of Therapeutic Goods (ARTG) (Table 4). DM is currently available through some public and private radiology service providers in Australia.

Table 4 Digital mammography equipment listed on the ARTG [a] as at May 2007

Product name [a]	DM type	ARTG #	Product #	Sponsor
FCR ClearView-1	Computed radiography	118214	198739	Fuji Medical Systems
Kodak DirectView	Computed radiography	115560	195317	Kodak Australasia
PCR Eleva	Computed radiography	92461	163075	Philips Electronics Australia Ltd
CR Regius 190 and 110	Computed radiography	118970	199760	Shimadzu
Senographe 2000D	Indirect flat-panel detector	111376	190743	GE Healthcare Australia Pty Ltd
Hologic Lorad Selenia	Direct flat-panel detector	99194	171994	Medi Consumables Pty Ltd
Hologic Lorad Selenia	Direct flat-panel detector	100125	174146	InSight Oceania Pty Ltd
Novation DR	Direct flat-panel detector	99143	171829	Siemens Ltd
MicroDose	Scanning photon counting system	106898	185664	Sectra Pty Ltd
PCM System	Phase-contrast mammography	61891	171789 [b]	Shimadzu
		99106		Siemens Ltd

a: PlanMed Sophie Digital – a direct flat-panel detector – is likely to be submitted for TGA registration late in 2007 by Medical Imaging Technologies Pty Ltd. b: Every company that wants to import the same product must apply to the TGA.

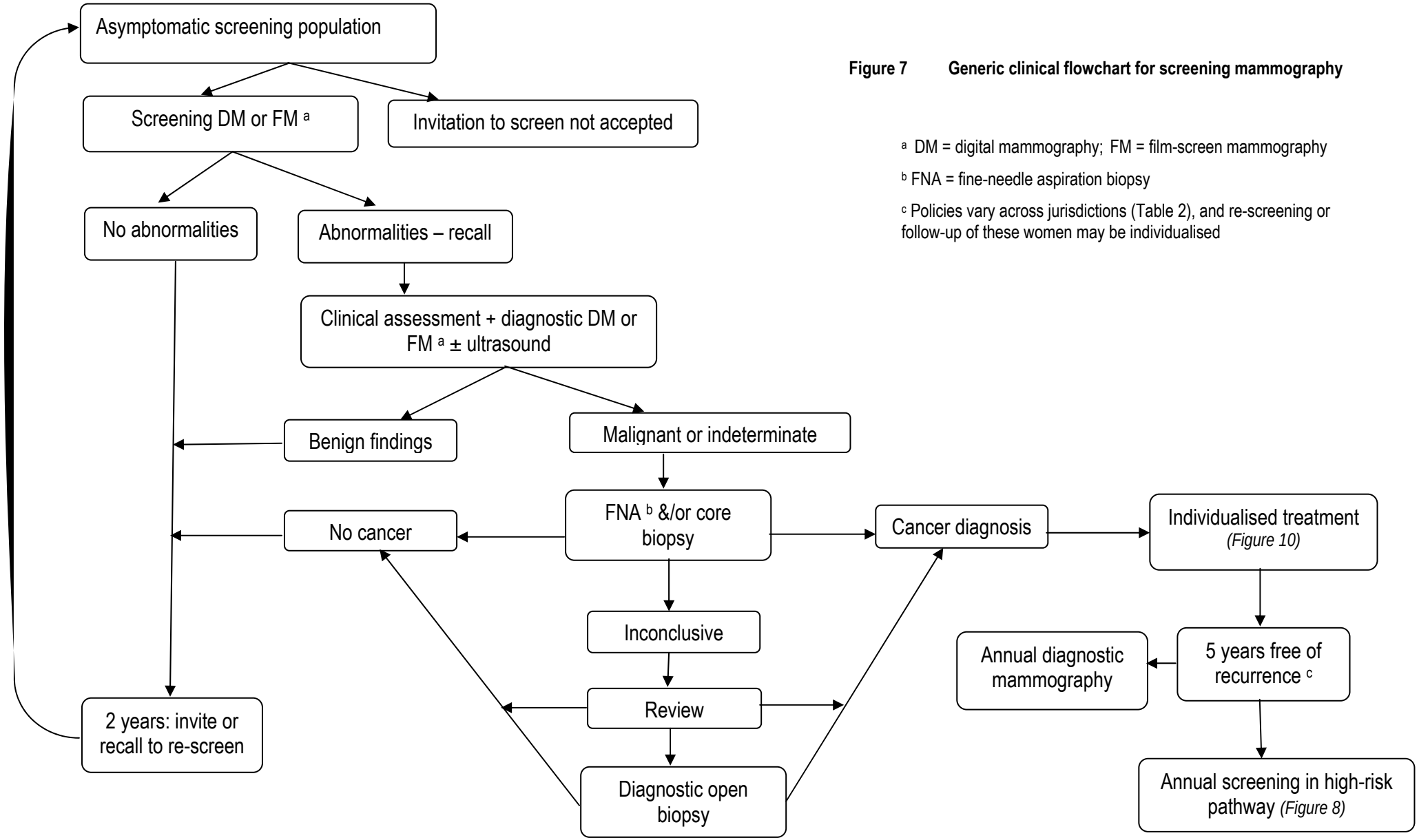


Figure 7 Generic clinical flowchart for screening mammography

[a] DM = digital mammography; FM = film-screen mammography

[b] FNA = fine-needle aspiration biopsy

[c] Policies vary across jurisdictions (Table 2), and re-screening or follow-up of these women may be individualised

Figure 8 Generic clinical flowchart for annual surveillance of women at potentially high risk of breast cancer [a]

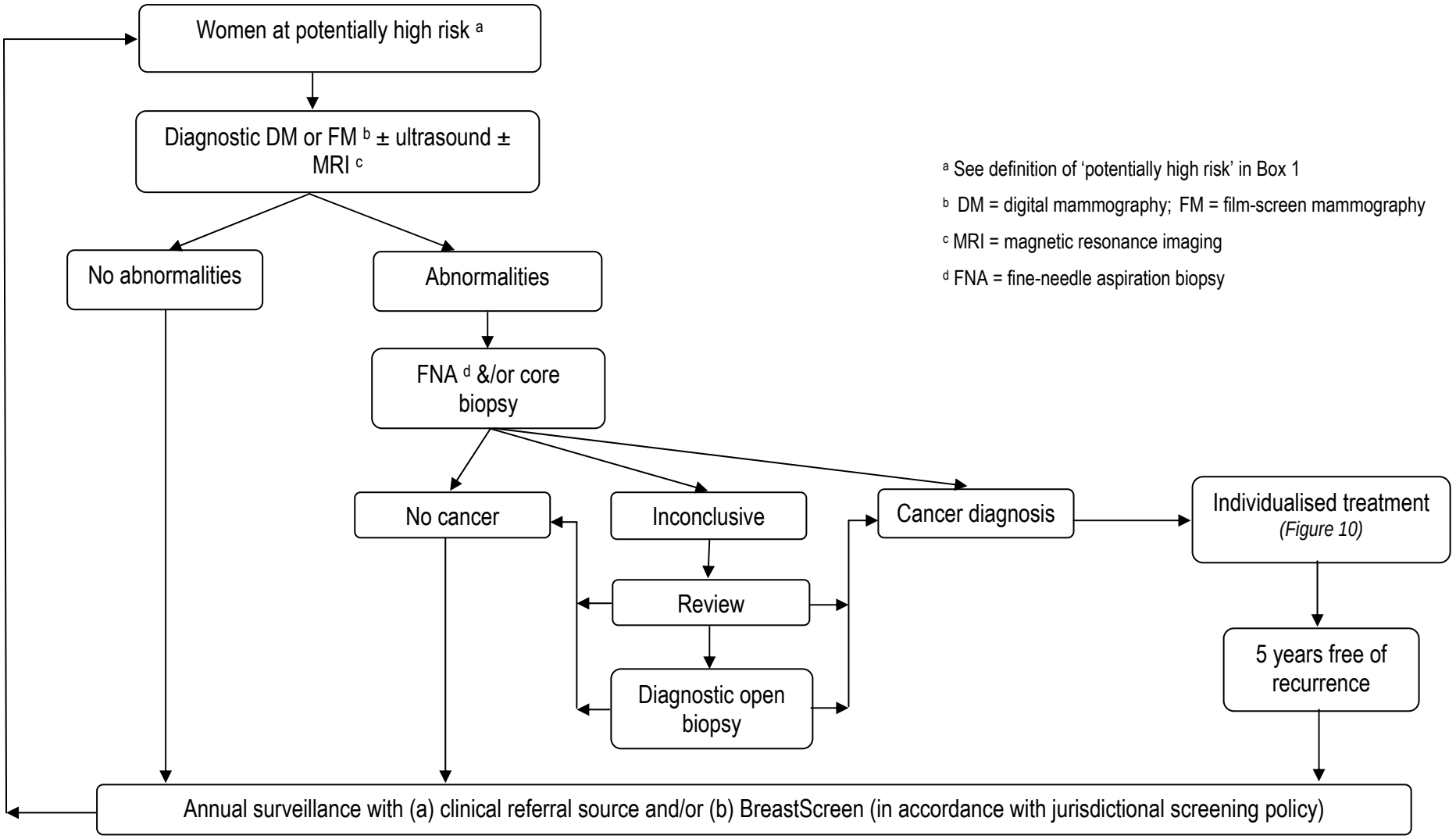
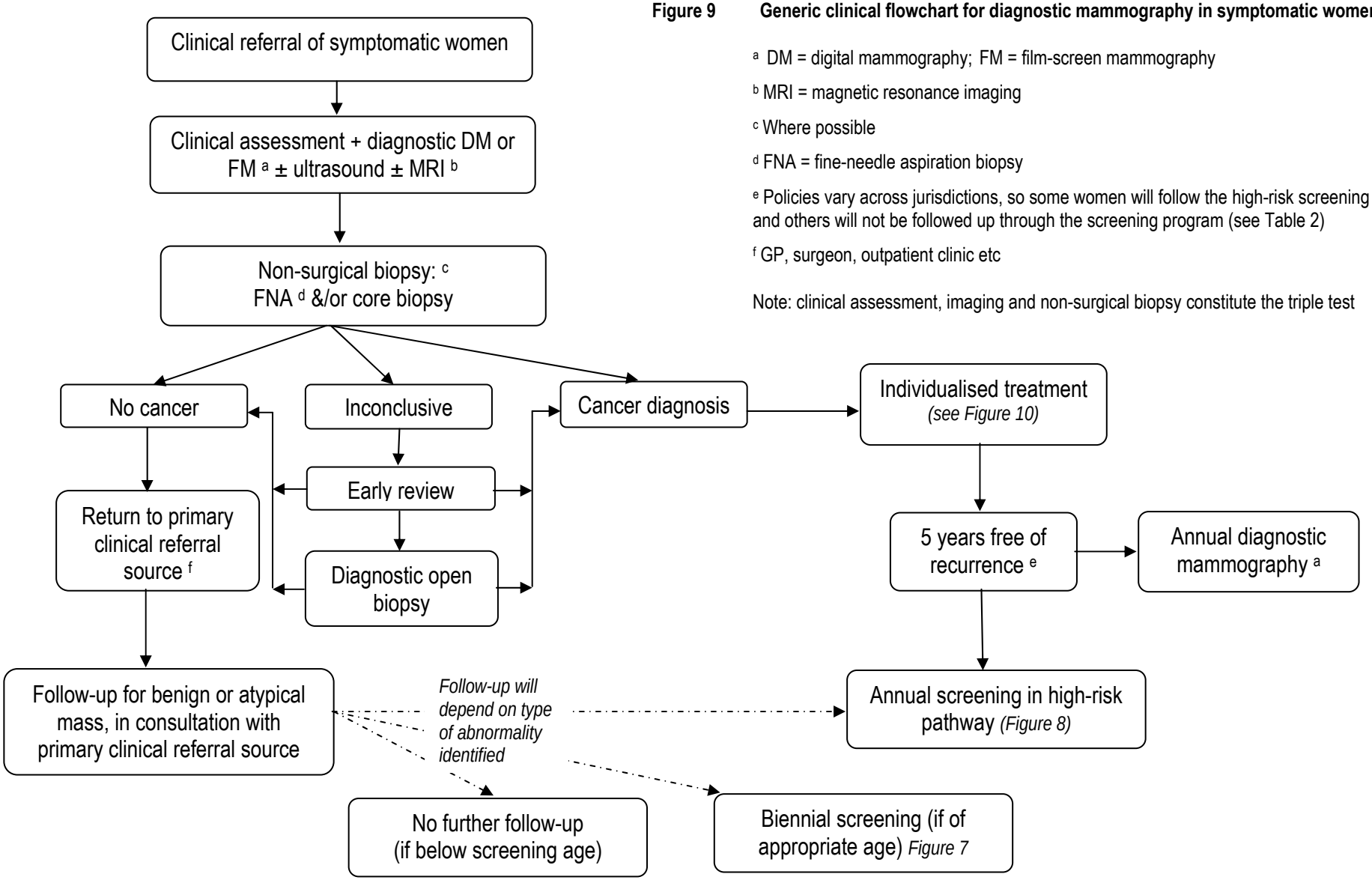


Figure 9 **Generic clinical flowchart for diagnostic mammography in symptomatic women**



[a] DM = digital mammography; FM = film-screen mammography

[b] MRI = magnetic resonance imaging

[c] Where possible

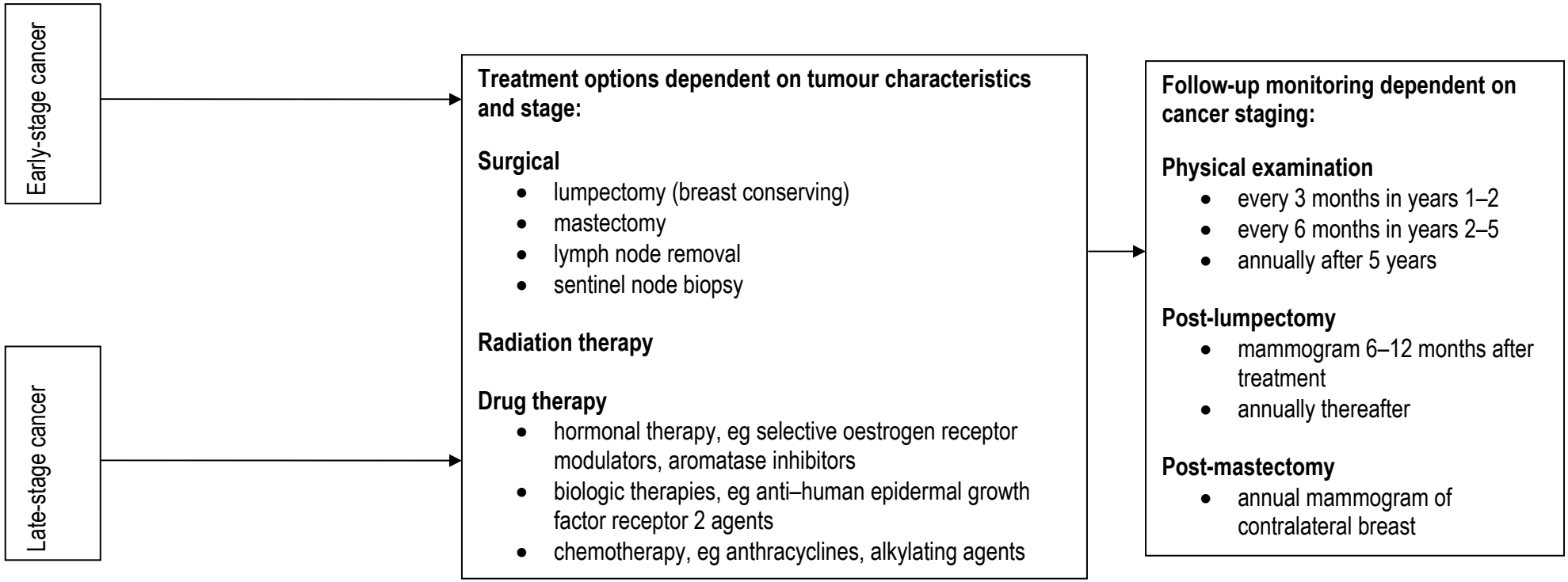
[d] FNA = fine-needle aspiration biopsy

[e] Policies vary across jurisdictions, so some women will follow the high-risk screening pathway and others will not be followed up through the screening program (see Table 2)

[f] GP, surgeon, outpatient clinic etc

Note: clinical assessment, imaging and non-surgical biopsy constitute the triple test

Figure 10 Clinical flowchart for breast cancer treatment [a]



[a] Treatment options are individualised according to patient characteristics, prognostic factors, response to therapy and preference. For more comprehensive evidence-based guidelines on breast cancer treatment see NCCN (2006) and NHMRC (2001).

Current reimbursement arrangement

Film mammography is currently funded by the joint Australian, state and territory governments' BreastScreen Australia program as a population-based screening test targeted at women aged 50–69 years. These women are invited and reminded to attend for free biennial mammographic screening. Women aged 40–49 years or 70 years and over are also eligible for a screening mammogram every 2 years under this program, although they do not receive reminders.

Mammography is also subsidised under the Medicare Benefits Scheme as a diagnostic test for individuals with symptoms, a previous history of breast cancer, or at potentially high risk of breast cancer due to family history (Medicare Benefits Schedule [MBS] item numbers 59300 and 59303; Table 5). The current item numbers cover DM for these indications. However, the current reimbursement arrangement was developed on the basis of the costs of FM, and these costs may differ from those of DM.

Table 5 MBS items for mammography of women with symptoms or at potentially high risk of breast cancer

MBS item numbers	Services
59300	**Mammography** of both breasts if there is a reason to suspect the presence of malignancy because of: (i) the past occurrence of breast malignancy in the patient or members of the patient's family; or (ii) symptoms or indications of malignancy found on an examination of the patient by a medical practitioner. Unless otherwise indicated, mammography includes both breasts (R) Fee: $89.50. Benefit: 75% = $67.15, 85% = $76.10
59303	**Mammography** of one breast if: (a) the patient is referred with a specific request for a unilateral mammogram; and (b) there is reason to suspect the presence of malignancy because of: (i) the past occurrence of breast malignancy in the patient or members of the patient's family; or (ii) symptoms or indications of malignancy found on an examination of the patient by a medical practitioner (R) Fee: $53.95. Benefit: 75% = $40.50, 85% = $45.90

Sourced from (MBS 2007) (R) = requested service

Approach to assessment

Objectives

To carry out a structured evaluation of DM for (1) breast cancer screening in the general population, (2) surveillance of women at potentially high risk of breast cancer, and (3) diagnosis in women with signs or symptoms of breast cancer.

To determine whether DM can *replace* FM, or whether it should be used as an *additional* method to identify breast abnormalities that have not been identified with FM (ie, used only in those women who test negative on FM). The latter usage scenario would most likely have application where women present with symptoms or are at potentially high risk of breast cancer.

Research questions

1. What are the safety, effectiveness and cost-effectiveness of DM, compared with FM, for screening for breast cancer in asymptomatic women aged 40 years and over attending a breast screening program?

2. What are the safety, effectiveness and cost-effectiveness of DM ± breast ultrasound scan ± MRI, compared with FM ± breast ultrasound scan ± MRI, for surveillance of breast cancer in women who are at potentially high risk of breast cancer?

3. What are the safety, effectiveness and cost-effectiveness of DM ± breast ultrasound scan ± MRI, compared with FM ± breast ultrasound scan ± MRI, for diagnosing breast cancer in women presenting with signs or symptoms of malignancy?

Diagnostic assessment framework

This assessment of DM is based on the framework outlined in the MSAC *Guidelines for the Assessment of Diagnostic Technologies* (MSAC 2005).

To assess the effectiveness of DM, we need to consider its diagnostic accuracy (in comparison with a reference standard), its impact on the clinical management of women, and its ultimate impact on the health outcomes of women. The first goal of this assessment was to find **direct evidence** of the effectiveness of DM on health outcomes. That is, one group of women would receive DM ± subsequent testing, treatment and follow-up and would be compared with another group receiving FM ± subsequent testing, treatment and follow-up, for a period of time until the impact on health outcomes (eg, survival) could be evaluated. No direct evidence of the impact of DM on final health outcomes was identified in the search, although one study (in a screening population) did report the impact on a surrogate health outcome (cancer detection rate). A **linked evidence** approach was therefore undertaken.

In some situations it is appropriate to narratively link evidence from studies that report on:

- diagnostic test performance (diagnostic accuracy) – sensitivity, specificity and accuracy

- the impact on clinical decision-making – does clinical decision-making (patient management) change as a result of the test?
- the impact of the treatment of diagnosed patients on health outcomes – do patients receiving a change in management benefit in terms of health outcomes?

to infer the effect of the diagnostic test on patient health outcomes.

A methodological decision algorithm for the use of linked evidence in the assessment of diagnostic tests was applied to this assessment of DM for screening, surveillance and diagnosis of breast cancer (Table 6).

With respect to the use of DM for screening asymptomatic women, we recognised that the test was proposed as a *replacement* for the current screening strategy and thus would not need to be evaluated in the context of a screening test being implemented in a previously unscreened population.

Table 6 Decision algorithm for use of linked evidence in the assessment of diagnostic tests[2]

Index test relative to comparator test	Evidence requirement for analysis	Possible decision on use of test [a]
Less accurate	Evidence of diagnostic accuracy Comparative assessment of test invasiveness & safety considerations	If as safe, index test is not a suitable alternative to comparator test If safer, there may be trade-off between safety and effectiveness
As accurate	Evidence of diagnostic accuracy Comparative assessment of test invasiveness & safety considerations	If as safe, index test is potential alternative to comparator If safer, index test is potential replacement for comparator If not as safe, index test is not a suitable replacement or alternative
More accurate, identifies different disease spectrum in populations, but same treatment options available for both alternatives	Evidence of diagnostic accuracy Comparative assessment of test invasiveness & safety considerations Evidence of change in management (optimisation of diagnostic strategy or treatment)	If as safe, index test is potential replacement for comparator, or used in addition to comparator [b] If safer, index test is potential replacement for comparator If not as safe, there is a trade-off between safety and effectiveness – potentially used in addition to comparator [b]
More accurate, identifies different disease spectrum in populations, different treatment options available for both alternatives	Evidence of diagnostic accuracy Comparative assessment of test invasiveness & safety considerations Evidence of change in management (optimisation of diagnostic strategy or treatment) Evidence of new or alternative treatment effectiveness in the identified population with the different disease spectrum Note: this is a proxy for direct evidence	If as safe, index test is potential replacement for comparator If safer, index test is potential replacement for comparator If not as safe, there is a trade-off between safety and effectiveness – potentially used in addition to comparator [b] or alternative to comparator

a: Cost-effectiveness is another consideration once the decision on safety and effectiveness is made. b: In those patients who tested negative on comparator test.

[2] Decision algorithm developed by Tracy L. Merlin (2007).

Review of literature

The medical literature was searched to identify relevant studies and reviews for the period from 1990 (or if inception of the database was later, from that date) until January–February 2007. Search alerts were maintained over the duration of the review to identify any new major research published and indexed in the major databases since the completion of the search. In such an event, the full search would have been updated. Appendix D describes the electronic databases that were used for this search and other sources of evidence that were investigated. Grey literature[3] was included in the search strategy. Unpublished literature, however, was not canvassed as it is difficult to search for this literature exhaustively and systematically, and trials that are difficult to locate are often smaller and of lower methodological quality (Egger et al 2003). It is, however, possible that these unpublished data could alter the results of this assessment.

The search terms, presented in Appendix D, were used to identify literature on the safety, effectiveness and cost-effectiveness of using DM for the screening, surveillance and diagnosis of breast cancer.

Inclusion/exclusion criteria

In general, studies were excluded if they:

- did not address the research question
- did not provide information on the pre-specified target population, eg, were phantom or computer modelling studies, or outside the age range
- did not include one of the pre-specified interventions; for reports that indicated mammography was performed but did not state which type, the conservative assumption was made that the report related to conventional FM
- did not compare results with the pre-specified comparator
- did not address one of the pre-specified outcomes or provided inadequate data on these outcomes (in some instances, a study was included to assess one or more outcomes but had to be excluded for other outcomes owing to data inadequacies)
- were written in other languages and gave a lower level of evidence than that available in English
- did not have the appropriate study design.

Where two or more papers reported on different aspects of the same study, such as the methodology in one and the findings in the other, they were treated as one study. Similarly, if the same data were presented in multiple articles, results from the most comprehensive or most recent article only were included.

---

[3] Literature that is difficult to find, including published government reports, theses, technical reports and non-peer-reviewed literature.

The criteria for including studies in this evaluation are presented in relevant areas of the Results section. The criteria relevant to determining the safety of DM for screening, surveillance and diagnosis of breast cancer are presented in Box 2. The criteria for including studies relevant to determining the direct screening effectiveness of DM are presented in

Box 4. The criteria for evaluating the linked evidence components in the assessment of the value of DM as a screening tool are presented in Box 3, Box 5 and Box 9. Similarly, the criteria for evaluating the linked evidence for the effectiveness of DM for surveillance of women at potentially high risk and for diagnosis of symptomatic women are presented in Box 6, Box 8 and Box 9.

Search results

The process of study selection for this report went through five phases:

1. All reference citations from all literature sources were collated into an Endnote 8.0 database.

2. Duplicate references were removed.

3. Studies were excluded on the basis of the citation information if it was obvious that they did not meet the inclusion criteria. Citations were assessed independently by two reviewers. Studies marked as requiring further evaluation by either reviewer were retrieved for full-text assessment (after discussion).

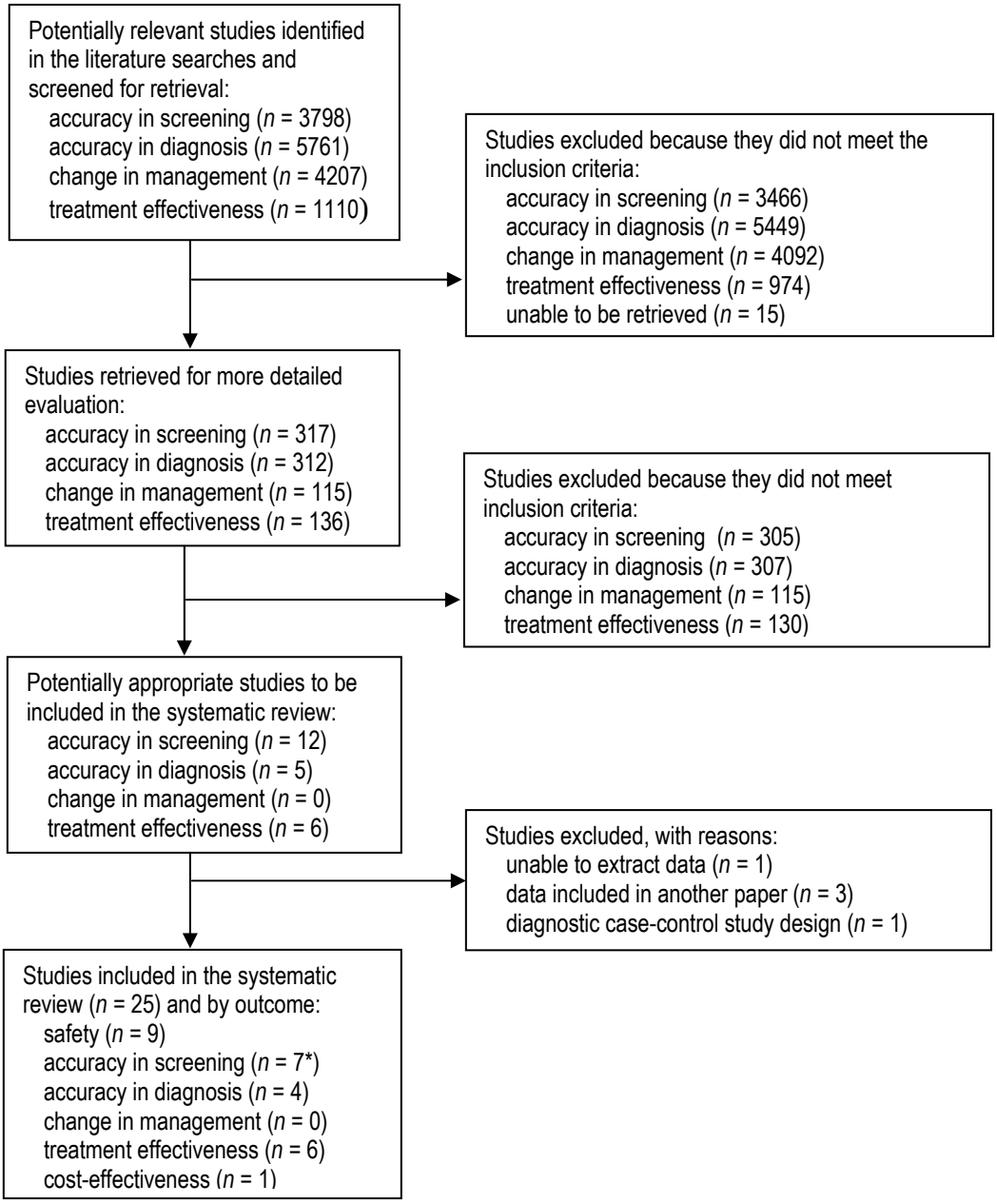
4. Studies were included if they met the criteria according to two reviewers who independently read the full-text articles. Those articles meeting the criteria formed part of the evidence base. The remainder provided background information.

5. The reference lists of the included articles were pearled for additional relevant studies. These were retrieved and assessed according to phase 4.

The evidence base consisted of articles from phases 4 and 5 that met the inclusion criteria.

Any doubt concerning inclusions at phase 4 was resolved by consensus between the two reviewers. A third reviewer was included to arbitrate where necessary. The results of the process of study selection are provided in Figure 11.

Figure 11 Summary of the process used to identify and select studies for the assessment of DM for screening effectiveness in an asymptomatic population



* Three citations were publications that provided interim, follow-up or additional information on another included study. There were four studies overall.

Adapted from (Moher et al 1999)

Data extraction and analysis

A profile of key characteristics was developed for each included study (Appendix E). Each study profile described the level of evidence, design and quality of the study, authors, publication year, location, study population characteristics, type of intervention (ie, type of DM and equipment used; assessment technique used, eg, double reader or single reader + CAD[4]), comparator

[4] CAD = computer-assisted diagnosis.

intervention (where relevant), reference standard, inclusion/exclusion criteria, outcomes assessed and follow-up period.

Studies that could not be retrieved or that met the inclusion criteria but contained insufficient or inadequate data for inclusion are listed in Appendix F. Definitions of all technical terms and abbreviations are provided in the Glossary. Descriptive statistics were extracted or calculated for all safety and effectiveness outcomes in the individual studies.

Assessing diagnostic accuracy

FM has already been shown to reduce mortality as a consequence of early detection of breast cancer in screening studies. Therefore, comparisons between DM and FM do not necessarily need to assess differences in patient survival (direct evidence) in order to determine whether one is more effective than the other. Cancer detection rate is a reasonable surrogate outcome measure (Irwig et al 2004).

To calculate the diagnostic accuracy of DM for the purposes of determining whether it should *replace* or *supplement* FM, the basic study design must require women to independently receive both DM and FM (Irwig et al 2004). The aim is to calculate the cancer detection rate (sensitivity) and false positive rate (1 – specificity) of each test, with reference to the standard of histopathology, and to determine how these rates differ between the tests.

To assess the diagnostic accuracy of each of the tests, where possible the sensitivity, specificity, negative and positive predictive values (NPV, PPV) of the tests, false negative and false alarm rates, and 95 per cent confidence intervals (CIs) were calculated. Data were extracted into a classic 2 × 2 table, in which the results of the index diagnostic test were cross-classified against the results of the reference standard (Armitage et al 2002; Deeks 2001), and Bayes' Theorem was applied:

		Breast cancer status (based on reference standard: histopathology)		
		Disease +	*Disease –*	
Index test *(DM or FM)*	*Test +*	true positive	false positive	Total test positive
	Test –	false negative	true negative	Total test negative
		Total with breast cancer	Total without breast cancer	

The **sensitivity** of the index test (DM or FM) was calculated as the proportion of women with breast cancer who have positive diagnostic test results:

Sensitivity (true positive rate) = true positive / total with breast cancer

The **specificity** of the index test (DM or FM) was calculated as the proportion of women without breast cancer who have normal diagnostic test results:

Specificity (true negative rate) = true negative / total without breast cancer

When a 95 per cent CI was not provided, this was calculated by exact binomial methods.

In some studies in the evidence base, DM was compared with FM rather than with the reference standard (histopathology) for the assessment of diagnostic accuracy. Although these data were extracted and included in the Results section of this report, they must be interpreted with caution, as FM is an imperfect reference standard, and so we cannot be certain that positive

results reported from DM, but not FM, are a consequence of a false positive reading or actually show improved sensitivity.

False positive rate = false positive / total without breast cancer (determined by histopathology)

False negative rate = false negative / total with breast cancer (determined by histopathology)

In breast cancer screening and surveillance studies it is often difficult to determine the true disease status of women testing negative on the index test as it requires long-term follow-up of their health. Long-term follow-up is necessary because there are ethical issues associated with doing an invasive test (histopathology from a biopsy) in asymptomatic, test negative women. Even then, the disease may have developed during follow-up rather than before the index test was conducted. Thus, an accurate false negative rate is difficult to calculate as the correct numerator and denominator may not be known, whilst an accurate false positive rate is difficult to determine as the correct denominator is often unclear.

As a consequence of the inability to determine the true disease status of the women in these studies, rates dependent on the test results alone are often reported:

False alarm rate = false positive / total test positive

False reassurance rate = false negative / total test negative

As FM is an imperfect reference standard it will fail to identify some cancers, and so a different study design is required to determine the differential impact of DM or FM on the **interval cancer rate** in breast cancer screening and surveillance populations (Irwig et al 2004). To assess the difference in interval cancer rates between DM and FM in screening and surveillance would require women to be randomised to the different tests (as was done when mammography was compared with no screening) rather than receiving both tests.

In screening and surveillance studies the cancer detection rate at a particular point in time is often reported:

Cancer detection rate = number of cancers detected / total population screened, at a point in time

Meta-analysis could not be conducted owing to the heterogeneous nature of the available evidence base assessing the diagnostic accuracy of DM for women at potentially high risk, symptomatic women and asymptomatic women. A narrative meta-synthesis of the data was therefore undertaken, with discussion of relevant results concerning those subgroups delineated beforehand.

All statistical calculations were undertaken using the biostatistical computer package Stata version 9.0 (Stata Corporation 2005).

Appraisal of the evidence

The evidence presented in the selected studies was assessed and classified using the dimensions of evidence defined by the National Health and Medical Research Council (NHMRC 2000).

These dimensions (Table 7) consider important aspects of the evidence supporting a particular intervention and include three main domains: strength of the evidence, size of the effect and relevance of the evidence. The first domain is derived directly from the literature identified as

informing a particular intervention. The last two require expert clinical input as part of their determination.

Table 7 Dimensions of evidence

Type of evidence	Definition
Strength of evidence	
Level	Study design used, as an indicator of the degree to which bias has been eliminated by design (Table 9).
Quality	The methods used by investigators to minimise bias within a study design.
Statistical precision	The *P*-value or, alternatively, the precision of the estimate of the effect. It reflects the degree of certainty about the existence of a true effect.
Size of effect	The distance of the study estimate from the 'null' value and the inclusion of only clinically important effects in the confidence interval.
Relevance of evidence	The usefulness of the evidence in clinical practice, particularly the appropriateness of the outcome measures used.

The three subdomains (level, quality and statistical precision) are collectively a measure of the strength of the evidence. With specific regard to diagnostic evidence, the individual studies assessing diagnostic effectiveness were graded according to the quality and applicability criteria (MSAC 2005) shown in Table 8. The designations of the levels of evidence are shown in Table 9. Study quality was assessed using the checklists shown in Table 10.

Table 8 Grading system used to rank included diagnostic studies

Validity criteria	Description	Grading system
Appropriate comparison	Did the study evaluate a direct comparison of the index test strategy versus the comparator test strategy?	C1 direct comparison CX other comparison
Applicable population	Did the study evaluate the index test in a population that is representative of the subject characteristics (age and sex) and clinical setting (disease prevalence, disease severity, referral filter and sequence of tests) for the clinical indication of interest?	P1 applicable P2 limited P3 different population
Quality of study	Was the study designed to avoid bias? High quality = no potential for bias based on predefined key quality criteria Medium quality = some potential for bias in areas other than those pre-specified as key criteria Poor quality = poor reference standard and/or potential for bias based on key pre-specified criteria	Study design: NHMRC level of evidence Study quality (QUADAS checklist): Q1 high quality (≥12/14) Q2 medium (10–11/14) Q3 poor reference standard, poor quality (<10/14) or insufficient information

Table 9 Designations of levels of evidence* according to type of research question (NHMRC, 2005)

Level	Intervention §	Diagnosis **	Screening
I *	A Systematic review of level II studies	A systematic review of level II studies	A Systematic review of level II studies
II	A Randomised controlled trial	A study of test accuracy with: an independent, blinded comparison with a valid reference standard, §§ among consecutive patients with a defined clinical presentation ††	A Randomised controlled trial
III-1	A Pseudorandomised controlled trial (i.e. alternate allocation or some other method)	A study of test accuracy with: an independent, blinded comparison with a valid reference standard, §§ among non-consecutive patients with a defined clinical presentation††	A Pseudorandomised controlled trial (i.e. alternate allocation or some other method)
III-2	A Comparative study with concurrent controls: Non-randomised, experimental trial † Cohort study Case-control study Interrupted time series with a control group	A comparison with reference standard that does not meet the criteria required for level II and III-1 evidence	A Comparative study with concurrent controls: Non-randomised, experimental trial Cohort study Case-control study
III-3	A Comparative study without concurrent controls: Historical control study Two or more single arm study ‡ Interrupted time series without a parallel control group	Diagnostic case-control study ††	A Comparative study without concurrent controls: Historical control study Two or more single arm study
IV	Case series with either post-test or pre-test/post-test outcomes	Study of diagnostic yield (no reference standard) ‡‡	Case series

* A systematic review will only be assigned a level of evidence as high as the studies it contains, excepting where those studies are of level II evidence.

§ Definitions of these study designs are provided on pages 7–8 How to use the evidence: assessment and application of scientific evidence (NHMRC 2000b).

** The dimensions of evidence apply only to studies of diagnostic accuracy. To assess the effectiveness of a diagnostic test there also needs to be a consideration of the impact of the test on patient management and health outcomes. See (MSAC 2005)

§§ The validity of the reference standard should be determined in the context of the disease under review. Criteria for determining the validity of the reference standard should be pre-specified. This can include the choice of the reference standard(s) and its timing in relation to the index test. The validity of the reference standard can be determined through quality appraisal of the study. See (Whiting P 2003)

†† Well-designed population based case-control studies (eg. population based screening studies where test accuracy is assessed on all cases, with a random sample of controls) do capture a population with a representative spectrum of disease and thus fulfil the requirements for a valid assembly of patients. However, in some cases the population assembled is not representative of the use of the test in practice. In diagnostic case-control studies a selected sample of patients already known to have the disease are compared with a separate group of normal/healthy people known to be free of the disease. In this situation patients with borderline or mild expressions of the disease, and conditions mimicking the disease are excluded, which can lead to exaggeration of both sensitivity and specificity. This is called spectrum bias because the spectrum of study participants will not be representative of patients seen in practice.

† This also includes controlled before-and-after (pre-test/post-test) studies, as well as indirect comparisons (ie. utilise A vs B and B vs C, to determine A vs C).

‡ Comparing single arm studies ie. case series from two studies.

‡‡ Studies of diagnostic yield provide the yield of diagnosed patients, as determined by an index test, without confirmation of the accuracy of this diagnosis by a reference standard. These may be the only alternative when there is no reliable reference standard.

Note 1: Assessment of Comparative harms/safety should occur according to the hierarchy presented for each of the research questions, with the proviso that this assessment occurs within the context of the topic being assessed. Some harms are rare and cannot feasibly be captured within randomised controlled trials; physical harms and psychological harms may need to be addressed by different study designs; harms from diagnostic testing include the likelihood of false positive and false negative results; harms from screening include the likelihood of false alarm and false reassurance results.

Note 2: When a level of evidence is attributed in the text of a document, it should also be framed according to its corresponding research question eg. level II intervention evidence; level IV diagnostic evidence.

Table 10 Quality checklists

Study type	Checklists
Systematic reviews, health technology assessment reports	NHMRC Checklist Table 1.4 (NHMRC 2000)
Randomised controlled trials	NHMRC Checklist Table 1.4 (NHMRC 2000)
Cohort study	NHMRC Checklist Table 1.4 (NHMRC 2000)
Case-control	NHMRC Checklist Table 1.4 (NHMRC 2000)
Diagnostic test cross-sectional study	QUADAS quality assessment tool (Whiting P 2003)
Intervention case series	NHS CRD quality assessment scale (Khan et al 2001) (Box 5.9)

Statistical precision

Statistical precision was determined by using standard statistical principles. Small CIs and low *P*-values suggest that a reported effect is real (NHMRC 2000).

Size of effect in individual studies

It is important to establish whether statistically significant differences are also clinically important. The size of the effect needs to be determined, as well as whether the 95 per cent CI includes only clinically important effects (NHMRC 2000).

Relevance of evidence in individual studies

Similarly, the outcome being measured in the studies should be appropriate and clinically relevant. Inadequately validated (predictive) surrogate measures of a clinically relevant outcome should be avoided (NHMRC 2000). Cancer detection rate was considered a valid surrogate outcome.

Assessment of the body of evidence

Once the results of the studies were synthesised, the overall conclusion as derived from the body of evidence (Table 11) was presented to answer each clinical question – see Discussion section.

Table 11 Body of evidence assessment matrix

Component	A Excellent	B Good	C Satisfactory	D Poor
Evidence base	Several level I or II studies with low risk of bias	one or two level II studies with low risk of bias or an SR/multiple level III studies with low risk of bias	Level III studies with low risk of bias, or level I or II studies with moderate risk of bias	Level IV studies, or level I to III studies with high risk of bias
Consistency	All studies consistent	Most studies consistent, and inconsistency may be explained	Some inconsistency reflecting genuine uncertainty around clinical question	Evidence is inconsistent
Clinical impact	Very large	Substantial	Moderate	Slight or restricted
Generalisability	Population/s studied in body of evidence are the same as the target population	Population/s studied in the body of evidence are similar to the target population	Population/s studied in body of evidence different to target population but it is clinically sensible to apply this evidence to target population	Population/s studied in body of evidence different to target population and hard to judge whether it is sensible to generalise to target population
Applicability	Directly applicable to Australian health care context	Applicable to Australian health care context with few caveats	Probably applicable to Australian health care context with some caveats	Not applicable to Australian health care context

SR = systematic review

Expert advice

An advisory panel with expertise in oncology, surgery, radiography, radiology, screening, diagnostics, medical administration and consumer issues was established to evaluate the evidence and provide advice to MSAC from a clinical perspective. In selecting members for advisory panels, MSAC's practice is to approach the appropriate medical colleges, specialist societies and associations, and consumer bodies for nominees. Membership of the advisory panel is listed in Appendix B.

Results of assessment

Is digital mammography safe?

Summary – Safety of digital mammography

A total of nine studies reported on the safety of digital mammography.

Three studies compared radiation exposure levels between the two imaging modalities. DM using the Senographe 2000D mammogram unit was estimated to use very similar levels of radiation exposure to FM. One study found that radiation levels could be reduced to approximately half the amount used with the standard AOP, while still maintaining image quality.

Two studies provided enough information to determine false positive rates. The better-quality Oslo I study found that DM results in slightly more false positives than FM, although the statistical significance of this difference was unable to be calculated. Conversely, the Colorado Screening Trial found that DM had lower rates of false positives. One further study provided sufficient information to calculate false alarm rates, which were not significantly different between the two types of mammography.

Overall, DM appears as safe as FM.

The safety of DM was assessed from studies identified through predetermined criteria shown in Box 2.

A total of nine studies (eleven citations) were identified that provided information on the safety of DM. Safety outcomes reported were either radiation exposure or unnecessary intervention or harm from a false alarm or a false positive.

Box 2 Inclusion criteria for identification of studies relevant to an assessment of safety of DM

Characteristic	Criteria
Publication type	Randomised or non-randomised controlled trials, cohort studies, registers, case series, case reports or systematic reviews of these study designs. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded
Patient	1. Asymptomatic women aged 40 years and over attending a breast screening program Subgroups: 40–49 y, 50–69 y, 70+ y 2. Women diagnosed with breast cancer who presented with: symptoms or a clinical abnormality of the breast(s) suggesting malignancy OR a potentially high risk of breast cancer due to: 1. Potentially high risk of ovarian cancer 2. two 1° or 2° relatives on one side of the family diagnosed with breast or ovarian cancer plus one or more of the following features on the same side of the family: – additional relative(s) with breast or ovarian cancer – breast cancer diagnosed before the age of 40 – bilateral breast cancer – breast and ovarian cancer in the same woman – Ashkenazi Jewish ancestry – breast cancer in a male relative 3. one 1° or 2° relative diagnosed with breast cancer at age 45 or younger plus another 1° or 2° relative on the same side of the family with sarcoma (bone or soft tissue) at age 45 or younger 4. being a member of a family in which the presence of a high-risk breast cancer gene mutation has been established, eg, *BRCA1*, *BRCA2*, *Tp53* 5. personal history of breast cancer 6. premalignant conditions, ie, lobal carcinoma *in situ* and atypical ductal hyperplasia Subgroups: (1) <50 y v. 50+ y [a] (2) very to extremely dense breast tissue v. less dense tissue (3) pre- or perimenopausal v. postmenopausal [a]
Intervention or test	1. Screening DM, consisting of 2 full-field x-ray views of each breast 2. Diagnostic DM, consisting of additional full-field x-ray views of each breast ± breast ultrasound scan ± breast MRI
Comparator	1. Screening FM, consisting of 2 x-ray views of each breast 2. Diagnostic FM, consisting of additional x-ray views of each breast ± breast ultrasound scan ± breast MRI
Outcome	Physical harms from mammography (eg, adverse events, radiation exposure); physical harms from unnecessary further intervention (due to false positive or alarm)
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

MRI = magnetic resonance imaging. a: These subgroups are surrogate measures for breast tissue density, as density can be measured only via mammogram.

Radiation risk

The glandular tissue of the breast is sensitive to radiation, and there is some evidence that high doses of radiation may induce breast cancer (Faulkner 2005). The onset of cancer is typically at least 10 years after the radiation exposure (Faulkner 2005). For this reason, although patients may not experience any immediate negative effects from radiation, it is important to keep the level of radiation as low as reasonably achievable to minimise the carcinogenic risk (Samei et al 2007).

The radiation dose within FM is determined by the amount of radiation required to achieve the correct exposure on the analogue film (Gennaro et al 2004; Samei et al 2007). DM has the advantage of separating the detection and display functions. This allows the image to be manipulated to enhance the contrast after images have been acquired (Samei et al 2007). In DM,

the patient radiation dose is therefore determined by the signal-to-noise ratio rather than by the contrast levels required (Gennaro et al 2004).

Most radiation doses cannot be measured directly (except for skin doses), so are estimated on the basis of phantoms matched to clinical parameters. Studies that assessed radiation doses without a clinical population as a basis for parameters such as breast thickness and glandularity were excluded.

Seven studies estimated the radiation exposure from DM, and three of these compared the radiation dose to that in FM (see Table 12). Radiation doses were reported as either average glandular dose (AGD) or entrance surface air kerma (ESAK). AGD is deduced by converting the ESAK by a factor *g*, which largely depends on the thickness and density of the breast (Hermann et al 2002). Air kerma was measured as an exponential function of the tube voltage (Gennaro et al 2004). All of the studies that provided estimates of radiation dose from DM used the Senographe 2000D FFDM system (GE Medical Systems). The level of radiation exposure associated with DM was determined from the automatic optimisation of parameters (AOP) operating mode, automated on the basis of pre-exposure (Gennaro & di Maggio 2006). There are three different AOP modes, depending on the priority: contrast, dose reduction or a combination of both (standard mode) (Hermann et al 2002).

Using the standard automated mode of operation, one Italian study (Gennaro & di Maggio 2006) reported that women undergoing DM were exposed to less radiation than those who underwent FM. Contrary to this, a randomised trial (Fischmann et al 2005) reported that women received a higher radiation dose from DM than from FM, although this difference was not statistically significant. Moran et al (2005) reported that DM resulted in less radiation exposure than slow-screen mammography in a symptomatic population, but higher exposure than fast-screen mammography in an asymptomatic population.

Gennaro et al (2004) compared the radiation exposure levels between four different Italian sites, two of which used the standard mode of AOP (sites 1 and 2), and two of which preferred the contrast enhancing mode (sites 3 and 4). The mean radiation levels estimated from the standard mode were 17 to 28 per cent lower than in the contrast enhancing mode (Gennaro et al 2004).

Hemdal et al (2005a) investigated whether image quality could be retained with a lower glandular dose. Twenty-eight women who underwent routine mammographic screening had one breast imaged with standard AOP DM techniques (using automatically determined tube loading, anode–filter combination and tube potential, based on breast thickness and density), and the other imaged without AOP, using a manual mode, with the same anode–filter combination and tube potential as the first breast, but with approximately half the tube loading. Observers were able to detect a difference between the standard dose and half-dose images, but considered that all of the reduced-dose images were of sufficient quality to diagnose or detect breast cancer (Hemdal et al 2005a).

All studies reported AGDs lower than the NHMRC diagnostic reference level for FM of ≤2 mGy and the BreastScreen requirement that AGDs be kept below 3 mGy (McLean et al 2007).

Table 12 Safety results: physical harms from mammography due to radiation exposure

Study location	Study design quality	Study population	Radiation exposure (mGy): mean ± SD (range)	
			DM	FM
(Fischmann et al 2005) Tübingen, Germany	Randomised controlled trial Level II interventional evidence 1/3 (NHMRC Quality Assessment Scale) Not double blinded Allocation not concealed 199/200 women included	*N* = 199 females Asymptomatic population Age ≥40 y	AGD MLO 1.70 ± 0.40 (1.00–3.16) CC 1.63 ± 0.38 (0.85–3.36)	AGD MLO 1.52 ± 0.58 (0.46–3.49) CC 1.45 ± 0.55 (0.35–3.60)
(Moran et al 2005) Madrid, Spain	Cohort study Level III-2 interventional evidence 2/6 (NHMRC Quality Assessment Scale)	*N* = 27 545 images DM = 20137 images (5034 females) Symptomatic population Mean age = 56 ± 11 y FM = 7401 images Screening population	AGD [mean (range)] 1.88 (0.4–8.1) ESAK [mean (range)] 8.1 (1.8–44)	ESAK [mean] *Slow screen type* 11.3 *Fast screen type* 5.9
(Gennaro & di Maggio 2006) Padua, Italy	Cohort study Level III-2 interventional evidence 2/6 (NHMRC Quality Assessment Scale)	*N* = 596 Population not stated Age not stated DM breast thickness = 48.5 mm FM breast thickness = 45.3 mm	AGD 1.40 ESAK 6.06	AGD 1.92 ESAK 8.25
(Yamada et al 2004) Sendai, Japan	Case series Level IV interventional evidence 5/6 (NHS CRD Quality Assessment Scale)	*N* = 480 females Mixed asymptomatic/ symptomatic population Age ≥50 y	AGD 1.48	n/a
(Hemdal et al 2005a) Malmö, Sweden	Case series Level IV interventional evidence 4/6 (NHS CRD Quality Assessment Scale)	*N* = 28 females; 56 images Asymptomatic population Mean age = 62 (52–74) y	AGD 1st breast [a] 1.58 (1.27–1.94) AGD 2nd breast [b] 0.75 (0.57–0.93)	n/a
(Gennaro et al 2004) Padua, Italy	Case series Level IV interventional evidence 3/6 (NHS CRD Quality Assessment Scale)	*N* = 800 images Population n/a Age n/a	AGD [c] Site 1, 1.25 ± 0.24 Site 2, 1.37 ± 0.55 Site 3, 1.57 ± 0.43 Site 4, 1.73 ± 0.45 AGD [d] Site 1, 1.37 ± 0.26 Site 2, 1.49 ± 0.60 Site 3, 1.71 ± 0.47 Site 4, 1.89 ± 0.49 ESAK Site 1, 5.90 ± 1.79 Site 2, 6.00 ± 2.70 Site 3, 7.09 ± 2.28 Site 4, 8.20 ± 2.78	n/a
(Hermann et al 2002) Göttingen, Germany	Case series Level IV interventional evidence 3/6 (NHS CRD Quality Assessment Scale)	*N* = 591 females; 1116 images Mixed population Mean age = 55 ± 10 (34–81) y	AGD 1.51 ± 0.34 (0.66–4.05) ESAK 7.05 ± 2.48 (1.62–25.65)	n/a

DM = digital mammography; FM = film-screen mammography; ESAK = entrance surface air kerma; AGD = average glandular dose; y = years; n/a = not applicable; MLO = mediolateral oblique; CC = craniocaudal; AOP = automated optimisation of parameters; STD = standard. a: AOP/STD mode, including pre-exposure. b: Manual mode, no pre-exposure, using same anode–filter combination and tube potential, but only about half of the tube loading used for the first breast. c: Assuming a breast composition of 50% adipose, 50% glandular tissue. d: Assuming a breast composition of 70% adipose, 30% glandular tissue.

False positive rates and false alarm rates

Women who are recalled owing to suspected breast cancer undergo further investigations such as additional imaging and biopsy. If these investigations cannot find an abnormality or they find a lump that is benign, the woman may have undergone unnecessary psychological distress and potential physical harm. While no physical harms were reported in any of the studies identified, three studies reported the number of women falsely identified through DM and FM as potentially having cancer (Table 13 and Table 14).

False positive rates indicate the proportion of women who do not have breast cancer but who have undergone further investigations unnecessarily on account of a positive test result. False positive rates are able to be calculated only when the true disease status of all women in a study is known. It would be unethical to require women who had a negative mammogram to undergo a biopsy, so the majority of studies included in this review were not able to ascertain the true disease status of the whole population, except those that traced interval cancers on a population basis, with 100 per cent surveillance (such as the Oslo I study) (Skaane et al 2005; Skaane et al 2003). This good-quality study found that DM resulted in a higher false positive rate than FM (Table 13). The statistical significance of this difference could not be calculated as the data for DM and FM were non-independent (both types of mammography were conducted on the same individuals). Statistical testing could be conducted only on patient-level data, which were not available. A second study (Lewin et al 2002) provided information on the disease status of all study participants, but did not specify the reference standard. This study found minimal differences between FM and DM in false positive rates.

Table 13 Safety results: physical harms from unnecessary further intervention (due to false positive)

Study location	Study design quality	Study population	False positives	
			DM rate (%)	FM rate (%)
(Skaane et al 2005; Skaane et al 2003) Oslo, Norway Oslo I study	Level III-2 diagnostic evidence CX, P1 Q1 (QUADAS = 12/14)	*N* = 3683 Asymptomatic population Mean age = 58.2 y [50–69]	16.1% (585/3626)	11.3% (411/3626)
(Lewin et al 2002) (Lewin et al 2001) Colorado screening study I	Level III-2 diagnostic evidence CX, P1 Q2 (QUADAS = 10/14)	*N* = 4489 females 6736 paired examinations Asymptomatic population Mean age = 55.6 y	11.5% (768/6694)	14.5% (968/6694)

DM = digital mammography; FM = film-screen mammography

The rate of false alarms is calculated from the number of women who were falsely identified as having cancer as a proportion of all women who tested positive on a mammogram. The true disease status is determined by a reference standard such as histopathology from a biopsy. However, as not all women receive this reference standard (ie, those women who tested negative on the mammogram), there is only partial verification of the diagnostic accuracy of mammography in the study population. This measure, therefore, is not as robust as the false positive rate as an intermediate measure of unnecessary further intervention and potential harm.

One study was unable to provide the rate of false positives, but was able to provide sufficient information to compare the false alarm rates of DM and FM (Table 14). There was a slightly higher rate of false alarms from DM than from FM, but the difference was not statistically significant (Yamada et al 2004).

Table 14 Safety results: physical harms from unnecessary further intervention (due to false alarm)

Study location	Study design quality	Study population	False alarm		
			DM rate (%)	FM rate (%)	Difference (*P* value)
(Yamada et al 2004) Sendai, Japan	Level III-2 diagnostic evidence CX, P1 Q3 (QUADAS = 7/14)	480 females *N* = 40 tested positive on DM *N* = 28 tested positive on FM Mixed asymptomatic/ symptomatic population Age ≥50 y	90% (36/40)	85.7% (24/28)	*P* = 0.69 *(Fisher's exact test)*

DM = digital mammography; FM = film-screen mammography

Is digital mammography effective in a screening population?

Summary – Linked evidence of diagnostic effectiveness of screening DM

Four studies provided information on the diagnostic accuracy and screening effectiveness of DM in comparison with FM in a screening or asymptomatic population. Two of the studies were of high quality, one of medium quality and one of poor quality.

Differences in accuracy and cancer detection between these studies were observed. In the initial studies DM had poorer diagnostic accuracy and lower cancer detection rates than FM, although the differences were not statistically significant. In the later studies DM performed better than FM in diagnostic accuracy and cancer detection rate, although overall the differences were again not statistically significant.

The observed differences between the studies appear to relate to study quality, the rate at which radiologists learn to read the digital display, developments in the digital technology and the display, statistical power of the studies, and different screening protocols between the studies.

The largest and highest-quality study with adequate statistical power determined that DM had similar overall diagnostic accuracy to that of FM. However, DM also had significantly better accuracy in pre-specified subgroups of women, specifically those aged under 50 years, or pre- or perimenopausal, or with heterogeneously dense or extremely dense breasts. The cancer detection rate in these subgroups was higher when imaged by DM.

The relative impact of DM and FM on interval cancer rates has yet to be properly established.

The evidence base was examined to compare the diagnostic accuracy of DM and FM with reference to the diagnostic standard of histopathology. Studies were included if they met the criteria outlined in Box 3 or if they assessed the screening effectiveness of the two methods (Box 4).

Box 3 Inclusion criteria for identification of studies relevant to an assessment of diagnostic accuracy of DM in a screening (asymptomatic) population

Characteristic	Criteria
Publication type	Cross-sectional studies where patients are cross-classified on the index test and comparator and/or reference standard. Case-control diagnostic studies were acceptable only if cross-sectional studies were not available. Systematic reviews of cross-sectional studies. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded.
Patient	Asymptomatic women aged 40 years and over attending a breast screening program Subgroups: 40–49 y, 50–69 y, 70+ y
Intervention/test	Screening DM, consisting of 2 full-field x-ray views of each breast
Comparator	Screening FM, consisting of 2 x-ray views of each breast
Reference standard	Histopathology
Outcome	Sensitivity and specificity (and therefore rates of false positives and negatives), positive and negative likelihood ratios, positive and negative predictive values (and therefore false alarm and reassurance rates), diagnostic odds ratios, ROC curves, AUC, accuracy
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

Box 4 Inclusion criteria for identification of direct evidence studies concerning the effectiveness and cost-effectiveness of DM for screening

Characteristic	Criteria
Publication type	Effectiveness: Randomised or non-randomised controlled trials or cohort studies or systematic reviews of these study designs. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded. Cost-effectiveness: Economic studies, modelling, economic analyses
Patient	Asymptomatic women aged 40 years and over attending a breast screening program Subgroups: 40–49 y, 50–69 y, 70+ y
Intervention/test	Screening DM, consisting of 2 full-field x-ray views of each breast
Comparator	Screening FM, consisting of 2 x-ray views of each breast
Outcome	Effectiveness: Primary – *Patient-relevant*: reduction in mortality, quality of life; *Surrogat*e: interval cancers, cancer (invasive or ductal carcinoma *in situ*) detection rates, tumour size or stage Secondary – recall rates, re-attendance rates, biopsy rates, false positive rates, false negative rates, image quality or interpretation, duration of procedure, delayed intervention (due to false negative) Cost-effectiveness: Cost, incremental cost-effectiveness ratio (eg, cost/LYG, QALY, DALY), workforce issues
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

LYG = life years gained; QALY = quality-adjusted life year; DALY = disability-adjusted life year

Included studies

Four studies (seven papers: two each from three studies) were included for the assessment of diagnostic effectiveness of DM in a screening population. Three of the studies pertained to the diagnostic accuracy of DM in this population, and one assessed the screening effectiveness of DM (although only by surrogate outcome measures).

Colorado screening study (reports I and II)

The Colorado screening study (Lewin et al 2001; Lewin et al 2002) was the first to compare DM with FM in a screening or asymptomatic population. It was also the first to attempt full verification of cancer status in the study cohort (although, as reported in Lewin 2001, this could be determined for only 59% of the cohort). The study enrolled 4489 women but included 1665 women who enrolled twice and 291 women who enrolled three times, meaning that some of the data were analysed in terms of examinations rather than individuals. Women received both FFDM (prototype indirect flat-panel detector without the automated processing algorithms that are now available) and FM. Most women (91%) received FM first followed immediately by DM; otherwise, the DM occurred within 3 days. CC and MLO views were taken. Two radiologists, one per method, then took a single reading each. A positive reading was indicated by a BIRADS[5] score of 0, 2, 3, 4 or 5. Concordant readings between the radiologists resulted in immediate work-up. Discordant readings were discussed at a recall meeting at which both types of mammogram were assessed. The rate of findings detected by DM and dismissed at this meeting was higher (12.6%) than that of FM findings dismissed (1.9%) (Lewin et al 2001). Lewin et al (2001) gave interim results ('Colorado I') and Lewin et al (2002) gave the final results ('Colorado II'). The study was of average methodological quality.

[5] Breast Imaging Reporting and Data System. See Box 7 for full BIRADS classification system.

Oslo I study

The Oslo I study was conducted by Skaane et al in 2000 on 37 per cent of the asymptomatic women attending the Norwegian Breast Cancer Screening Program (*n* = 3683). The women received both FFDM (indirect flat-panel detector) and FM, with two views (CC and MLO) of each breast taken. Radiologists independently double-read each type of mammogram. A positive reading for each type of mammogram was recorded when at least one of the two readers gave a score of 2 or higher on a 5-point scale (1 = normal or definitely benign, 2 = probably benign, 3 = indeterminate finding, 4 = probably malignant, 5 = malignant). Women were recalled for a diagnostic work-up on the basis of a score of 3 or higher, or when four radiologists gave a score of 2 at a consensus meeting. Diagnostic work-up occurred within 2 weeks of the consensus meeting and included one or more of spot compression, magnification views, ultrasound, MRI and FNA cytology. Data on breast cancers immediately identified at study screening were collected (Skaane et al 2003). As data on the women attending the Norwegian Breast Cancer Screening Program are linked to the Norwegian Cancer Registry, at follow-up 2½ years later (after the 2-yearly screening interval), data were collected on interval cancers and on cancers identified at the subsequent screening round for those women involved in the study. Radiologists assessed the interval and subsequent screening-round cancers to determine whether they were visible on the original study mammograms or were true incident cancers (Skaane et al 2005). The incident cancers were excluded from the analysis.

Oslo II study

The Oslo II study was a screening effectiveness study in which women were randomised to either DM (*n* = 12 473) or FM (*n* = 30 956) at invitation to screening (Skaane & Skjennald 2004). Of those invited, 25 263 women attended screening, 7209 for DM and 18 054 for FM. This is the intention-to-screen population. However, as it appears that allocation to screening method was not concealed in this trial – women were allocated on the basis of the last digit of the screening invitation letter – and women and staff were not blinded to the mammographic method allocated, a large proportion of crossovers (*n* = 351) were observed. These crossovers included 102 women who were randomised to FM but were offered DM because the staff considered it a superior diagnostic technique for women with breast implants. Some crossovers occurred because women were unwilling to undergo 'ordinary' mammography, and in other cases the allocated equipment was not available. The authors analysed the trial data as per protocol (ie, excluding the crossovers) as a result. An intention-to-screen analysis for cancer detection has also been conducted in the evaluation. Despite this study's limitations, if the authors were to conduct a follow-up study that included interval cancer rates, this would provide the best evidence to date of the difference in interval cancer detection between the two screening methods (currently other studies are confounded by the application of both methods to the same women).

DMIST study

One of the largest breast cancer screening studies performed was conducted in 33 academic and community practices in the United States and Canada. The Digital Mammography Imaging Screening Trial, known as the DMIST study, compared DM with FM in the same women (Pisano et al 2005b). The high-quality study recruited 49 528 women, of whom 42 760 contributed data to the analysis. The strength of this study was its applicability to current diagnostic practice, comparing several different types of DM and FM units. The method of digital display – soft-copy or hard-copy display – also differed among the breast imaging centres. All mammograms were read by a single radiologist. Each woman effectively had a double reading – one with DM and one with FM. Single reading is the norm in the United States. The study attempted full verification of all breast cancer cases occurring with 455 days of study commencement. However, without data linkage with a breast cancer registry or 100 per cent re-

attendance for mammography screening within the 1-year screening interval, the ascertainment of cancer status could not be assured. Attempts at verification were comprehensive, but to correct for any potential systematic under-ascertainment, the authors conducted statistical analyses to correct for verification bias and to test the robustness of the original results. According to the authors, results with and without correction for verification bias were qualitatively similar.

Diagnostic accuracy

The diagnostic accuracy results presented in the Colorado screening study are likely to have been affected by partial verification bias, as only 59 per cent of women in Colorado I had their cancer status verified, and most of those verified were likely to have been women with originally abnormal mammograms (Lewin et al 2002; Lewin et al 2001). Nevertheless, the results suggest (Table 15) that the PPV of DM was slightly higher than that of FM, although not above chance (χ^2 test, $P > 0.3$). Test sensitivity was similar between the two methods (χ^2 test, $P > 0.5$) (Lewin et al 2001), as was the AUC (χ^2 test, $P = 0.18$) (Lewin et al 2002).

The Oslo I follow-up study (Skaane et al 2005; Skaane et al 2003) was a high-quality level III-2 diagnostic study that demonstrated that DM had lower diagnostic accuracy than FM when it was calculated on the basis of the initial double-reading result (as opposed to the further consensus recall meeting) and included data on all interval cancers. On measures of sensitivity, specificity, false positive, false negative, PPV and NPV, FM out-performed DM, although in some cases the differences were very slight. This was reflected in an AUC of 0.66 in DM [95% CI 0.59–0.72] and 0.72 in FM [95% CI 0.65–0.78], with overlapping confidence intervals (Table 15). The observed difference in diagnostic accuracy was therefore not statistically significant, although the clinical importance of the differences is unclear. Positive and negative predictive values (PPVs, NPVs) were very similar for both FM and DM.

Only PPVs and NPVs could be calculated from the per-protocol data presented in the Oslo II study, as the study design was a randomised controlled screening trial rather than a cross-classification of women receiving both methods (Skaane & Skjennald 2004). The PPV and NPV of DM in this poor-quality study were very similar to those of FM. There were no differences in PPV between the mammography trial arms, for each age group assessed, although within the trial arms the PPV was considerably lower in women aged 45–49 as a consequence of the low prevalence of cancer in this age group (Table 15).

The high-quality and largest multicentre diagnostic accuracy study (the DMIST study) determined that DM had slightly better diagnostic accuracy versus the reference standard than FM, although the difference was not statistically significant (AUC difference 0.03 [95% CI –0.02, 0.08], $P = 0.18$) (Pisano et al 2005b). The study had enough statistical power that this result appears valid. The study also found DM to be more accurate in specific subgroups of women who were identified *a priori*. These differences were statistically significant for subgroups of women aged under 50 years, premenopausal and perimenopausal women, and women with dense breasts (Table 15). Calculations of sensitivity, specificity and PPV were done using two different scales – a 7-point malignancy scale and the BIRADS scale, both of which were dichotomised into a positive or negative mammographic finding. The former was presented in the paper with longer follow-up (455 days) than the latter (365 days). The 455-day follow-up results are presented in Table 15. Mean sensitivity of DM relative to FM was very similar overall, although slightly higher in the pre-specified subgroups. Average specificity and PPV, however, did not differ between the two methods, either overall or in the subgroup analysis.

Overall, PPVs were higher in the two Oslo studies than in the other studies as a likely consequence of the method of recall. The Oslo studies generally invited all recalls for further work-up on the basis of consensus or arbitration from either two or four radiologists, unlike the American studies that either invited women for recall on the basis of a single radiologist's reading or a consensus reading from two radiologists.

Cancer detection rate

Immediate detection

The poor-quality Oslo II randomised controlled screening trial determined that the cancer detection rate – when analysis was done as per protocol, ie, when crossovers between screening arms were excluded – was higher with DM (0.59%) than with FM (0.41%), although the difference was not statistically significant (χ^2 test, $P = 0.06$) (Skaane & Skjennald 2004). An intention-to-screen analysis increased the cancer detection rate for FM (0.43%) and slightly decreased the rate for DM (0.58%) (Table 16). The difference in per-protocol cancer detection rates approached statistical significance (χ^2 test, $P = 0.053$) for the subgroup of women aged 50–69. No statistical differences were detected for the subgroup of women aged 45–49 years as a consequence of the small number of cancers in this group.

The medium-quality Colorado screening study (Lewin et al 2002; Lewin et al 2001) observed lower cancer detection rates with DM than with FM, although the difference was not statistically significant (χ^2 test, $P > 0.1$), perhaps owing to a lack of statistical power. These results are likely affected by only partial verification of cancer status during the study period.

Interval detection

Although the four diagnostic accuracy studies were able to report on the surrogate screening effectiveness outcome measure of immediate cancer detection rate, they were not the best design to assess differences in interval cancer rates. Interval cancers that arise after screening would be those missed by both tests in diagnostic accuracy studies (Irwig et al 2004), rather than from one or the other test, as recall for diagnostic work-up in these studies was done on the basis of findings from both DM and FM. Unfortunately, at this time the single available screening effectiveness study (Skaane & Skjennald 2004) has reported only immediate cancer detection rates and has yet to follow up with a report on interval cancer rates.

One diagnostic accuracy follow-up study (Skaane et al 2005) retrospectively assessed the difference in interval cancer rates by ignoring the recall decision and instead associating the immediate and interval cancer detection with the original DM or FM classification. Although this is useful information, its validity is hindered by the retrospective and hypothetical nature of the evaluation. The method of recall in the Oslo I study decreased rates of cancer detection by both methods when the recall data alone were considered. Thirty-one cancers were detected immediately following the initial screen as a result of the consensus recall meeting, and a further 26 'initially prevalent' cancers were detected during the screening interval and at the subsequent screening round 2 years later: that is, the cancers identified were seen on the original mammogram in retrospect. Rates of cancer detection by both methods would have been higher if recall had been done on the basis of at least one of the pair of double readers indicating a positive result (Table 16). The hypothetical cancer detection rate when the full data set (immediate and interval cancers) is used would have been higher with FM (0.84%) than with DM (0.73%), although the difference was not statistically significant (McNemar's test, $P = 0.48$). It is

unclear whether this was a consequence of a lack of statistical power to detect an effect. This is a difference of approximately 1 in 1000 cancers being missed.

The DMIST study compared cancer detection rates between DM and FM (Pisano et al 2005b). Although all efforts were made to ensure full verification of cancer status, the study did not link women who were screened to a national cancer registry, so full ascertainment of cancer status could not be assured. Statistical correction for partial verification bias found that the corrected results did not qualitatively differ from the 'raw' results (ie, where there was no correction for partial verification bias). The study followed up the cohort of women for 455 days (approximately 11–15 months from study entry) to allow the inclusion of interval cancers. The overall cancer detection rate was very similar between DM and FM, with a slightly higher rate given by DM in the pre-specified subgroups of women aged under 50 years, or who were pre- or perimenopausal, or with heterogeneously dense or extremely dense breasts (Table 16).

Recall rate

In the Oslo I study, the recall rate with DM was higher than with FM (4.56% and 3.48%, respectively), although the cancer detection rate was the reverse (0.54% and 0.71%, respectively) (Skaane et al 2003). The higher cancer detection rate with DM in the Oslo II study (Skaane & Skjennald 2004), however, appeared to be linked to the higher recall rate with DM (3.8%) than with FM (2.7%) (Table 16).

In contrast to both of the Oslo studies, the Colorado screening study gave significantly lower recall rates (Lewin et al 2001; Lewin et al 2002) with DM than with FM (χ^2 test, $P < 0.001$). This result may have been affected by the greater proportion of dismissals of positive findings for DM, at the recall meeting for discordant results, than from FM (Lewin et al 2001).

Biopsy rate

The Colorado screening study (Lewin et al 2002; Lewin et al 2001) reported that the biopsy rate associated with DM was lower than that associated with FM – 4.6 per cent (94/2031 positive findings) versus 7.0 per cent (143/2031), respectively. The difference was significantly different (χ^2 test, $P < 0.001$).

Biopsy rates reported in the DMIST study (Pisano et al 2005b) were essentially the same (3.2%) for both methods, as would be expected given that the work-up (and analysis) was based on the patient, rather than on the test findings.

Tumour size

In the Oslo II randomised controlled screening trial, tumour size was assessed in those women who were diagnosed with cancer. The two methods gave no clinically important differences in median tumour size (Table 16) (Skaane & Skjennald 2004).

Interpretation time

In the Oslo I study, the mean interpretation time for a normal digital mammogram with soft-copy reading (after exclusion of outliers) ranged from 31 to 63 s (mean = 45 s). The data were unreliable, however, as the wide variation in reading times may indicate interruptions during the

interpretation process (Skaane et al 2005). Data on film-screen mammogram interpretation time were not collected.

Table 15 Diagnostic accuracy results: screening in asymptomatic women

Study location	Study design quality	Study population	Diagnostic tests	Results						
				AUC ± SE [95% CI]	Sensitivity % [95% CI]	Specificity % [95% CI]	False negative rate %	False positive rate %	PPV % [95% CI]	NPV % [95% CI]
(Pisano et al 2005a; Pisano et al 2005b) Multicentre (33 sites) United States and Canada American College of Radiology Imaging Network DMIST	Level III-2 diagnostic evidence CX, P1 Q1 (QUADAS = 12/14)	*N* = 42 760 (PP) Age at enrolment (*N* = 49 333) = 54.6 [IQR 47–61]	**DM – FM difference** [a] Women <50 Pre- or perimenopausal Heterogeneously dense or extremely dense breasts	0.03 [–0.02, 0.08] *P* = 0.18 0.15 [0.05, 0.25] *P* = 0.0002 0.15 [0.05, 0.24] *P* = 0.002 0.11 [0.04, 0.18] *P* = 0.003						
					Mean ± SE	Mean ± SE			Mean ± SE	
			DM v. histopathology [a] Women <50 Pre- or perimenopausal Heterogeneously dense or extremely dense breasts	0.78 ± 0.02 0.84 ± 0.03 0.82 ± 0.03 0.78 ± 0.03	41 ± 3 49 ± 6 47 ± 5 38 ± 4	98 ± 0.1 97 ± 0.1 97 ± 0.1 97 ± 0.1			12 ± 1 8 ± 1 10 ± 1 10 ± 1	
					Mean ± SE	Mean ± SE			Mean ± SE	
			FM v. histopathology [a] Women <50 Pre- or perimenopausal Heterogeneously dense or extremely dense breasts	0.74 ± 0.02 0.69 ± 0.05 0.67 ± 0.05 0.68 ± 0.03	41 ± 3 35 ± 6 38 ± 5 36 ± 4	98 ± 0.1 98 ± 0.1 98 ± 0.1 97 ± 0.1			13 ± 1 7 ± 1 9 ± 1 10 ± 1	
(Skaane et al 2005; Skaane et al 2003) Oslo, Norway Oslo I Study	Level III-2 diagnostic evidence CX, P1 Q1 (QUADAS = 12/14)	*N* = 3683 Age = 58.2 y [50–69]	**DM v. histopathology** [b] *Hypothetical based on initial positive result* [c]						*3.76*	
			Based on recalls (consensus method used)						11.90	
			Hypothetical based on initial positive result and including interval cancers [c]	*0.66 [0.59, 0.72]*	*47.37 [34, 61]*	*83.87 [83, 85]*	*52.63 (30/57)*	*16.13 (585/3626)*	*4.41 [3, 6]*	*99.02 [98.6, 99.3]*
			FM v. histopathology [b] *Hypothetical based on initial positive result* [c]						*6.33*	

Study location	Study design quality	Study population	Diagnostic tests	Results						
				AUC ± SE [95% CI]	Sensitivity % [95% CI]	Specificity % [95% CI]	False negative rate %	False positive rate %	PPV % [95% CI]	NPV % [95% CI]
			Based on recalls (consensus method used)						20.31	
			Hypothetical based on initial positive result and including interval cancers [c]	*0.72 [0.65, 0.78]*	*54.39 [41, 68]*	*88.67 [88, 90]*	*45.61 (26/57)*	*11.33 (411/3626)*	*7.01 [5, 10]*	*99.20 [98.8, 99.5]*
(Lewin et al 2002; Lewin et al 2001) Denver, Colorado Colorado screening study I (interim) and II (final results)	Level III-2 diagnostic evidence CX, P1 Q2 (QUADAS = 10/14)	*N* = 4489 (1665 women enrolled 2×; 291 women enrolled 3×); 6736 examinations Age = 55.6 y	**DM v. histopathology**	0.74	60		40.5 (17/42)	11.5 (768/6694)	3.4 (27/793)	
			FM v. histopathology	0.80	63		21.4 (9/42)	14.5 (968/6694)	3.3 (33/1001)	
(Skaane & Skjennald 2004) Oslo, Norway Oslo II Study	Level II screening evidence C1, P1 Q3 (NHMRC = not double-blinded, no allocation concealment, no ITS analysis)	*N* = 25 263 DM = 7209 (ITT) DM = 6998 (PP) 50–69 grp = 59.1 mean years 45–49 grp = 47.4 mean years	**DM v. histopathology** 50–69 age group (PP)						21.6 (33/153)	78.4 (120/153)
			45–49 age group (PP)						7.1 (8/112)	92.9 (104/112)
			Overall (PP)						15.5 (41/265)	84.5 (224/265)
		FM = 18 054 (ITT) FM = 17 913 (PP) 50–69 grp = 59.2 mean years 45–49 grp = 47.4 mean years	**FM v. histopathology** 50–69 age group (PP)						22.1 (56/253)	77.9 (197/253)
			45–49 age group (PP)						7.4 (17/231)	92.6 (214/231)
			Overall (PP)						15.1 (73/484)	84.9 (411/484)

AUC = area under the curve in ROC analysis; SE = standard error; PPV = positive predictive value; NPV = negative predictive value; DM = digital mammography; FM = film-screen mammography. a: Calculated using 7-point malignancy scale at 455 days' follow-up, dichotomising a positive result from scores of 4, 5 ,6 and 7 and a negative result as 1, 2 or 3. Analysis using the BIRADS scale at 365 days' follow-up is also presented in the study with results in a similar direction. b: DM v. FM could be calculated but was considered unnecessary and likely misleading given data v. reference standard was available. c: Calculated in the evaluation; ITS = intention-to-screen; ITT = intention-to-treat; PP = per-protocol analysis; IQR = interquartile range.

Note: Confidence intervals were only calculated where possible.

Table 16 Screening effectiveness results: asymptomatic women

Study location	Screening program characteristics	Screening tests	Results		
			Tumour size mm (median)	Recall rate (%)	Cancer detection rate (%)
(Pisano et al 2005a; Pisano et al 2005b) Multicentre (33 sites) United States and Canada American College of Radiology Imaging Network DMIST	*Screening interval* – Annual. 455 days' follow-up *Views* – 2 (CC, MLO) of each breast with each method *Reading* – Single reading of DM or FM for each woman by independent radiologists. Therefore double reading per woman. Prior films or soft copy available • 4 scales: probability of malignancy (0%–100%); BIRADS (0–5), ordinal malignancy scale (1–7); call-back scale (1–6) • Work-up on basis of either DM or FM radiologist indicating abnormal mammogram <u>*Cancer status verification*</u> Histopathology + ~11–15 months of follow-up of interval cancers of those indicated at work-up for short-term follow-up + at next annual screen and other follow-up where possible (no cancer registry or linked data available). Sensitivity analysis undertaken with *post hoc* statistical correction for partial verification bias	**DM** (hard or soft copy, 5 types of DM) Overall			PP 0.43 (185/42 760)
		Women <50			0.33 (48/14 335)
		Pre- or perimenopausal			0.41 (65/15 803)
		Heterogeneously dense or extremely dense breasts			0.47 (94/19 897)
		FM (hard copy) Overall			0.41 (174/42 760)
		Women <50			0.22 (32/14 335)
		Pre- or perimenopausal			0.27 (43/15 803)
		Heterogeneously dense or extremely dense breasts			0.37 (73/19 897)
(Skaane et al 2005; Skaane et al 2003) Oslo, Norway Oslo I Study	*Screening interval* – 2 years *Views* – 2 (CC, MLO) of each breast with each method *Reading* – Batch double reading each of DM and FM for each woman by independent radiologists. Therefore 4 readings per woman • ≥1 positive [a] reading in each double reading = positive or assess for recall • Recall for DM and FM assessed independently. Recall on basis of ≥3 score by ≥1 reader; or consensus of 4 radiologists assessing DM or FM with 2 score. Previous films available. <u>*Cancer status verification*</u> Histopathology + 2-year follow-up of interval cancers through linked registry data	**DM** (soft copy, FFDM) *Hypothetical based on initial positive result* [b]		ITS	ITS *0.62 (23/3683)*
		Based on actual recalls		4.56 (168/3683)	0.54 (20/3683)
		Hypothetical based on initial positive result and including interval cancers [b]			*0.73 (27/3683)*
		FM (hard copy) *Hypothetical based on initial positive result* [b]		ITS	ITS *0.76 (28/3683)*
		Based on actual recalls		3.48 (128/3683)	0.71 (26/3683)

Study location	Screening program characteristics	Screening tests	Results		
			Tumour size mm (median)	Recall rate (%)	Cancer detection rate (%)
		Hypothetical based on initial positive result and including interval cancers [b]			*0.84 (31/3683)*
(Lewin et al 2002; Lewin et al 2001) Denver, Colorado Colorado screening study I (interim) and II (final results)	*Screening interval* – Annual. Follow-up of cohort was maintained to 12 months after study entry *Views* – 2 (CC, MLO) of each breast with each method Single reading of DM or FM for each woman by independent radiologists. Therefore double reading per woman. Prior films available • 2 scales: probability of malignancy (0%–100%); BIRADS (0, 2, 3, 4, 5) • Work-up on basis of both DM and FM radiologist indicating abnormal mammogram. Discordant findings were assessed together at a recall meeting <u>*Cancer status verification*</u> Histopathology + 12 months' follow-up via next screen. Colorado I also included telephone interview re interval cancers – 59% of cohort had verification. Colorado II does not mention verification status	**DM** (soft copy, prototype FFDM) Based on positive findings		21.8 (979/4489)	n/a
		Based on positive mammographic examinations		11.8 (793/6736)	0.37 (25/6736 exams)
		FM (hard copy) Based on positive findings		30.0 (1345/4489)	n/a
		Based on positive mammographic examinations		14.9 (1001/6736)	0.49 (33/6736 exams)
(Skaane & Skjennald 2004) Oslo, Norway Oslo II Study	*Screening interval* – 2 years for women outside of Oslo area; all women aged 45–49 (screened only in Oslo) had 1-year screening interval; women aged 50–69 living in Oslo had 1-year screening interval *Views* – 2 views (CC, MLO) of each breast *Reading* – Batch double reading each of DM or FM by independent radiologists. In first 4 months, 4 readings of DM undertaken to ensure cancer detection rate no lower than FM (stopping rule). Only first 2 readings for each DM case used in analysis during this period • ≥1 positive [a] reading in each double reading = positive or assess for recall • Recall on basis of ≥3 score by ≥1 reader; or consensus of ~2 radiologists assessing DM or FM with 2 score. Previous films available	**DM**			
		50–69 age group (PP)	15	3.8 (153/3985)	0.83 (33/3985)
		45–49 age group (PP)	10	3.7 (112/3012)	0.27 (8/3012)
		Overall (PP)		3.8 (265/6997)	0.59 (41/6997)
		Overall (ITS)			0.58 (42/7209)
		FM			
		50–69 age group (PP)	13	2.5 (253/10 304)	0.54 (56/10 304)
		45–49 age group (PP)	11	3.0 (231/7607)	0.22 (17/7607)
		Overall (PP)		2.7 (484/17 911)	0.41 (73/17 911)
		Overall (ITS))			0.43 (78/18 054)

Study location	Screening program characteristics	Screening tests	Results		
			Tumour size mm (median)	Recall rate (%)	Cancer detection rate (%)
	Cancer status verification Histopathology + assessment of cancer status during study period (13 months) through linked registry data				

DM = digital mammography; FM = film-screen mammography; BIRADS = Breast Imaging Reporting and Data System; n/a = not available; FFDM = full-field digital mammography; CC = craniocaudal; MLO = mediolateral oblique; PP = per-protocol analysis; ITS = intention-to-screen. a: Positive = score of 2–5 on 5-point scale, 1 = normal or definitely benign, 2 = probably benign, 3 = indeterminate finding, 4 = probably malignant, 5 = malignant. b: Calculated in the evaluation.

Does digital mammography change patient management in a screening population?

If DM is more sensitive than FM, a greater proportion of women will have breast cancer detected, and could consequently receive treatment at an earlier stage than those who are missed and either are picked up through subsequent screening or report at a later stage with symptoms (interval cancers). Similarly, if DM is more specific than FM, a greater proportion of women will not need to undergo unnecessary diagnostic follow-up, eg, biopsy procedures. The evidence base was assessed to see whether the use of DM in preference to FM resulted in any changes in clinical decision-making or patient management. Studies were assessed against the criteria outlined in Box 5. No studies were identified that met the inclusion criteria.

Box 5 Inclusion criteria for identification of studies relevant for an assessment of change in management as a consequence of DM for screening

Characteristic	Criteria
Publication type	Randomised or non-randomised controlled trials (including before-and-after studies) or cohort studies or systematic reviews of these study designs. Uncontrolled pre-test/post-test case series. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded
Population	Asymptomatic women aged 40 years and over attending a breast screening program Subgroups: 40–49 y, 50–69 y, 70+ y
Intervention/test	Screening DM, consisting of 2 full-field x-ray views of each breast
Comparator	Screening FM, consisting of 2 x-ray views of each breast
Outcome	Rates of diagnostic tests, time to diagnosis, rate of referral to specialist, time to treatment, treatment rates, method of treatment
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

Is digital mammography effective for surveillance of women at potentially high risk of breast cancer and for diagnosis of women with symptoms?

Summary – Is DM effective for surveillance of women at potentially high risk and for diagnosis of symptomatic women?

The impact of DM on the outcomes of women at potentially high risk of breast cancer and symptomatic women was assessed using a linked evidence approach.

Four studies reported on the diagnostic accuracy of DM in women at potentially high risk and symptomatic women. Results provide limited evidence as to the effectiveness of DM in these groups of women.

Seo et al (2006) reported PPVs of 0.48 for FM and 0.47 for DM. The sensitivity of DM in detecting breast cancer ranged between 66 and 72 per cent across the three studies that reported on the outcome. The sensitivity of FM was slightly higher across the same studies, ranging between 70 and 89 per cent. In the two studies that attempted to quantify the outcome, the specificity of mammography ranged between 55 and 67 per cent for DM and 53 and 60 per cent for FM. These results suggest that DM has similar diagnostic properties to FM for assessing women at potentially high risk for breast cancer and symptomatic women.

Is digital mammography accurate in these populations?

Studies of the diagnostic accuracy of DM for the surveillance of women at potentially high risk of breast cancer and the diagnosis of symptomatic women were included for assessment if they met the criteria outlined in Box 6. Five studies met the inclusion criteria. Among these, a health technology assessment (Blue Cross Blue Shield Association 2006) was subsequently excluded from further analysis to avoid the duplication of results. The results of the remaining four studies are presented in Table 17.

Only one study (Venta et al 2001) mentioned the use of ultrasound in conjunction with the comparator, FM. It is possible that ultrasound or MRI was used in conjunction with mammography in the other three studies, but this was not explicitly stated. These complementary diagnostic tools are used in the surveillance of women at potentially high risk of breast cancer and in the diagnosis of symptomatic women.

Box 6 Inclusion criteria for identification of studies relevant to an assessment of diagnostic accuracy of DM for surveillance and diagnosis

Characteristic	Criteria
Publication type	Cross-sectional studies where patients were cross-classified on the test and comparator(s) and/or reference standard. Case-control diagnostic studies were acceptable only if cross-sectional studies were not available. Systematic reviews of cross-sectional studies. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded
Population	Women with symptoms or a clinical abnormality of the breast suggesting malignancy; OR women at potentially high risk of breast cancer due to: 1. Potentially high risk of ovarian cancer 2. two 1° or 2° relatives on one side of the family diagnosed with breast or ovarian cancer plus one or more of the following features on the same side of the family: – additional relative(s) with breast or ovarian cancer – breast cancer diagnosed before the age of 40 – bilateral breast cancer – breast and ovarian cancer in the same woman – Ashkenazi Jewish ancestry – breast cancer in a male relative 3. one 1° or 2° relative diagnosed with breast cancer at age 45 or younger plus another 1° or 2° relative on the same side of the family with sarcoma (bone or soft tissue) at age 45 or younger 4. being a member of a family in which the presence of a high-risk breast cancer gene mutation has been established, eg, *BRCA1*, *BRCA2*, *Tp53* 5. personal history of breast cancer 6. premalignant conditions, ie, lobal carcinoma *in situ* and atypical ductal hyperplasia Subgroups: (1) <50 y v. 50+ y [a] (2) very to extremely dense breast tissue v. less dense tissue (3) pre- to perimenopausal v. postmenopausal [a]
Intervention/test	Diagnostic DM, consisting of additional full field x-ray views of both breasts ± breast ultrasound scan, ± breast MRI
Comparator	Diagnostic FM, consisting of additional x-ray views of both breasts ± breast ultrasound scan, ± breast MRI
Reference standard	Histopathology
Outcome	Sensitivity and specificity (and therefore rates of false positives and negatives), positive and negative likelihood ratios, PPVs and NPVs (and therefore false alarm and reassurance rates), diagnostic odds ratios, ROC curves, AUC, accuracy
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

MRI = magnetic resonance imaging. a: These subgroups are surrogate measures for breast tissue density, as density can be measured only via a mammogram.

All four studies included in the assessment were of level III-2 diagnostic evidence. With the exception of one poor-quality study (Hendrick et al 2001), the other three studies were of high quality. Each study contained a mixture of women at potentially high risk, symptomatic women and women referred for diagnostic mammography because of an abnormal screening result. Three of the four studies used histopathology as the reference standard for diagnosing malignancy. The remaining study (Hendrick et al 2001) failed to report the reference standard used. Mammograms in all studies were assessed by a single radiologist and were rated according to BIRADS (American College of Radiology 2003); see Box 7. In three of the four studies, BIRADS scores of 1 or 2 were classified as negative for malignancy, and scores of 3, 4 or 5 were classified as positive. In the remaining study (Venta et al 2001) BIRADS scores of 1, 2 or 3 defined a negative result for malignancy, and scores of 4 or 5 defined a positive result.

Box 7 BIRADS categorisation of mammograms (American College of Radiology 2003)

BIRADS category	Assessment
0	Assessment incomplete
1	Negative finding
2	Benign finding
3	Probably benign finding – initial short-interval follow-up recommended
4	Suspicious abnormality – biopsy should be considered
5	Highly suggestive of malignancy – biopsy and appropriate treatment recommended
6	Known biopsy – proven malignancy

Seo et al (2006) compared the PPV of FM to that of DM in a group of 11 621 women undergoing diagnostic mammography. Depending on the availability of units at the time of presentation, each woman underwent either FM or DM. The true disease status was confirmed by histopathology in 1121 women (1400 lesions in total). In lesions that were classified as positive by mammography, the PPVs of FM and DM were 48 per cent [95% CI 44, 52] and 47 per cent [95% CI 41, 53] respectively ($P = 0.86$). No statistically significant differences between the PPVs of the two methods were reported when analyses were also stratified according to breast density.

Two of the four studies presented the test sensitivity of FM and DM in detecting biopsy-proven malignancies. In a study of 692 women who underwent both types of mammography, Venta et al (2001) reported a test sensitivity of 89 per cent for FM [95% CI 65, 99] and 72 per cent for DM [95% CI 49, 90]. With only 50 women undergoing biopsy, however, the study was unable to demonstrate a statistically significant difference in test sensitivity between the two methods ($P = 0.20$). In a study involving 242 women who underwent diagnostic mammography, Cole et al (2004) reported a test sensitivity of 74 per cent for FM and 66 per cent for DM. In addition to biopsy results, the authors defined actual breast cancer status according to the results of 1-year clinical follow-up using FM. Patients who did not undergo biopsy and had normal results on FM at follow-up were assumed to be free of cancer. Using this approach, the study reported a test specificity of 60 per cent for FM and 67 per cent for DM. The AUC was calculated to be 0.77 for FM and 0.72 for DM [difference in AUC = –0.05, 95% CI –0.10, 0.002].

In the poor-quality study by Hendrick et al (2001), 625 women underwent diagnostic mammography with both FM and DM. Although the authors provided enough information to calculate a variety of diagnostic accuracy measures, it is unclear how the cancer status of the women was ascertained. From the reported cancer status, however,

the test sensitivity of FM and DM in detecting cancer cases was calculated to be 70 per cent [95% CI 0.63, 0.76] and 68 per cent [95% CI 0.62, 0.74], respectively. Likewise, the test specificity of FM and DM for correctly classifying negative cancer cases was calculated to be 53 per cent [95% CI 0.51, 0.55] and 55 per cent [95% CI 0.53, 0.56], respectively. Overlapping confidence intervals suggest a lack of statistically significant differences in diagnostic accuracy between the two methods in this population.

Table 17 Diagnostic accuracy results: women at potentially high risk and symptomatic women

Study location	Study design quality	Study population	Diagnostic tests	Results				
				AUC	Sensitivity % [95% CI]	Specificity % [95% CI]	PPV % [95% CI]	NPV % [95% CI]
(Seo et al 2006) North Carolina, USA	Level III-2 diagnostic evidence CX, P1 Q1 (QUADAS = 12/14)	*N* = 11 621 DM, not stated FM, not stated Ref. standard = 1121 (1400 lesions) Symptomatic women, women at potentially high risk and women with abnormal screening mammograms Age: mean 56.5 y (range 28–85)	**DM v. partial histo** **FM v. partial histo** **High breast density:** DM v. partial histo FM v. partial histo **Low breast density:** DM v. partial histo FM v. partial histo				47.0 [41, 53] 48.0 [44, 52] 51.0 [43, 59] 45.0 [39, 51] 42.0 [32, 52] 51.0 [45, 57]	
(Venta et al 2001) Chicago, IL, USA	Level III-2 diagnostic evidence CX, P1 Q1 (QUADAS = 12/14)	*N* = 692 DM = 692 FM = 692 Ref. standard = 50 Symptomatic women, women at potentially high risk, women with abnormal screening mammograms, women with breast augmentation and women referred for diagnostic mammography due to anxiety or proximity to clinic Age: mean 53.0 y (range 40–88)	**DM v. FM** **DM v. partial histo** **FM v. partial histo**		47.0 [32, 62] 72.0 [49, 90] 89.0 [65, 99]	98.0 [97, 99]	47.0 [32, 62]	98.0 [97, 99]
(Cole et al 2004) Multicentre study, North America	Level III-2 diagnostic evidence CX, P1 Q1 (QUADAS = 12/14)	*N* = 676 DM = 247 FM = 247 Ref. standard not stated Symptomatic women, women with abnormal screening mammograms and women recommended or scheduled to undergo histopathology Age: not stated	**DM v. partial histo** **FM v. partial histo**	0.72 0.77	66.0 74.0	67.0 60.0		

Study location	Study design quality	Study population	Diagnostic tests	Results				
				AUC	Sensitivity % [95% CI]	Specificity % [95% CI]	PPV % [95% CI]	NPV % [95% CI]
(Hendrick et al 2001) Multicentre study, USA	Level III-2 diagnostic evidence CX, P1 Q3 (QUADAS = 7/14)	*N* = 625 DM = 625 FM = 625 Ref. standard not stated Symptomatic women, women with abnormal screening mammograms and 20 consecutive cancer cases generated from a screening population Age: median 55.0 y (range 40–86 y)	**DM v. FM**		69.0 [67, 71]	74.0 [72, 76]	72.0 [69, 74]	72.0 [69, 74]
			DM v. cancer status	0.72	68.0 [62, 74]	55.0 [53, 56]	10.0 [9, 12]	96.0 [95, 97]
			FM v. cancer status	0.72	70.0 [63, 76]	53.0 [51, 55]	10.0 [9, 12]	96.0 [95, 97]

PPV = positive predictive value; NPV = negative predictive value; DM = digital mammography; FM = film-screen mammography; partial histo = partial verification with histopathology (ie, on 'positives' or abnormal results identified by index test); AUC = area under the curve.

Does digital mammography change patient management for women at potentially high risk or for symptomatic women?

No studies met the inclusion criteria identified *a priori* (Box 8) for determining the change in management consequent to diagnostic DM in women at potentially high risk or symptomatic women.

Box 8 Inclusion criteria for identification of studies relevant to an assessment of change in management as a consequence of DM for surveillance and diagnosis

Characteristic	Criteria
Publication type	Randomised or non-randomised controlled trials (including before-and-after studies) or cohort studies or systematic reviews of these study designs. Uncontrolled pre-test/post-test case series. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded
Population	Women with symptoms or a clinical abnormality of the breast suggesting malignancy; OR women at potentially high risk of breast cancer due to: 1. Potentially high risk of ovarian cancer 2. two 1° or 2° relatives on one side of the family diagnosed with breast or ovarian cancer plus one or more of the following features on the same side of the family: – additional relative(s) with breast or ovarian cancer – breast cancer diagnosed before the age of 40 – bilateral breast cancer – breast and ovarian cancer in the same woman – Ashkenazi Jewish ancestry – breast cancer in a male relative 3. one 1° or 2° relative diagnosed with breast cancer at age 45 or younger plus another 1° or 2° relative on the same side of the family with sarcoma (bone or soft tissue) at age 45 or younger 4. being a member of a family in which the presence of a high-risk breast cancer gene mutation has been established, eg, *BRCA1*, *BRCA2*, *Tp53* 5. personal history of breast cancer 6. premalignant conditions, ie, lobal carcinoma *in situ* and atypical ductal hyperplasia Subgroups: (1) <50 y v. 50+ y [a] (2) very to extremely dense breast tissue v. less dense tissue (3) pre- to perimenopausal v. postmenopausal [a]
Intervention/test	Diagnostic DM, consisting of additional full-field x-ray views of both breasts ± breast ultrasound scan, ± breast MRI
Comparator	Diagnostic FM, consisting of additional x-ray views of both breasts ± breast ultrasound scan ± breast MRI
Outcome	Rates of diagnostic tests, time to diagnosis, rate of referral to specialist, time to treatment, treatment rates, method of treatment
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

MRI = magnetic resonance imaging. a: These subgroups are surrogate measures for breast tissue density, as density can be measured only via a mammogram.

Does treatment change health outcomes in women identified with breast abnormalities?

For any assessment of a diagnostic or screening test there needs to be some assurance that the test in some way will benefit the health of the patient. As mentioned in the 'Approach to assessment' section, it is unnecessary for DM to demonstrate direct health benefits to women (ie, breast cancer mortality) when compared with FM, as cancer detection rate is an appropriate surrogate outcome measure (Irwig et al 2004). As the cancer detection rates are largely the same between DM and FM, it is unlikely that any differential health benefits would be measurable, except perhaps for women aged under 50 years, women who are pre- or perimenopausal, or women who have dense breasts. In these subgroups, DM had better overall diagnostic performance, and cancer detection was slightly higher. However, to detect actual differential health benefits in these subgroups as a consequence of the use of DM compared with FM would require a study of enormous size and resources. The assumption is that early cancer detection in these subgroups will manifest as health benefits.

Below is a summary of high-level synthesised data indicating the benefits of early detection and treatment of breast cancer in a mammography (FM only) screening population. There were no studies that assessed health outcomes from mammography specifically in a high-risk population. However, the outcomes provided below should be transferable to DM for population-based screening (asymptomatic women) and surveillance (women at potentially high risk) purposes, with a caveat that in some subgroups there are likely to be additional health benefits.

For women who present symptomatically, it is unlikely that there would be any differences in treatment as a consequence of DM, given the similar diagnostic accuracy to that of FM, even if subgroups effects are eventually identified. Treatment would occur no earlier than currently.

Summary – Do early identification and treatment change health outcomes in a breast cancer screening population?

Six meta-analyses were identified that combined studies comparing the mortality rates from breast cancer after screening with those in populations who did not undergo screening. A comparison of the included meta-analyses is complicated, as not all included the same trials in their analysis. In addition, not all meta-analyses reported results at a defined time of follow-up (Glasziou et al 1995; Hendrick et al 1997; Smart et al 1995).

Quality of meta-analyses also varied, as not all conducted systematic reviews to identify relevant trials, and not all performed a critical appraisal of the included studies. Even when a critical appraisal was performed, high-quality trials were not always included in the meta-analysis (Humphrey et al 2002).

In this context, all analyses which assessed the effectiveness of mammography in randomised controlled trials reported a relative risk reduction in women aged 50 years or more, with follow-up ranging from 7 to 14 years. The relative reduction varied slightly depending on the duration of follow-up and the studies considered in the analyses.

The relative risk of death from breast cancer is more complicated in women aged less than 50 years. The meta-analyses by Kerlikowske et al (1995), which used early follow-up data, did not report any significant reduction in relative risk in women who were screened relative to those who were not. In the analysis by Smart et al (1995), which included additional follow-up data, a relative risk reduction was reported only when the Canadian study was removed from the analysis. This may be explained by the fact that the Canadian study enrolled volunteers who were expected to be healthier than the general population. Hendrick et al (1997), who used additional follow-up data to supplement the analysis performed by Smart et al (1995), reported a significant relative risk reduction in women aged 40–49 years with the inclusion of the Canadian study (RR 0.82 [95% CI 0.71, 0.95]).

The Cochrane review by Gøtzsche and Nielsen (2007) did not report a relative risk reduction in women aged less than 50 years at 13 years' follow-up when considering only trials that were adequately randomised. However, when including trials which were quasi-randomised, they did report a statistically significant, although marginal, relative risk reduction of death from breast cancer (RR 0.84 [95% CI 0.72, 0.99]). This result is very similar to the one found by Hendrick and colleagues at a mean of 12.7 years' follow-up.

In summary, it would appear that a protective effect from early detection and treatment by screening in women aged 40–49 years is not apparent until 13 years after breast cancer is diagnosed. However, for women aged 50 years or over when diagnosed, the protective effect of early treatment is apparent between 7 and 14 years after detection.

The effectiveness of a diagnostic and screening test depends on whether it improves patient outcomes. To determine whether there are improved patient outcomes for women receiving DM, it is necessary to assess treatment as a consequence of a positive mammogram. To this end, the effectiveness of early versus late treatment was assessed.

In practice, early treatment is given to women diagnosed through breast cancer screening. These women are asymptomatic, and it is not possible for them to be diagnosed earlier without a screening mammogram. In contrast, women who do not participate in breast cancer screening are unlikely to receive early treatment, as they will not be diagnosed until they present with clinical symptoms. Using this reasoning, we assessed treatment effectiveness by comparing the health outcomes of women diagnosed with breast cancer in a screening population and those who were not diagnosed through screening. *A priori* criteria for studies to be included in this assessment are described in Box 9.

Box 9 Inclusion criteria for identification of studies relevant for an assessment of treatment effectiveness (patient health benefit) as a consequence of early detection of breast cancer

Characteristic	Criteria
Publication type	Systematic reviews or meta-analyses of randomised controlled trials; evidence-based clinical practice guidelines. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded
Patient	Women diagnosed with breast cancer who presented asymptomatically, aged 40 years and over, at a breast screening program Subgroups: 40–49 y, 50–69 y, 70+ y
Intervention/test	Early breast cancer treatment including surgery, radiotherapy, drug therapy, hormonal therapy and chemotherapy
Comparator	Late or no breast cancer treatment
Outcome	Primary – *Patient-relevant:* survival rates, reduction in mortality (all-cause and breast cancer), morbidity, quality of life, adverse effects Secondary – tumour size or stage, recurrence rates, disease progression
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

Meta-analyses that met the inclusion criteria determined *a priori* are shown in Table 18. These studies compared the risk of death in patients who underwent mammography screening with that in populations who did not receive screening.

Table 18 Summary of results of meta-analyses included in the assessment of treatment effectiveness

Meta-analyses	Duration of follow-up	Summary relative risk of death from breast cancer in screening v. no screening mammography population [95% CI]		
		<50 years	>50 years	Overall
(Glasziou et al 1995)	Unknown	0.95 [0.77, 1.18]	–	–
(Kerlikowske et al 1995)	7–9 y	1.02 [0.82, 1.27]	0.73 [0.63, 0.84]	0.78 [0.69, 0.89]
	10–12 y	0.83 [0.65, 1.06]	0.76 [0.67, 0.87]	0.77 [0.68, 0.86]
	Overall	0.92 [0.75, 1.13]	0.77 [0.69, 0.87]	0.79 [0.71, 0.87]
(Smart et al 1995)	Range: 7–18 y	0.84 [0.69, 1.02] Without Canadian study: 0.76 [0.62, 0.95]	–	–
(Hendrick et al 1997)	Average = 12.7 y Range: 10.5–18 y	0.82 [0.71, 0.95]	–	–
(Humphrey et al 2002)	14 y	0.85 [0.73, 0.99] Without Canadian study: 0.81 [0.73, 0.89]	0.78 [0.70, 0.87]	0.84 [0.77, 0.91] Without Canadian study: 0.80 [0.67, 0.96]
(Gøtzsche & Nielsen 2006)	13 y	Adequately randomised: 0.91 [0.71, 1.18] Quasi-randomised: 0.80 [0.64, 0.98] Total: 0.84 [0.72, 0.99]	Adequately randomised: 0.77 [0.69, 0.86] Quasi-randomised: 0.70 [0.62, 0.80] Total: 0.77 [0.69, 0.86]	Adequately randomised: 0.93 [0.80, 1.09] Quasi-randomised: 0.75 [0.67, 0.83] Total: 0.80 [0.73, 0.88]

Glasziou, Woodward and Mahon conducted a meta-analysis of eight published randomised controlled trials in 1995 to determine the effectiveness of mammography in screening for breast cancer in women under the age of 50 years. At the time, the most recently published and fully described publications were used, and all trials were critically appraised and given a quality score out of 6. This identified trials conducted in Canada

(Canadian National Breast Screening Study 1) and Sweden (Malmö study) as being of the highest quality. The overall relative risk of death from breast cancer was calculated to be 0.95 [95% CI 0.77–1.18]. In addition, the authors conducted a sensitivity analysis of these results and established that the findings were robust to the method of statistical analysis and were not affected by the quality of study design. However, comparison with other meta-analyses is difficult, as it is not apparent at what time during follow-up the calculated risk applies.

In the same year, Kerlikowske et al (1995) used meta-analysis to determine the effectiveness of mammographic screening in women stratified by age. Their defined inclusion criteria identified eight randomised controlled trials and four case-control studies. The authors did not report any appraisal of the quality of the studies. In the analysis of randomised controlled trials alone, the relative risk of death from breast cancer was determined according to duration of follow-up in each age-specific group of women (Table 19) (Kerlikowske et al 1995).

Table 19 Results of Kerlikowske et al's (1995) meta-analysis

Variable	Relative risk of death from breast cancer in women aged 40–49 years [95% CI]	Relative risk of death from breast cancer in women aged 40–74 years [95% CI]	Relative risk of death from breast cancer in women aged 50–74 years [95% CI]
Randomised controlled trials, overall	0.92 [0.75, 1.13]	0.79 [0.71, 0.87]	0.77 [0.69, 0.87]
7–9 years' follow-up	1.02 [0.82, 1.27]	0.78 [0.69, 0.89]	0.73 [0.63, 0.84]
10–12 years' follow-up	0.83 [0.65, 1.06]	0.77 [0.68, 0.86]	0.76 [0.67, 0.87]

Among women aged 40–49 years, there was a 17 per cent reduction in the relative risk of death from breast cancer at 10–12 years' follow-up, but this was not statistically significant. In addition, there appeared to be a marginal relative increase (2%) in risk of death at 7–9 years' follow-up, but this also was not statistically significant. In contrast, there was a considerable reduction in the relative risk of death from breast cancer for women aged 50–74 years at both 7–9 years' (27%) and 10–12 years' follow-up (24%).

A third meta-analysis conducted in 1995 by Smart et al (1995) based on data available at the time looked at the relative risk of death from breast cancer in women aged between 40 and 49 years in a breast cancer screening program. Inclusion criteria for the meta-analysis were limited to available data from randomised controlled trials of breast cancer screening in women aged 40–49 years (Table 20). The authors did not provide details of the sources used for searches of primary studies, or of how they selected articles for which there were multiple publications relating to the same study. It is also not clear which criteria, if any, they used to assess the quality of the trials. In addition, they did not report the use of tests for heterogeneity in the meta-analysis. With this in mind, it is not possible to determine whether the results of this meta-analysis are valid.

Smart et al (1995) also included eight randomised controlled trials in their meta-analysis. However, inclusion differed from the Kerlikowske et al (1995) study in that the second Canadian study was not included, as it did not assess the effectiveness of screening in women aged 40–49 years. This analysis combined studies which used mammography alone with studies using mammography and CBE for screening. Table 20 summarises the included trials in this study. Invited participants received mammography screening, while the control groups received usual care. The frequency of screening varied between trials, ranging from annual screening to screening every 28 months.

Table 20 Randomised controlled trial results for women aged 40 to 49 years (Smart et al 1995)

Study	Screening method	Frequency of screening	Follow-up (years)	Participants		Relative risk [95% CI][a]
				Invited	Control	
HIP	2 V MM + CBE	Annually, 4 rounds	18	14 432	14 701	0.77 [0.53, 1.11]
Malmö	1 or 2 V MM	18–24 months, 5 rounds	12	3658	3679	0.51 [0.22, 1.17]
Kopparberg	1 V MM	24 months, 4 rounds	12	9582	5031	0.73 [0.37, 1.41]
Östergötland	1 V MM	24 months, 4 rounds	12	10 262	10 573	1.02 [0.52, 1.99]
Edinburgh	1 or 2 V MM	24 months, 4 rounds	11	5913	5810	0.78 [0.46, 1.51]
Stockholm	1 V MM	28 months, 2 rounds	8	14 375	7103	1.04 [0.53, 2.05]
Göteborg	2 V MM	18 months, 4 rounds	7	10 600	12 800	0.73 [0.27, 1.97]
CNBSS-1	2 V MM + CBE	12 months, 5 rounds	7	25 214	25 216	1.36 [0.84, 2.21]

2 V MM = 2 view mammography; CBE = clinical breast examination; 1 V MM = 1 view mammography, CNBSS = Canadian National Breast Screening Study

a: Note that multiple publications of the same study may have reported different risks.

The authors analysed the data in four ways:

1. with the inclusion of all then-current published follow-up data from randomised controlled trials of women aged 40 to 49 years participating in breast screening mammography
2. as in (1) but with the exclusion of the CNBSS-1 study
3. as in (1) but with the inclusion of presented or unpublished data
4. as in (2) but with the inclusion of presented or unpublished data.

The results of these analyses are shown in Table 21.

Table 21 Summary relative risks of death from breast cancer for women aged 40 to 49 years (Smart et al 1995)

Analyses	Relative risk of death by breast cancer [95% CI]
All published data as of 1 June 1994	0.90 [0.73, 1.10]
All published data as of 1 June 1994 – excluding CNBSS-1 study	0.82 [0.66, 1.03]
All published and unpublished data as of 1 June 1994	0.84 [0.69, 1.02]
All published and unpublished data as of 1 June 1994 – excluding CNBSS-1 study	0.76 [0.62, 0.95]

CNBSS = Canadian National Breast Screening Study

These analyses combine all follow-up data and therefore do not indicate the relative risk at a particular time of follow-up. They concluded that there were non-significant relative risk reductions of 16 per cent when all eight trials were combined and of 24 per cent when the Canadian study was excluded.

A subsequent meta-analysis by the same group (Hendrick et al 1997) included data with longer follow-up of at least 10.5 years (average follow-up = 12.7 years) (Table 22).

Table 22 Results of included randomised trials in meta-analysis by Hendrick et al (1997)

Study	Screening method	Frequency of screening	Follow-up (years)	Participants		Relative risk [95% CI][a]
				Invited	Control	
HIP	2 V MM + CBE	Annually, 4 rounds	18	14 432	14 701	0.77 [0.53, 1.11]
Malmö	1 or 2 V MM	18–24 months, 5 rounds	12.7	13 528	12 242	0.64 [0.45, 0.89]
Kopparberg	1 V MM	24 months, 4 rounds	15.2	9650	5009	0.73 [0.37, 1.41]
Östergötland	1 V MM	24 months, 4 rounds	14.2	10 240	10 411	1.02 [0.59, 1.77]
Edinburgh	1 or 2 V MM	24 months, 4 rounds	12.6	11 755	10 641	0.81 [0.54, 1.20]
Stockholm	1 V MM	28 months, 2 rounds	11.4	14 185	7985	1.01 [0.51, 2.02]
Göteborg	2 V MM	18 months, 5 rounds	12	11 724	14 217	0.56 [0.32, 0.98]
CNBSS-1	2 V MM + CBE	12 months, 4–5 rounds	10.5	25 214	25 216	1.14 [0.83, 1.56]

2 V MM = 2-view mammography; CBE = clinical breast examination; 1 V MM = 1-view mammography

CNBSS = Canadian National Breast Screening Study

a: Note that multiple publications of the same study may have reported different risks.

The authors of this updated meta-analysis did not explain some discrepancies in the raw data used; in particular, the small differences in the number of participants in the Kopparberg, Östergötland and Stockholm trials. Also, the Göteborg trial has an unexplained extra round of mammography. This might be explained by the different sources of raw data between the meta-analyses.

The authors also did not indicate their search strategy for identifying relevant trials, but they indicated that in cases of multiple publications, they chose data with the longest follow-up. This was determined by the data with the greatest number of deaths from breast cancer among both groups. They did not indicate whether they conducted a critical appraisal of the included studies. In their analysis they reported summary relative risks for all eight randomised controlled trials, and then again for the five Swedish trials alone, at an average of 12.7 years' follow-up. Only the results of the meta-analysis of all eight randomised controlled trials have been reported here.

Hendrick et al (1997) showed that mammographic screening resulted in a significant reduction in relative risk from death from breast cancer in women aged 40–49 years of 18 per cent [95% CI 0.71, 0.95]. Through sensitivity analysis, they were able to conclude that these findings were robust to the study design and study heterogeneity.

Humphrey and colleagues conducted a meta-analysis of eight randomised controlled trials of screening in 2002 (Humphrey et al 2002). The review's objectives were to critically appraise and synthesise evidence about the effectiveness of breast cancer screening overall and among women younger than 50 years. The authors were interested in randomised controlled trials as the study design and mortality from breast cancer as the outcome. From each of the studies they extracted patient population, design, quality, data analysis and published results at each reported length of follow-up. Additionally, two authors independently assessed the internal validity of the trials against predefined criteria developed by the US Preventive Services Task Force. The eight randomised trials were made up of four Swedish trials (Göteborg, Stockholm, Malmö and Swedish Two-Country study), two Canadian studies (CNBCSS-1 and -2), a US study (HIP trial) and one study conducted in Edinburgh, Scotland. The eight studies varied in recruitment method, population and comparator. All but the two Canadian studies randomly assigned women to either an invitation-to-screen group or a group receiving 'usual care'. The two Canadian studies recruited a sample of volunteers identified through mass media. The intervention group also differed among the studies in that only the four Swedish trials

compared mammography alone with usual care. The other four studies included patients receiving mammography plus CBE. Lastly, differences in the age at enrolment are clearly accounted for in the results. Table 23 outlines the effectiveness of mammography in reducing breast cancer mortality.

Only one of the four Swedish studies, that is, the Swedish Two-County Trial, demonstrated statistically significant reductions in the risk of death from breast cancer.

The meta-analysis conducted by Humphrey et al (2002) of all age groups estimated a relative risk of death from breast cancer of 0.84 [95% CI 0.77, 0.91] at 14 years' follow-up. The meta-analysis excluded the Edinburgh study for its poor internal validity. Further, to determine the effectiveness of an invitation-to-screen program, the two Canadian studies were excluded on account of their recruitment of volunteers. This resulted in a relative risk of 0.81 [95% CI 0.73, 0.89].

The effectiveness of screening in women aged 40 to 49 was determined by calculating the relative risk of death from breast cancer in this age group, which was estimated to be 0.85 [95% CI 0.73, 0.99] at 14 years' follow-up with the exclusion of the Edinburgh trial. When CNBSS-1 was also excluded, the relative risk was 0.80 [95% CI 0.67, 0.96]. The relative risk for women aged 50 or older was 0.78 [95% CI 0.70, 0.87]. Insufficient evidence was available to perform a meta-analysis of the effectiveness of screening older women in reducing breast cancer mortality.

The most recent review was performed by Gøtzsche and Nielsen (2006). This was a Cochrane review and looked at randomised controlled trials assessing the effectiveness of mammography for breast cancer screening in women who had not been previously diagnosed with breast cancer (Gøtzsche & Nielsen 2006).

Unlike a number of other meta-analyses, this review included the Canadian trial in all analyses. In addition, Gøtzsche and Nielsen presented the results of their meta-analysis according to the two adequately randomised trials (Canadian and Malmö), the five quasi-randomised trials (Göteborg 1982, Kopparberg 1977, New York 1963, Stockholm 1981 and Östergötland 1978) and when all trials were combined (see Table 24 for results at 13 years' follow-up). According to the analysis of the adequately randomised trials, there was no reduction in the risk of mortality from breast cancer at 13 years' follow-up in women screened by mammography. However, inclusion of the quasi-randomised trials, resulted in a significant reduction in relative risk of death from breast cancer in women screened with mammography, particularly in women aged 50 years or more.

Table 23 Results of randomised controlled trials of mammography among women 39 to 74 years of age (Humphrey et al 2002)

Study	Age (years)	Median follow-up (years)	Breast cancer deaths / total women		Breast cancer death rate / 1000 women		Relative risk of death from breast cancer [95% CI]	Absolute risk reduction per 1000 women	Number needed to invite to screening to prevent 1 death
			Screened group	Control group	Screened group	Control group			
Mammography alone									
Stockholm	40–64	13.8	82/39 139	20/20 978	2.10	2.38	0.91 [0.65, 1.27]	0.288	3468
Göteborg	39–59	12.8	62/20 724	113/29 200	2.99	3.87	0.76 [0.56, 1.04]	0.878	1139
Malmö	45–70	17.1	161/21 088	198/21 195	7.63	9.35	0.82 [0.67, 1.00]	1.712	584
Swedish two-country trial	40–74	17	319/77 080	333/55 985	4.14	5.95	0.68 [0.59, 0.80]	1.809	553
Mammography plus CBE									
CNBSS-1	40–49	13	105/25 214	108/25 216	4.16	4.28	0.97 [0.74, 1.27]	0.12	–
CNBSS-2	50–59	13	107/19 711	105/19 694	5.43	5.33	1.02 [0.78, 1.33]	–0.097	–
HIP	40–64	16	232/30 239	281/30 256	5.46	6.89	0.79*	1.438	883
Edinburgh	45–64	13	156/22 926	167/21 342	6.80	7.82	0.79 [0.60, 1.02]	1.020	980

*Confidence intervals were not reported. CNBSS = Canadian National Breast Screening Study.

Table 24 Results of meta-analyses conducted by Gøtzsche and Nielsen (2006) at 13 years' follow-up

Analyses	Relative risk of death from breast cancer at 13 years' follow-up [95% CI]	Relative risk of death from breast cancer at 13 years' follow-up for women below 50 years of age [95% CI]	Relative risk of death from breast cancer at 13 years' follow-up for women ≥50 years of age [95% CI]	Overall mortality, 13 years' follow-up
Adequately randomised trials				
Canadian trial 1980a	0.97 [0.74, 1.27]	0.97 [0.74, 1.27]	–	1.00 [0.87, 1.14]
Canadian trial 1980b	1.02 [0.78, 1.33]	–	1.02 [0.78, 1.33]	1.06 [0.96, 1.18]
Malmö 1976	0.81 [0.80, 1.09]	0.52 [0.22, 1.20]	0.86 [0.64, 1.16]	0.98 [0.93, 1.04]
Sub-total	0.93 [0.80, 1.09]	0.91 [0.71, 1.18]	0.94 [0.77, 1.15]	1.00 [0.96, 1.04]
Quasi-randomised trials				
Göteborg 1982	0.75 [0.58, 0.97]	0.70 [0.46, 1.06]	0.83 [0.60, 1.15]	0.89 [0.83, 0.95]
Kopparberg 1977	0.58 [0.45, 0.76]	0.72 [0.38, 1.37]	0.55 [0.42, 0.73]	1.03 [0.99, 1.08]
New York (HIP) 1963	0.83 [0.70, 1.00]	0.78 [0.56, 1.08]	0.78 [0.60, 1.01]	0.99 [0.94, 1.05]
Stockholm 1981	0.73 [0.50, 1.06]	0.96 [0.48, 1.91]	0.64 [0.41, 1.01]	–
Östergötland 1978	0.76 [0.61, 0.95]	1.03 [0.58, 1.84]	0.71 [0.56, 0.91]	1.00 [0.96, 1.04]
Subtotal	0.75 [0.67, 0.83]	0.80 [0.64, 0.98]	0.70 [0.62, 0.80]	0.99 [0.97, 1.01]
Total	**0.80 [0.73, 0.88]**	**0.84 [0.72, 0.99]**	**0.77 [0.69, 0.86]**	**–**

Other relevant considerations

Future trends in mammography

Currently the major radiologic technique in breast imaging is FM. However, FM is being replaced inexorably by DM. The major breast imaging and developing equipment vendors have ceased research into new films, processors and mammography machines. Current equipment is being sold off by the major manufacturers, who are concentrating on digital development. It is anticipated that within a decade there will be no FM machines or processors in major imaging facilities in Australia.

This trend towards DM is the final stage of the digital revolution that began with computed tomography (CT) scanning in the 1980s, followed by ultrasound and MRI. Plain-film radiology is now virtually extinct in major hospitals and private radiology practices. Digital imaging has enabled the incorporation of PACS, which has eliminated the need for film storage and processing and enabled significant improvements in radiography and radiology workflow.

Mammography could be said to be at the high point of film-screen technology. Tissue resolution and film contrast and detail are at their finest. It is extremely unlikely that there would have been any further improvements, considering that recent developments were only minimal modifications to the advancement that came with the release of the Kodak 2000 film series in the late 1990s. DM, however, will allow further improvements, including the development of CAD and 3D imaging technology, notably tomosynthesis.

Hardware

Digital images are preferentially viewed electronically. However, owing to the density of information, specialised monitors are required (James 2004). Initially, high-resolution monochrome cathode ray tube monitors were used, but they consumed a lot of energy and required a large space. The use of flat-panel monitors, which need less space, is becoming more widespread. These monitors are said to have a longer life than the cathode ray tube monitors, although the lights are likely to need replacing within 5 years (Dershaw 2006; Jong & Yaffe 2005). But the technological life of this hardware may be no longer than 5 years in any event owing to technological developments.

In screening, four standard views are provided to the radiologist – the left and right breast CC and MLO images. In incident screening rounds, the radiologist may also assess the four standard views of the previous screen. Thus, eight standard images may be required for viewing or reading for each woman. The order in which the views are assessed is idiosyncratic to the radiologist. In DM, unless the images are laser printed, there will be no hanging of a hard copy, so the image ordering needs to be incorporated into the digital or soft-copy display of images. Technological improvements in this area are occurring apace (eg, Mammoview), and likely uptake of DM will be dependent on the available options for image display (Hemminger 2003).

Workforce issues

One of the major reasons that DM is currently being considered is the shortage of qualified radiologists and radiographers within Australia. This, in conjunction with the increasing demand for mammography, means that there is an imbalance in the supply of and demand for mammography services.

DM has the potential for reducing labour in four areas:

- **Image development.** Sometimes re-intensity windowing, re-processing or re-shooting of films may be required owing to poor image quality or contrast. Digital display incorporates automatic preprocessing algorithms to optimise initial viewing (Hemminger 2003).

- **Image hanging.** Manual labour is required to hang the hard-copy films according to the appropriate protocol for the radiologist. Soft-copy display permits this to be done from a keyboard (Hemminger 2003).

- **Film processing.** With the indirect and direct flat-panel systems (although not the CR systems), there is no need to handle cassettes or to develop films (Comité d'Evaluation et de Diffusion des Innovations 2002). This has occupational health and safety benefits for radiographers, who are no longer exposed to the chemicals associated with the processing of conventional films, and who do not have to handle film cassettes (DR mammography only, not CR mammography).

- **Archive labour.** Film handling to store, archive, retrieve and re-archive films would be unnecessary should soft-copy mammography become the norm. The radiographer or radiologist can retrieve previous digital or digitised images via a workstation. Archived hard-copy films would no longer need to be retrieved from a centralised storage facility.

The use of DM and PACS means that a network of radiologists specialising in DM could be used, thus facilitating workforce flexibility and maximising capacity across both metropolitan and rural localities.

Use of soft-copy display by radiologists has shown both a longer reviewing time (Ciatto et al 2006) and a shorter reviewing time (Pisano et al 2005b) than hard-copy display. This variation may be due to the use, in some instances, of manual image alteration (interactive intensity windowing) as opposed to the use of automatic preset algorithms for enhancing soft-copy display (Hemminger 2003). Both options are generally available. Alternatively, the differences may be due to the experience of the radiologists with soft-copy display.

Training of radiologists in soft-copy reading for DM is likely to be a major concern should it replace or be used in addition to FM. Consistently high image quality using the lowest radiation dose possible is achievable only by clinical and technical professionals with a high level of skill and expertise. Research has been conducted assessing the 'training needs for professionals (radiology technicians, physicists and radiologists) in digital radiology' under the DIMOND III multicentre European project (Peer et al 2005). A survey determined that although a high level of *technical* quality control was available (>70%), organised measures for quality control of staff were limited (14%–30%). Many professionals trained on the job (30%–35%), of whom between 23 and 28 per cent trained with the assistance of the vendor. Published literature was the main resource used

by staff (45%–67%). Thus, it may be reasonable that the use of DM should be considered only within centres of excellence that provide adequate training, supervision and support.

Digital storage

DM and the electronic storage of images may be inevitable if there is forced redundancy of FM by the manufacturers. One health benefit of a soft-copy display to radiographers would be the removal of the need to use processing chemicals.

Electronic storage of mammograms is not a trivial issue. A typical screening session involves approximately 120 Mb of data. If records from the previous screening session are required, then a further 120 Mb of data is needed (Lloyd et al 2005). With 1.6 million Australian women receiving screening mammography in a given year, the data storage capacity would need to be considerable. Diagnostic and surveillance mammograms, which may involve additional views, could involve higher volumes of electronic data per case, although without the large throughput. Lossless data compression (that retains image quality) can, however, be achieved satisfactorily.

In the display environment for reading of mammograms it is not recommended to have a film hard copy on a light box next to a digital soft copy mammogram on a workstation – for comparison purposes – due to the issue of ambient light. Ambient lighting conditions are more important for monitor viewing due to the lower luminescence levels provided by monitors. Windows need to be covered and room lighting must be indirect to prevent direct light from falling on a monitor (McLean et al 2007).

For the transition to DM, either digital mammograms will have to be printed for comparison with previous hard-copy film-screen mammograms, or hard-copy film-screen mammograms will have to be digitised for comparison with digital soft-copy images. This is likely a one-off consideration but will have associated costs and administration.

Telemammography

Telemammography is the real-time, off-site management and interpretation of digital mammograms through wireless, fibre-optic or wire-based networks between the mammography site, usually a remote area, and a centre where mammograms are interpreted. The primary aim of this approach is to improve access to mammography and hence increase early cancer detection rates. A digital system can be fitted into a mobile unit for visiting areas where there is a lack of radiologists or mammography units. Advantages of this process include efficiency for radiologists to cover multiple sites, better consultation with the patient or between physicians, less time to diagnosis and consequently less duration of patient anxiety. The system also has issues of expense, the feasibility of sending large data sets, network and storage, and protecting patient privacy (Jong & Yaffe 2005; Gitlin et al 2003; Leader et al 2006).

Computer-aided diagnosis

CAD has been designed to enhance the performance of radiologists by drawing attention to potential abnormalities on the mammogram that may have been overlooked by the reader. A computer algorithm is applied to the digital data to identify any abnormalities,

and the radiologist can then determine whether or not the patient should be recalled for additional evaluation. The following flow diagram shows the place of CAD in the context of other reading methods, in terms of diagnostic accuracy:

1 generalist radiologist < 1 generalist radiologist + CAD < 1 sub-specialised breast radiologist < 1 generalist radiologist + 1 sub-specialised breast radiologist < 2 sub-specialised radiologists.

The use of CAD is more efficient when used in conjunction with DM, as the use of FM requires film digitising before CAD (Jong & Yaffe 2005; Feig & Yaffe 1998).

Advanced applications of digital mammography

Contrast-enhanced DM involves intravenous administration of an iodine-based contrast medium in FFDM to enhance the contrast. This permits the detection of cancerous mass obscured by dense tissue and reduces the radiation dose associated with standard DM.

Tomosynthesis aims to eliminate the obscuring effect that superimposed structures have on the perception of lesions. The movement of a radiation beam in an arc about the stationary breast provides a series of low-dose DM images. The images are shifted and added together to achieve tomographic views of the breast. Nevertheless, it has 1 to 1.5 times more radiation exposure than a mammogram (Jong & Yaffe 2005).

Consumer issues

Privacy and data security

The electronic storage of data raises privacy and ethical concerns associated with the safety of electronic transfer and the storage of personal image files. Electronic data are more readily accessed and extracted than hard-copy data, so encryption, control mechanisms, certificates and keys are likely to be required (Lloyd et al 2005). Informed consent is likely to be a prerequisite to participation in digital screening programs.

For diagnostic and surveillance mammograms purchased under Medicare, women's retention of access to their personal data (originally films, now files) may also be pertinent. Consumer access to personal DM records may need to be considered.

Impact of film redundancy

Manufacturers are unlikely to support FM in the near future, as vendors and developers are promoting digital technology as the successor to FM. Currently the consumer is likely to bear any out-of-pocket costs associated with this technological change.

Rural and remote patients

As the suitability of a DM image (whether under-exposed or over-exposed) can be immediately ascertained and corrected, it is likely that technical recalls will be fewer with

DM than with FM. This will benefit women in rural and remote areas who have to travel large distances to be reassessed.

Similarly, the long waiting times between having a mammogram and being informed of the results may be reduced with the transmission of digital images, via telemammography, to a radiologist who can read the image in real time.

Breast compression

When low-dose FM was introduced, most women found mammography uncomfortable and would have preferred less compression of their breasts during the procedure (Kimme-Smith 1999). One study found that a substantial minority (17%) reported that the pain experienced during the procedure lessened their likelihood of attending mammography in the future (Sapir et al 2003). Compression of the breast will always be necessary to limit movement and to spread the tissue, but a 20 per cent reduction in breast compression may be useful in some women (Kimme-Smith 1999).

Phantom studies have suggested that it is possible to increase the kilovolt peak (kVp) and reduce the mean glandular radiation dose in digital imaging, relative to film-screen imaging, improving the signal-to-noise ratio (Kimme-Smith 2000). It has therefore been proposed that if the mean glandular dose is kept at the same level as in FM, breast compression during a DM procedure could be decreased (by increasing the kVp) (Kimme-Smith 2000). Furthermore, a potential advantage of DM is the ability to digitally enhance the contrast of images later (Kimme-Smith 2000). Thus, retrospective contrast enhancement might be used to compensate for any loss of contrast due to decreased compression (Kimme-Smith 2000).

Diagnostic and screening procedures aim to keep radiation doses as low as possible to limit potential safety issues. The possibility of reducing radiation dose through DM could be weighed against the ability to reduce breast compression for the sake of reducing discomfort, with the possible benefit of increasing compliance with further screening. However, the evidence assessing the potential for a reduction in breast compression through the use of DM is currently limited, and more definitive research is required.

What are the economic considerations?

Background

In assessing a new service, MSAC considers not only the comparative effectiveness and safety of the service under review, but also the comparative economic implications. An economic evaluation is performed in which the service under MSAC consideration is substituted for the main comparator. The purpose of the economic evaluation is to inform decision-makers of the additional costs and gains (health or other socially relevant outcomes) of the proposed service over the comparator when used in the Australian health care system. This is to ensure that society's ultimately scarce resources are allocated to those activities from which we will get the most value. That is, it seeks to enhance economic efficiency.

What are the economic considerations for digital mammography?

On the basis of the systematic review presented in this report, the advisory panel determined that DM was at least as safe and as effective as FM overall. The advisory panel also agreed that DM, as a means of breast cancer screening, appeared to be associated with improved diagnostic accuracy in women under the age of 50 years, women with radiographically dense breasts, and pre- or perimenopausal women, relative to FM. Therefore, the costs of DM and FM for the overall population were compared in settings of breast cancer screening, diagnosis and surveillance. In addition, in a screening setting the cost-effectiveness for the subgroups in which DM is more accurate than FM was analysed.

The use of DM as an alternative to conventional FM has numerous economic implications. Cost categories of importance include the professions involved in conducting mammography, the differential workload and occupational safety; the capital costs of the two systems and associated equipment; the use of consumables associated with the two technologies; and the costs of data storage.

Implications for the associated workforce

DM is associated with improved efficiency and improved safety for radiographers, as the use of chemicals in film development and the handling of film cassettes are avoided. The work of assistants who organise the film evaluation by putting film mammograms on the roller viewer for interpretation by the radiologists would be reduced as well. The workload for radiologists in terms of time spent on a mammogram could increase as a result of the increased viewing options for digital images and extra electronic reading assistance (Berns et al 2006; Ciatto et al 2006; Legood & Alastair 2004).

Differential capital costs

Capital inputs are defined as inputs that continue to deliver services for more than 1 year. Economic costs focus on the cost of resources used over a specific period (1 year in this

analysis), rather than at the time they are purchased. The economic cost takes into account the value of alternative opportunities for using the resources tied up in the capital inputs. The annual economic cost is higher than the annual financial cost, because the investment of funds 'up front', to pay for the equipment in full at the start of its use, has an opportunity cost. The difference in economic costs of the capital inputs will have an impact on the cost comparison of the technologies. Currently, DM is the more expensive technology, so it has a higher annual economic cost than FM. Apart from mammography units, the other capital inputs associated with FM include film processors, cassettes and roller viewers; those associated with DM include viewing or reporting monitors and the PACS. Expert opinion suggests that the prices of DM machines and PACS will continue to decrease. Therefore, in the future the incremental cost of DM may need to be re-estimated.

A laser printer might be required to print digital images as hard copies, particularly in a diagnostic setting. In addition, in the transitional stage where a woman's previous mammograms are conventional film and the new mammograms are digital, digitisers might be needed to digitise the film images. This is of more relevance in the screening and surveillance settings than in the diagnostic setting.

Comparative use of consumables

One economic advantage of DM is the reduction in consumables needed to perform one procedure. In conventional FM, films and chemicals are essential. In DM these items may be redundant. However, for the sake of data safety and convenience, digital images may be printed on a laser printer and on film, particularly in the diagnostic setting, where digital storage is often not used. Hard-copy printouts of digital mammograms might still cost less in consumables than conventional FM, but the advantage of DM in such a scenario is less obvious.

Data storage – physical versus digital

Another economic advantage of DM can be seen in the digital storage of data. The physical storage of conventional films requires expenditure on storage space, capital equipment, folders, energy, cleaning, maintenance and administration. These resources might be saved when digital mammograms are stored as digital files only. However, the use of PACS requires expenditure on set-up and annual maintenance or updating.

Effectiveness and changes in health-related quality of life

As there was no evidence to indicate a significant overall difference between DM and FM in terms of diagnostic accuracy in either a screening or diagnostic setting, no effectiveness data were included in the economic analysis, and a cost comparison was done.

Given that DM appears to be more accurate than FM in screening for breast cancer in women under the age of 50 years, women with radiographically dense breasts, and pre- or perimenopausal women, the cost-effectiveness for these populations was analysed. As the primary outcome of the DMIST trial (Pisano et al 2005b) was the difference between DM and FM in the AUC, this outcome was used in the cost-effectiveness analysis. However, this effectiveness measure is not intuitive and may be difficult to interpret, and is not a

directly clinically relevant outcome. Thus, the number of additional cancers detected by DM was also used as a measure of effectiveness.

The increased cancer detection rate (as an intermediate outcome measure) from DM is likely to be associated with clinically relevant outcomes such as increased survival, and consequently quality-adjusted survival, of those patients whose malignancies have been detected early. As reported by the DMIST trial, the cancers detected by DM but missed by FM in women under the age of 50 years, women with heterogeneously or extremely dense breasts, and pre- or perimenopausal women included many invasive and high-grade *in situ* cases; these are the lesions that must be detected early to save lives through screening (Pisano et al 2005b). However, there was no rigorous evidence of survival benefit or quality-adjusted survival due to DM in these subgroups. Consequently, effectiveness in terms of quality-adjusted survival was not included in this analysis.

Existing literature

When this assessment was performed, no economic analyses of DM in Australia were available for digital mammography. There were, however, cost-comparison analyses reported in the international literature. Circumstances vary considerably from country to country; for example, the costs of labour, capital and consumables, health care practices and the type of technology used might differ. Variability in economic evaluations of DM in different international settings may reflect these differences.

An Italian study (Ciatto et al 2006) found that *radiographers'* time per procedure did not differ significantly between DM and conventional FM, but *radiologists'* time for evaluation differed significantly, between 14 seconds for conventional FM and 25 seconds for DM. In contrast, a UK study found significant time savings for radiographers of approximately 6 minutes per woman (Legood & Alastair 2004). A US study found significantly lower acquisition time and significantly higher interpretation time of DM images (Berns et al 2006). Another US study found that radiologists spent less time interpreting soft-copy images (Pisano et al 2002).

These results show that evidence on the economic impact of DM is inconclusive and not transferable to the Australian setting, reflecting differences in national practice styles, applied methodologies, and the actual procedures and technologies compared.

Methods of economic evaluation

Populations used and types of analyses

The type of economic analysis conducted is conditional on the results of the systematic literature review of the safety and effectiveness of DM. The populations and corresponding types of analyses are summarised in Table 25.

In the cost comparison, all cost categories described earlier are included. The outcome is measured as the incremental cost per examination per year of DM over FM.

In the cost-effectiveness analysis of subgroups of women younger than 50 years, pre- or perimenopausal women, and women with heterogeneously or extremely dense breasts in

a screening setting, as the difference between the AUC was the primary result of the DMIST trial (Pisano et al 2005b), the cost-effectiveness of DM compared with FM was measured in terms of the incremental cost per additional AUC. To enhance the interpretation of the cost-effectiveness data, a more clinically relevant outcome in terms of the incremental cost per additional cancer detected is also presented.

Table 25 Populations and corresponding type of economic analysis

Mammography purpose	Population		Type of analysis	Outcome measure
Screening	Asymptomatic women	Overall	Cost comparison	Incremental cost/examination/year
		Age <50 years	Cost-effectiveness	Incremental cost/extra AUC; incremental cost/extra cancer detected
		Premenopausal and perimenopausal	Cost-effectiveness	Incremental cost/extra AUC; incremental cost/extra cancer detected
		With heterogeneously dense or extremely dense breasts	Cost-effectiveness	Incremental cost/extra AUC; incremental cost/extra cancer detected
Diagnosis	Symptomatic women – overall		Cost comparison	Incremental cost/examination/year
Surveillance	Women at high risk – overall		Cost comparison	Incremental cost/examination/year

Approach to the economic evaluation

Costs were analysed from the perspective of Australian society. Hence, all non-trivial differential costs were taken into account. Both costs and effectiveness were estimated over a 1-year period. All costs were expressed in Australian dollars at December 2006 prices. Costs estimated from an earlier year were adjusted with the Consumer Price Index reported by the Australian Bureau of Statistics.

Inputs to the economic evaluation

Assumptions used in the economic evaluation

As mammography practice varies from place to place, a number of assumptions have been made to confine the economic evaluation to the most likely scenarios. Based on discussions with the advisory panel, the following assumptions were made:

- The cost of space occupied by the two technologies and the allocation of the costs of overheads (including administration, cleaning and power) are assumed to be the same for both DM and FM.
- All DM images are to be printed on hard copies in the diagnostic setting, but only the examinations that need to be further assessed will most probably be printed in the screening setting (in this case, cancer detection rate is used to determine the proportion of examinations that need to be printed).
- For DM, prior conventional film images are assumed not to need digitisation, as film digitisation should be a one-off cost.
- Given the considerable variation in costs between CR and DR, two scenarios are analysed for the diagnostic setting: one using CR without PACS and one using DR with PACS.

- The throughput of DR is assumed to be 1.4 times that of FM in a screening setting, and 1.8 times in a diagnostic setting[6]. The throughput of CR is the same as that of FM.

- Different types of PACS will be used for screening and diagnostic DM: enterprise PACS in the screening setting and clinical PACS in the diagnostic setting.

Potential cost categories and cost items in the economic evaluation

Tables 25–28 summarise cost categories and cost items of importance for the economic comparison of DM with FM.

The cost inputs of human resources are presented in Table 26. DM is associated with a longer reading time but a shorter acquisition time. The time of film handling assistants (film-hangers) is saved.

In estimating the cost of the contribution of each profession per mammography examination, the full cost of an employee is represented by the individual's gross earnings, including the take-home pay plus any additional benefits such as superannuation and incentive payments. In this assessment, a 29 per cent oncost is included in addition to the salaries presented in Table 26.

Table 26 Comparative workforce implications of DM and conventional FM

Profession	Average gross salary per annum (plausible range)	Average time spent on one patient (plausible range)
DM:		
Radiologists	$350 000 ($200 000–$500 000)	5 [3–7] min for double-reading screening mammography (Berns et al 2006), 24 [20–26] min for diagnostic mammography [a]
Radiographers	$60 000 ($50 000–$70 000)	10 [8–12] min
Assistants	$40 000 ($35 000–$45 000)	n/a
FM:		
Radiologists	$350 000 ($200 000–$500 000)	3 [2–4] min for double-reading screening mammography, 15 [13–17] min for diagnostic mammography
Radiographers	$60 000 ($50 000–$70 000)	15 [13–17] min
Assistants	$40 000 ($35 000–$45 000)	5 [2–10] min

a: Estimates based on a 57% longer interpretation time for DM than for FM for screening asymptomatic women (Berns et al 2006).

The above estimates are based on guidance from the MSAC advisory panel or BreastScreen Victoria, except where noted.

Numbers in parentheses are the plausible ranges for that variable used for the sensitivity analysis.

n/a = not applicable

As noted earlier, economic costs focus on the resources used over a specific period (1 year in this analysis), rather than at the time they are purchased. Therefore, the capital costs need to be annualised to obtain an annual economic cost. Annualisation means that capital costs are not simply depreciated, but that the foregone interest on the initial

[6] This is based on estimates provided by the Advisory Panel of 10 minutes per procedure for film-screen mammography and 7 minutes per procedure for DR in a screening setting; and 15 minutes for film-screen mammography and 8 minutes for DR in a diagnostic setting. This relative productivity of the two technologies has been confirmed by a study in the USA (Healthcare, 2007).

investment is taken into account. The annualisation factor is determined by the discount rate and useful life of capital inputs. It is calculated from the following formula:

$F = ((1 + r)^t - 1) / r(1 + r)^t$

where r is the discount rate and t is the number of years of useful life of the capital items.

The annual economic cost of each item is the current value of the item divided by the annualisation factor.

The discount rate is assumed to be 5 per cent at the base case. An alternative of 3 per cent is examined in a sensitivity analysis. The purchasing prices of the capital inputs and their corresponding average annual maintenance costs and useful lifetime are summarised in Table 27.

Table 27 Comparative capital costs of mammography systems and associated equipment

Cost item or relevant variable	DM (plausible range)	FM (plausible range)
Mammography system		
Capital cost of mammography system	DR: $450 000 ($400 000–$500 000) CR: $225 000 ($220 000–$265 000)	$80 000 ($75 000–$120 000)
Lifetime of mammography system (years)	7 (6–10)	10 (7–11)
Average maintenance cost/year	DR: $40 000 ($35 000–$45 000) CR: $20 000 ($16 000–$22 000)	$7000 ($6000–$10 000)
Number of patients/mammography/year	Screening: 5600 (4200–16 800 [a]) Diagnosis (DR): 4680 (3600–14 400 [a])	Screening: 4000 (3000–12 000 [a]) Diagnosis: 2600 (2000–8000 [a])
Picture archiving and communication system (PACS)[b]		
Set-up cost of PACS	Enterprise: $800 000 ($680 000–$920 000) Clinical: $150 000 ($127 500–$172 500)	n/a
Average lifetime of PACS (years)	6 (5–7)	n/a
Average updating and maintenance cost/year	Enterprise: $80 000 ($68 000–$92 000) Clinical: $20 000 ($17 000–$23 000)	n/a
Capacity of PACS (in number of patients)	Enterprise: 75 000 (70 000–100 000) Clinical: 5200 (4000–10 000)	n/a
Laser printer (for hard copies of digital images)		
Capital cost of printer	$26 000 ($20 000–$40 000)	n/a
Lifetime of printer (years)	5 (3–7)	n/a
Average maintenance cost/year	$2600 ($2000–$4000)	n/a
Workload of printer/year (the number of examinations)	4680 (3600–14 400)	n/a
Viewing or reporting monitor		
Capital cost of reporting monitor	$35 000 ($30 000–$40 000)	n/a
Average lifetime of reporting monitor (year)	5 (3–7)	n/a
Workload of reporting monitor (number of patients)	Screening: 5600 (4200–16 800)	n/a

Cost item or relevant variable	DM (plausible range)	FM (plausible range)
	Diagnosis: 4680 (3600–6000)	
Film processor		
Capital cost of film processor	n/a	$55 000 ($50 000–$60 000)
Lifetime of film processor	n/a	7 (5–10)
Average maintenance cost/year	n/a	$5500 ($5000–$10 000)
Workload of film processor/year (number of patients)	n/a	5200 (4000–16 000)
Roller-viewer		
Capital cost of roller viewer	n/a	$25 000 ($20 000–$40 000)
Lifetime of roller viewer	n/a	10 (8–15)
Average maintenance cost/year	n/a	$2000 ($1500–$4000)
Workload of roller viewer/year (number of patients)	n/a	Screening: 4000 (3000–12 000) Diagnosis: 2600 (2000–8000)
Cassettes		
Capital cost of cassette (18 × 24 cm)	n/a	$360 ($300–$400)
Capital cost of cassette (24 × 30 cm)	n/a	$400 ($350–$450)
Proportion of women screened or diagnosed with 24 × 30 cm film	n/a	15%
Lifetime of cassette	n/a	5 (4–6)
Number of cassettes used per procedure	n/a	8
Number of procedures per year	n/a	2600 (2000–8000)

n/a = not applicable

a: The maximum estimated annual throughput of 12 000 patients/year in a screening setting and 8000 patients/year in a diagnostic setting for FM are based on 10 or 15 minutes/procedure, 8 hours/day and 50 weeks/year. This maximal throughput is unlikely to be realised. The annual throughput of DM is calculated as 1.4 times that of FM in a screening setting and 1.8 times in diagnostic setting.

b: Estimates are based on discussions with the companies that provide PACS. To protect the companies' pricing confidentiality, the estimates are only indicative. Costs of hardware and network communication are not included.

The above estimates are based on guidance from the MSAC advisory panel or BreastScreen Victoria, except where noted.

Based on the assumption that the throughput of DM would be 1.4 times that of FM in a screening setting and 1.8 times in a diagnostic setting, the maximum annual throughput of DM could reach 16 800 and 14 400 respectively. However, this high workload may not be realisable since, firstly, radiographers' ability to work at this pace may raise occupational health and safety concerns and, secondly, there may not be sufficient numbers of patients, particularly in a single hospital. The use of this maximum estimate in the sensitivity analysis aims to illustrate the impact of throughput on the incremental cost per examination of DM.

As the costs of PACS depend on a number of factors, including the number of examinations per year, the number of workstations attached to the PACS, the number of modalities sending images to the PACS, the number of sites connected to the core PACS and the data storage format, it is not possible to estimate cost without pre-specifying the components of the PACS under evaluation. Therefore, the following scenarios have been assumed for the cost estimates associated with PACS:

Enterprise PACS (for use with mammographic screening and surveillance):

- 75 000 examinations per year
- Workflow Information Service Engine (WISE) database

- 6 ImageServer/s online data stores ('s' = short-term storage) for 1 core site and 5 additional fixed sites
- 1 ImageServer/fs archive ('fs' = file system) for central archiving
- Unlimited mammographic unit support
- Unlimited workstation support
- Control tower for analysis of system usage, workstation usage etc
- Integration with radiology information systems (RIS) via Health Level 7 (HL7) communication protocol
- Digital Imaging and Communication in Medicine (DICOM standard) work list
- 12 viewing or reporting workstations with CAD support and keypads
- 12 radiographer workstations with CAD support for PACS quality assurance
- 3 telemammography workstations for mobile vans to send images back to the main PACS

Clinical PACS (for use with diagnostic mammography):

- 5200 examinations per year
- WISE database
- 1 ImageServer/s for main site
- 1 ImageServer/fs for central archiving
- Up to 4 workstations
- 2 mammographic unit connections
- HL7 RIS interface
- 2 reporting workstations with CAD support and keypads
- 2 radiographer workstations for PACS quality assurance

Expert opinion suggests that PACS is likely to be used by the whole radiological practice in the diagnostic setting, with DM sharing about 10 per cent of the overall usage. Therefore, 10 per cent of the total cost of PACS has been allocated to DM to estimate the cost per examination when DR is used as a means of breast cancer diagnosis. In the screening setting, PACS will be used solely for DM, and therefore the full cost of a PACS has been used to derive the cost per examination.

The costs of PACS presented in Table 27 do not include hardware or network communication. The costs of hardware depend on the size and configuration of each site. It is also a prerequisite for both Enterprise and Clinical PACS to have remote access to

PACS clients and core via a virtual private network connection with the PACS help desk, which will impose an additional cost. However, as the PACS is the major contributor to the total cost, the exclusion of these costs is unlikely to have a substantial impact on the cost per examination.

The costs of PACS provided in Table 27 are conditional on the specified scenarios. Any changes in the actual components of PACS may result in a change in the total cost of PACS, and consequently a change in the cost per examination.

Table 28 presents the inputs of consumables used in DM and FM. More consumable inputs are required for FM than for DM.

Table 28 Consumables used in conventional FM and DM

Cost items or relevant variable	DM(plausible range)	FM(plausible range)
Cost per film 18 × 24 cm	n/a	$0.65 ($0.6–$0.8)
Cost per film 24 × 30 cm	n/a	$1.10 ($1.0–$1.2)
Percentage of procedures with 24 × 30 cm films	n/a	15%
Number of films per standard procedure (screening or diagnosis)	n/a	5 (4–6)[a]
Cost of chemicals per procedure	n/a	$0.75 ($0.6–$0.8)
Cost per film for printouts of digital images	$0.6 ($0.5–$1.0)	n/a
Number of images per printout	2 (1–4)	n/a
Number of films per standard digital procedure	2 (1–4)	n/a

n/a = not applicable

The above estimates are based on guidance from the MSAC advisory panel or BreastScreen Victoria.

a: Although the standard procedure for both screening and diagnosis is 4 images per patient, a base case of 5 is assumed, to take into account wastage from technical repeats and spoiled films.

Since mammograms are not stored in a diagnostic setting, the storage of data is relevant only to a screening setting, via PACS. Therefore, the inputs of data storage presented in Table 29 are exclusively for FM.

Table 29 Resource implications for storage of data in FM

Cost items or relevant variable	FM (plausible range)
Number of standard data sets [a] per square metre	125 (100–200) (Legood & Alastair 2004)
Capital costs per square metre per year [b]	$182 ($160–$250) (Bradley & Mozjerin 2000)
Overhead costs per square metre per year [b]	$27 ($10–$40) (Legood & Alastair 2004)

a: Per standard data set includes all the films per examination.

b: The costs are at December 2006 price adjusted with the Consumer Price Index (Australian Bureau of Statistics, 2007).

Storage costs could vary considerably, depending on the size of the storage room and the type of storage equipment. If films are stored on roller storage racks, which reduce the space required, the storage cost per data set will be reduced accordingly (Legood & Alastair 2004).

The effectiveness of DM compared with FM in the defined subgroups of women is summarised in Table 30.

Table 30 Effectiveness of DM compared with FM

Effectiveness		DM	FM	Increment [95% CI]
AUC	Women <50 years	0.84	0.69	0.15 [0.05, 0.25]
	Pre- or perimenopausal women	0.82	0.67	0.15 [0.05, 0.24]
	Women with heterogeneously or extremely dense breasts	0.78	0.68	0.11 [0.04, 0.18]
Cancer detection rate	Women <50 years	48/14 335	32/14 335	0.0011 [0.0004, 0.0018]
	Pre- or perimenopausal women	65/15 803	43/15 803	0.0014 [0.0006, 0.0022]
	Women with heterogeneously or extremely dense breasts	94/19 897	73/19 897	0.0011 [0.0002, 0.0019]

ROC = receiver operating characteristic; CI = confidence interval

Data are extracted from DMIST trial (Pisano et al 2005b). 95% CIs of differences in cancer detection rate were calculated during the evaluation.

Sensitivity analysis of the economic evaluation

Sensitivity analysis of all key variables and assumptions was conducted to examine the robustness of the cost comparison and cost-effectiveness analyses. The variables that had a considerable impact on the results were identified.

Results of the economic evaluation

Screening in the general population

The results of the cost comparison between DM and FM for breast cancer screening in the general population are summarised in Table 31. DM is $11.40 more expensive per examination per year than FM when used as a means of breast cancer screening.

Table 31 Cost comparison between DM and FM among asymptomatic women for breast cancer screening

Cost item	DM	FM	Increment
Cost of workforce/examination	$21.87	$21.39	$0.48
Cost of capital inputs/examination	$25.63	$8.70	$16.93
Cost of consumables/examination	$0	$4.34	–$4.34
Cost of hard-copy storage/examination	$0	$1.67	–$1.67
Total cost/examination	$47.50	$36.10	$11.40

DM appears more accurate than FM when used as a means of breast cancer screening among women younger than 50 years, pre- or perimenopausal women, and women with heterogeneously or extremely dense breasts, as demonstrated in the DMIST trial (Pisano et al 2005b). The results of the analysis of cost-effectiveness of DM in these subgroups are summarised in Table 32 and Table 33. As the AUC was the primary outcome of the DMIST trial but cancer detection rate is the more clinically relevant outcome, the cost-effectiveness in terms of both outcomes is presented.

Table 32 Cost-effectiveness of DM for breast cancer screening compared with FM in terms of incremental cost per extra AUC among asymptomatic subgroups of women

Subgroup	Incremental cost/ examination	Incremental AUC/examination [95% CI]	ICER (incremental cost/extra AUC) [95% CI]
Women <50 years	$11.40	0.15 [0.05, 0.25]	$76 [$46, $228]
Pre- or perimenopausal women	$11.40	0.15 [0.05, 0.24]	$76 [$48, $228]
Women with heterogeneously or extremely dense breasts	$11.40	0.11 [0.04, 0.18]	$104 [$63, $285]

ICER = incremental cost-effectiveness ratio.

Table 33 Cost-effectiveness of DM for breast cancer screening compared with FM in terms of incremental cost per extra cancer detected among asymptomatic subgroups of women

Subgroup	Incremental cost/ examination	Incremental cancer detected/examination [95% CI]	ICER (incremental cost/extra cancer detected) [95% CI]
Women <50 years	$11.40	0.0011 [0.0004, 0.0018]	$10 364 [$6333, $28 500]
Pre- or perimenopausal women	$11.40	0.0014 [0.0006, 0.0022]	$8143 [$5182, $19 000]
Women with heterogeneously or extremely dense breasts	$11.40	0.0011 [0.0002, 0.0019]	$10 364 [$6000, $57 000]

ICER = incremental cost-effectiveness ratio.

The confidence intervals for the incremental cost per extra cancer detected in each subgroup vary widely as a result of wide confidence intervals in the cancer detection rate (due to the small number of cancer cases detected). However, the base-case estimates for each subgroup are all less than or just above $10 000 per extra cancer detected, indicating that DM represents good value for money when compared with FM.

Diagnosis of symptomatic women

The results of the cost comparison between DM and FM as a means of breast cancer diagnosis in symptomatic women are summarised in Table 34. DM is $36 more expensive per examination than FM in the diagnostic setting when DR is used and $33 more expensive when CR is used. The incremental cost in the diagnostic setting is greater than that in the screening setting as diagnostic mammograms need a longer reading time, all soft-copy images are assumed to be printed on special films by laser printer, and the annualised cost of Clinical PACS per examination is more expensive than that of Enterprise PACS.

Table 34 Cost comparison between DM and FM for breast cancer diagnosis among symptomatic women

Cost item	DM	FM	Increment
If DR is used			
Cost of workforce/examination	$79.47	$57.77	$21.70
Cost of capital inputs/examination	$29.69	$11.84	$17.85
Cost of consumables/examination	$1.20	$4.34	–$3.14
Cost of hard-copy storage/examination	$0	$0	$0
Total cost/examination	$110.36	$73.95	$36.41
If CR is used			
Cost of workforce/examination	$79.47	$57.77	$21.70
Cost of capital inputs/examination	$26.21	$11.84	$14.38
Cost of consumables/examination	$1.20	$4.34	–$3.14
Cost of hard-copy storage/examination	$0	$0	$0
Total cost/examination	$106.88	$73.95	$32.94

Surveillance of women at potentially high risk of developing breast cancer

Women at a potentially high risk of breast cancer (<1% of the general female population) are encouraged to attend mammographic screening every year. Whether these women are eligible for annual screening through the national BreastScreen program depends on their jurisdictions (see Table 2). As there is no universal policy across Australia, three scenarios are assumed: (1) 100 per cent of women at high risk are screened annually through BreastScreen; (2) 100 per cent of women at high risk are screened annually through private clinics; and (3) 50 per cent are screened annually through BreastScreen and 50 per cent through private clinics.

In scenario 1, the incremental cost of DM per examination compared with FM is equal to $11.40, the same as for asymptomatic women in a screening setting. In scenario 2, the incremental cost is $36.41 when DR is used and $32.94 when CR is used. In scenario 3, the incremental cost is $23.91 (($11.40 + $36.41)/2) when DR is used and $22.17 (($11.40 + $32.94)/2) when CR is used.

Sensitivity analyses

All variables in Tables 25–28 were examined by one-way sensitivity analysis. The variables that have an important impact on the incremental cost per examination per year in screening and diagnostic settings are summarised in Tables 34–36. An important impact is defined arbitrarily as an increase or decrease in the incremental cost of at least $1 when the values of the variables are varied within the plausible ranges.

Table 35 Results of sensitivity analyses for cost comparison between DM and FM for screening

Variables	Incremental cost/examination/year		
	A	B	C
Average salary of radiologists	$8.81	$11.40	$14.00
Average time of radiologist/examination	$8.37	$11.40	$14.44
Average time of assistant/examination	$12.75	$11.40	$9.17
Average cost of FM	$11.57	$11.40	$10.11
Average cost of DM	$9.86	$11.40	$12.95
Average lifetime of DM	$13.35	$11.40	$7.92
Average workload of mammography system	$16.32	$11.40	$1.46

A = Based on the lower estimate in the plausible range for variables presented in Tables 25–28.

B = Base case

C = Based on the higher estimate in the plausible range for variables presented in Tables 25–28.

Table 36 Results of sensitivity analyses for cost comparison between DM and FM in a diagnostic setting when DR is used

Variables	Incremental cost/examination/year		
	A	B	C
Average salary of radiologists	$24.71	$36.41	$48.10
Average time of radiologist/examination	$30.34	$36.41	$36.41
Average time of assistant/examination	$37.75	$36.41	$34.17
Average cost of film mammography	$36.66	$36.41	$34.42
Average cost of DM	$34.56	$36.41	$38.25
Average lifetime of FM	$35.08	$36.41	$36.69
Average lifetime of DM	$38.73	$36.41	$32.24
Average maintenance cost of FM per year	$36.79	$36.41	$35.25
Average maintenance cost of DM per year	$35.34	$36.41	$37.48
Average workload of mammography system per year	$41.76	$36.41	$25.15
Average lifetime of reporting monitors	$37.43	$36.41	$35.97
Average number of printouts per DM examination	$35.81	$36.41	$37.61

A = Based on the lower estimate in the plausible range for variables presented in Tables 25–28.

B = Base case

C = Based on the higher estimate in the plausible range for variables presented in Tables 25–28.

Table 37 Results of sensitivity analyses for cost comparison between DM and FM in a diagnostic setting when CR is used

Variables	Incremental cost/examination/year		
	A	B	C
Average salary of radiologists	$21.24	$32.94	$44.63
Average time of radiologist/examination	$26.88	$32.94	$32.94
Average time of assistant/examination	$34.28	$32.94	$30.70
Average cost of FM	$33.19	$32.94	$30.95
Average cost of DM	$32.61	$32.94	$35.60
Average lifetime of FM	$31.60	$32.94	$33.22
Average lifetime of DM	$35.03	$32.94	$29.19
Average maintenance cost of FM per year	$33.32	$32.94	$31.78
Average maintenance cost of DM per year	$31.40	$32.94	$33.71
Average workload of mammography system per year	$37.25	$32.94	$24.02
Average lifetime of reporting monitors	$33.96	$32.94	$32.50
Average number of printouts per DM examination	$32.34	$32.94	$34.14

A = Based on the lower estimate in the plausible range for variables presented in Tables 25–28.

B = Base case

C = Based on the higher estimate in the plausible range for variables presented in Tables 25–28.

Tables 34–36 indicate that in both settings, the average workload of the mammography system has the most significant impact on the incremental cost of DM per examination per year. When the annual workload of a mammography system is increased to its maximum capacity, the incremental cost decreases from $11.40 to just $1.46 in a screening setting, from $36.41 to $25.15 in a diagnostic setting using DR, and from $32.94 to $24.02 in a diagnostic setting using CR.

Apart from the workload of the mammography system, the average gross salary of radiologists, the average time spent on one patient per radiologist and the average cost and useful lifetime of DM units have a substantial impact on the incremental cost per DM examination per year.

As the workload of a mammography system has the most significant impact on the results of cost comparison, its impact on the incremental cost-effectiveness ratios (ICERs) for the relevant population subgroups in a screening setting was examined by sensitivity analyses (Tables 37–38).

Table 38 Results of sensitivity analyses of cost-effectiveness of DM for breast cancer screening compared with FM in terms of incremental cost per extra AUC among asymptomatic subgroups of women

Subgroup	Incremental cost/ examination/year	Incremental AUC/examination [95% CI]	ICER (incremental cost/extra AUC) [95% CI]
Women <50 years	$1.46	0.15 [0.05, 0.25]	$10 [$6, $29]
	$11.40		$76 [$46, $228]
	$16.32		$109 [$65, $326]
Pre- or perimenopausal women	$1.46	0.15 [0.05, 0.24]	$10 [$6, $29]
	$11.40		$76 [$48, $228]
	$16.32		$109 [$68, $326]
Women with heterogeneously or extremely dense breasts	$1.46	0.11 [0.04, 0.18]	$13 [$8, $37]
	$11.40		$104 [$63, $285]
	$16.32		$148 [$91, $408]

$11.40/examination is the base case at a workload of 4000 patients/year per FM unit; $1.46/examination is based on a workload of 12 000 patients/year per FM unit, which is the maximum work capacity of FM in a screening setting; $16.32/examination is based on a workload of 3000 patients/year per FM unit. The workload of DM is assumed to be 1.4 times of that of FM.

CI = confidence interval. ICER = incremental cost-effectiveness ratio.

Table 39 Results of sensitivity analyses of cost-effectiveness of DM for breast cancer screening compared with FM in terms of incremental cost per extra cancer detected among asymptomatic subgroups of women

Subgroup	Incremental cost/ examination/year	Incremental cancer detected/exam [95% CI]	ICER (incremental cost/extra cancer detected) [95% CI]
Women <50 years	$1.46	0.0011 [0.0004, 0.0018]	$1327 [$811, $3650]
	$11.40		$10 364 [$6333, $28 500]
	$16.32		$14 836 [$9067, $40 800]
Pre- or perimenopausal women	$1.46	0.0014 [0.0006, 0.0022]	$1043 [$664, $2433]
	$11.40		$8143 [$5182, $19 000]
	$16.32		$11 657 [$7418, $27 200]
Women with heterogeneously or extremely dense breasts	$1.46	0.0011 [0.0002, 0.0019]	$1327 [$768, $7300]
	$11.40		$10 364 [$6000, $57 000]
	$16.32		$14 836 [$8589, $81 600]

$11.40/examination is the base case at a workload of 4000 patients/year per FM unit; $1.46/examination is based on a workload of 12 000 patients/year per FM unit, which is the maximum work capacity of FM in a screening setting; $16.32/examination is based on a workload of 3000 patients/year per FM unit. The workload of DM is assumed to be 1.4 times that of FM.

CI = confidence interval. ICER = incremental cost-effectiveness ratio.

The sensitivity analyses of the ICERs of DM and FM indicate that the incremental cost per extra cancer detected is up to $40 800 for women under 50 years, and up to $27 200 for pre- or perimenopausal women. The incremental cost per extra cancer detected in women with heterogeneously or extremely dense breasts varies widely (up to $81 600).

Financial implications

Mammography for surveillance and diagnosis

The current financial implications to the Medicare Benefits Schedule (MBS) of mammography services provided to symptomatic women and women at a high risk of breast cancer are given in Table 40. Mammography is a relatively frequently performed procedure, and the associated financial impact on Medicare is considerable.

Table 40 Current implications to the MBS of FM services provided to women at potentially high risk of breast cancer or for the diagnosis of symptomatic women

MBS item numbers	Services	Number of procedures 2005–06	Requested benefit 2005–06 (A$)
59300	Mammography of both breasts Fee: $89.50. Benefit: 75% = $67.15; 85% = $76.10	320 384	$24 075 311
59303	Mammography of one breast Fee: $53.95. Benefit: 75% = $40.50; 85% = $45.90	39 368	$1761 165
Total		**359 752**	**$25 836 476**

MBS item descriptions sourced from (MBS 2007). MBS item statistics sourced from (Medicare Australia 2006).

A change in the cost of the procedure therefore entails extensive financial implications for Medicare. Should DM replace FM, an additional $33 (CR) to $36 (DR) per examination, representing a 43 to 48 per cent increase from the current MBS scheduled fee ($76.10), would be borne by Australian society. This change in technology would represent a total additional cost of $10 to $13 million per year given the number of procedures performed in 2005–06. The financial implication of replacing FM with DM in a clinical setting is summarised in Table 41.

Table 41 Financial implications of replacing FM with DM in a clinical setting (for both high risk and symptomatic women)

Incremental cost/procedure	Number of procedures 2005–06 [a]	Incremental cost to Australian society
$36.41 (DR)	320 384	$11 665 181
	359 752	$13 098 570
$32.94 (CR)	320 384	$10 553 449
	359 752	$11 850 231

a: The number of procedures is sourced from (Medicare Australia 2006). Refer to Table 40 for the number of procedures.

Screening mammography

According to the 2003–2004 monitoring report of BreastScreen Australia, 1.1 million women aged 50–69 years were screened as part of the invited target population, and another 500 000 women aged 40–49 years were screened in the same period (AIHW 2007a). Given that the estimated incremental cost per procedure of DM is $11.40 in a screening setting, and that the number of screening examinations was 1.6 million in 2003–2004, replacing FM with DM would have a substantial incremental cost to Australian society of $18 million over 2 years for screening 1.6 million women, or around $9 million per year. The financial implication of replacing FM with DM in a screening setting is summarised in Table 42.

Table 42 Financial implications of replacing FM with DM for screening asymptomatic women

Incremental cost/ procedure [a]	Number of procedures 2003–2004 [b]	Incremental cost to Australian society over 2003–2004	Estimated *annual* incremental cost to Australian society
$11.40 (base case)	1 600 000	$18 240 000	$9 120 000
$1.46	1 600 000	$2 336 000	$1 168 000
$16.32	1 600 000	$26 112 000	$13 056 000

a: $11.40/examination is the base case at a workload of 4000 patients/year per FM unit; $1.46/examination is based on a workload of 12 000 patients/year per FM unit, which is the maximum work capacity of FM in a screening setting; $16.32/examination is based on a workload of 3000 patients/year per FM unit. The workload of DM is assumed to be 1.4 times of that of FM.

b: The number of procedures is sourced from (AIHW 2007a).

As the number of procedures sourced from Medicare and BreastScreen Australia includes women at potentially high breast cancer risk, no separate financial analysis has been conducted for these women.

Discussion

Is it safe?

Nine studies (eleven articles) on the safety of DM were identified. All of the studies that reported estimated radiation doses were based on an indirect flat-panel FFDM machine, the Senographe 2000D. When the standard automated optimisation of parameters (AOP) was used, the average glandular dose (AGD) reported ranged from 1.25 to 1.88 mGy (reported in six studies). Only two comparative studies provided AGD estimates for FM, which ranged from 1.45 mGy for a CC view to 1.92 mGy for an MLO view.

Overall, the radiation doses administered with the Senographe 2000D using standard AOP were equivalent to those by FM. A pilot study compared the image quality between images taken using standard AOP and those taken in manual mode with the same anode–filter combination and tube potential but only half the tube loading (Hemdal et al 2005a). Although the accuracy of any interpretations based on these images was not discussed, the three radiologists in the study suggested that the reduced-dose images were of sufficient image quality to allow breast cancer to be diagnosed.

Technological advances have been made in the area of DM, and the Senographe 2000D is now outdated (see 'Available digital mammography systems' in the 'Background'). The reported doses are therefore likely to be higher than now used. A phantom study using the Sectra MicroDose mammography unit, not using a clinical population as the basis for the clinical parameters, reported AGDs of 0.21 and 0.28 mGy for a 50-mm simulated compressed breast with 50 per cent glandularity (simulated by polymethylmethacrylate) (Hemdal et al 2005b). This is considerably lower than the levels estimated for the Senographe 2000D. The current mammography units are therefore likely to be safer in regards to radiation levels than FM.

Two large screening studies provided contrasting results on the rates of false positives. The larger Colorado study found that DM resulted in fewer false positives than FM, but was affected by partial verification bias, with inadequate verification of disease status for the whole cohort. The better-quality Oslo I study found that DM resulted in more unnecessary further investigations than FM. The statistical significance of this difference is unclear. However, the authors indicated that the radiologists were still learning to read digital mammograms, and thus this is likely to have affected the rate of false positive. One further study (Yamada et al 2004) supported the results of Oslo I, although the difference in false alarm rates was not statistically significant. No literature discussed any physical harm or psychological distress caused by unnecessary investigations.

Is it effective . . .

In a screening population?

The results of the four diagnostic accuracy studies and one screening effectiveness study varied widely. The Oslo I and Colorado I and II studies showed DM to have similar accuracy as but slightly less cancer detection than FM, although differences were not statistically significant. The Oslo II and DMIST studies, however, found that DM

performed similarly to FM, and had slightly but not significantly higher cancer detection overall. The latter study definitively showed significantly higher diagnostic accuracy and improved cancer detection with DM in certain population subgroups.

Skaane et al (2005) suggested that the results of the Colorado screening study and the Oslo I (2003) and II studies differed on account of the use of prototype DM equipment in the Colorado study, double reading of each mammographic type in the Oslo I and II studies and single reading in the Colorado study, and the poor display environment in the Oslo I study. To complicate matters further, most women in the Colorado study had been screened before (an incident screening population) and approximately 44 per cent were enrolled twice or thrice, so the data were partly analysed in terms of examinations, not individuals. Screening in the United States (Colorado and DMIST studies) generally occurs with single reading, as opposed to double reading in Europe and Australia.

Skaane et al (2005) also suggested that the differences in cancer detection rates and recall rates between the Oslo I and II studies (both higher in Oslo II) were likely a combination of differences in the display environments and better soft-copy reading experience in Oslo II.

Effectiveness results must be viewed in the context that diagnostic accuracy and cancer detection rate in mammography depend on radiologist reading. Inter-rater reliability of mammography reading by radiologists is low and dependent on experience (Berg et al 2002; Elmore et al 2003). It is clear from the screening diagnostic accuracy and screening effectiveness studies that experience with reading digital mammograms was much lower than with reading film-screen mammograms (Pisano et al 2005a; Skaane et al 2005). The studies thus compared a technology in which skills of interpretation and development are at their peak with a technology in which the skills are in their early stages. Developments in displays and the ability to manipulate and process digital images and improved access to training programs will likely only improve the performance of radiologists at reading digital mammograms in the future.

Even given the difference in radiologist experience between the two methods, it is clear that when the comparison was adequately powered there was no real difference in diagnostic accuracy overall. The largest high-quality diagnostic accuracy study (the DMIST study) determined that DM had slightly better diagnostic accuracy versus the reference standard than FM, but the difference was not statistically significant. The study was adequately powered for this analysis and the subgroup analyses, and as the results were also robust to statistical correction for partial verification bias, these results appear valid. The results also appear generalisable to most radiology practices, as several types of mammographic unit (DR, CR and film-screen) and display (digital in hard copy or soft copy) were included. Of considerable interest was the statistically significant improved performance of DM at detecting breast cancer in subgroups that are traditionally difficult to image: specifically, women under the age of 50 years, women who are pre- or perimenopausal, and women with heterogeneous or extremely dense breast tissue. This improved accuracy supports DM as a candidate to replace FM in these screening subgroups or in addition to FM in these subgroups (although at the risk of increased cumulative radiation dose).

The relative effects of the two methods on interval cancer rates cannot be assessed at present. The Oslo II study has the correct study design for evaluating differences in interval cancer detection rates (not confounded by the use of paired data, as in the

diagnostic accuracy studies such as DMIST and Oslo I), but has yet to present interval cancer data.

Effect on health outcomes

Using a linked evidence approach, treatment effectiveness was determined by comparing early treatment with late treatment. Early treatment refers to treatment of asymptomatic women with cancer detected by mammography screening. Late treatment refers to treatment of unscreened women presenting with clinical symptoms. Therefore, this assessment of the effect on health outcomes was limited to studies comparing screening mammography with no screening. Given the similar diagnostic accuracy of DM and FM in an asymptomatic population, it is likely that the FM screening results reported would be transferable to a population screened with DM. As discussed previously, however, asymptomatic women receiving DM who are aged under 50 years, are pre- or perimenopausal, or have heterogeneous or extremely dense breasts should receive additional health benefits.

The three studies that measured mortality risk reduction (Gøtzsche & Nielsen 2006; Humphrey et al 2002; Kerlikowske et al 1995) in women over 50 years identified statistically significant results 7–14 years after breast cancer diagnosis. A protective effect from mammography screening in women less than 50 years was not seen until 13 years after diagnosis. This lower risk reduction can be explained by the fact that these women have denser breast tissue, which makes identification of breast cancer more difficult using FM.

A limitation of this assessment of treatment effectiveness is that the available data reported only mortality from breast cancer and gave no indication of other relevant outcomes such as recurrence rates, disease progression, morbidity and quality of life.

Table 43 Body of evidence assessment matrix for screening with DM (derived from Table 11)

Component	A Excellent	B Good	C Satisfactory	D Poor
Evidence base			Level III studies with low risk of bias, or level I or II studies with moderate risk of bias	
Consistency		Most studies consistent and inconsistency may be explained		
Clinical impact			Moderate [a]	
Generalisability		Population/s studied in the body of evidence are similar to the target population		
Applicability		Applicable to Australian health care context with few caveats		

a: In women in certain subgroups; otherwise the clinical impact would be similar to that already being achieved with FM.

In a diagnostic population?

The four studies included in the assessment of diagnostic accuracy in women at potentially high risk and symptomatic women reported similar diagnostic properties between FM and DM. Although FM achieved higher test sensitivity than DM in those studies, no statistically significant differences between the two methods were observed. Furthermore, test specificity, AUC and PPV were similar for FM and DM.

A major limitation of the assessment stemmed from the patient groups enrolled in each of the four studies. Ideally, the diagnostic accuracy of DM should be analysed separately for women at potentially high risk and symptomatic women. These two groups of women have distinct risks of malignancy at the time of diagnostic mammography and follow very different clinical courses of management. Regrettably, each study recruited a mixture of women at potentially high risk, symptomatic women and women referred for diagnostic mammography for other reasons (including abnormal screening mammograms, patient anxiety, proximity to clinic and breast augmentation). In the absence of any relevant subgroup analyses, it is unclear whether the diagnostic properties reported for FM and DM would be applicable to either women at potentially high risk or symptomatic women.

In the three studies that provided sufficient information on the reference standard used to diagnose malignancy, biopsy was performed only in a subset of the women undergoing diagnostic mammography. Although it is clearly unethical to subject all women in a mixed diagnostic population to biopsy, the decision to analyse only a subset of women by the reference standard introduced partial verification bias to the results. Other more general criticisms of the four studies include a general lack of power to detect potentially real differences between FM and DM (Venta et al 2001), poorly described methods (Hendrick et al 2001), methodological shortcomings such as patients being assessed with only one method of mammography (Seo et al 2006), and a paucity of information on the age, breast density and menopausal status of the women in these studies.

Given the small number of studies available and their associated limitations, it is difficult to make any strong statements regarding the performance of DM in women at potentially high risk and symptomatic women relative to FM. However, women at high risk in particular are likely to present for surveillance at a younger age than the general population, and thus are more likely to be pre- or perimenopausal with dense breasts (MSAC 2007). Thus, the improved diagnostic accuracy of DM in these subgroups of women in a screening population in the DMIST study would appear to be transferable to women at potentially high risk – leading to perhaps even further increased benefits given that these women are starting at a higher baseline risk for breast cancer than the general population.

Effect on health outcomes

Using a linked evidence approach, treatment effectiveness was determined by comparing early treatment versus late treatment. Early treatment, in this case, refers to treatment of asymptomatic women detected by mammography screening. Late treatment was characterised by women not undertaking screening mammography, but presenting with clinical symptoms. As such, no studies could be identified that assessed early treatment in a symptomatic population. Among women who present symptomatically, it is unlikely that there would be any differences in treatment as a consequence of DM, given the similar diagnostic accuracy to FM, even if subgroups effects are eventually identified. The effectiveness of current treatments primarily depends on tumour stage and characteristics (National Comprehensive Cancer Network (NCCN) 2006; NHMRC 2001).

No systematic reviews or meta-analyses of trials assessing treatment effectiveness in women at potentially high risk were identified either. However, the significant mortality risk reduction found in the studies comparing mammographic screening with no screening in an asymptomatic population is likely to be higher in women who are at high risk of breast cancer, given they often present for surveillance at a younger age.

Table 44 Body of evidence assessment matrix for surveillance and diagnosis with DM (from Table 11)

Component	A Excellent	B Good	C Satisfactory	D Poor
Evidence base			Level III studies with low risk of bias, or level I or II studies with moderate risk of bias	
Consistency		Most studies consistent and inconsistency may be explained		
Clinical impact				Slight or restricted [a]
Generalisability		Population/s studied in the body of evidence are similar to the target population		
Applicability			Probably applicable to Australian health care context with some caveats	

a: On the basis of the evidence presented for this population. Extrapolation of evidence from the screening setting indicates a likely moderate clinical impact, over and above FM, for asymptomatic women at potentially high risk of breast cancer.

What are the economic considerations?

DM is more expensive than FM both as a means of screening and for diagnosis. Given that in a diagnostic setting the mammographic throughput is lower, images require a much longer reading time and all digital images are printed, the incremental cost of DM is greater than in a screening setting. Although a small proportion of digital images (at cancer detection rate) in a screening setting would also be printed for further assessment, the cost associated with this process had only a trivial impact on the incremental cost per examination of the two technologies given the very small cancer detection rate.

The cost-effectiveness analysis related to specific population subgroups used two outcomes: the incremental cost per additional AUC and the incremental cost per additional cancer detected. Although the AUC was the primary outcome of the DMIST trial, it is less clinically relevant than cancer detection rate. The cost-effectiveness analysis indicated that DM appears to represent good value for money, in terms of the incremental cost per additional cancer detected, particularly in subgroups of women under 50 years and pre- or perimenopausal women. A wide range of ICERs was found in the subgroup of women with heterogeneously or extremely dense breasts.

Sensitivity analyses demonstrated that the workload or throughput of a mammography system had the most significant impact on the comparative cost of the two technologies. The throughput may vary from state to state and from hospital to hospital. When

throughput increases, the incremental cost of DM decreases. Therefore, the operational efficiency of each DM system needs to be considered.

There are a number of limitations to this evaluation. Firstly, the costs are only indicative. To obtain a more accurate estimate, a time and motion study could observe a sample of radiologists, radiographers and assistants at work. Secondly, the cost estimates are based on a static centre. As the throughput of a mobile unit may be very different from that of a static centre, the estimated cost in this evaluation may not represent costs in mobile units. Moreover, the cost per procedure in a mobile unit is associated with an additional cost of transporting films to static centres. Therefore, the cost per procedure of FM in mobile units is higher than that in static centres, and consequently the incremental cost per procedure of DM in a mobile unit is likely to be less than that estimated in this report.

DM is associated with a higher cost per mammography unit, a longer soft-copy reading time and a high cost of PACS, whereas FM is associated with a longer image acquisition time, more staff time for handling, archiving and filing films, and physical storage of films. Electronic hardware and software continue to come down in price, at least those associated with high-volume usage, while the costs of personnel and real estate continue to increase (Ciatto et al 2006). Therefore, the incremental cost of DM is likely to change in the future and should be updated regularly.

The total costs of DM and FM are unknown. Only the differential costs were considered in this evaluation. The results of the economic evaluation are conditional on the assumptions and inputs of the evaluation. The purpose of the evaluation is to synthesise evidence and assumptions so as to allow decision-makers to understand the relations between assumptions and outcomes. Should a better estimate be available for any of the variables, the economic evaluation would need to be updated.

A number of advantages of DM have been alluded to in the 'Other relevant considerations' section but could not be quantified in this economic evaluation.

What are the other relevant considerations?

FM is unlikely to be supported by vendors and developers of breast imaging equipment in the near future. DM is replacing FM. Currently any out-of-pocket costs associated with this technological change are likely to be borne by the consumer.

The Australian shortage of radiologists and radiographers is another factor driving this change to DM, which maximises workforce flexibility and capacity. The accessibility of the same digital mammograms from different reading stations through a PACS could reduce the need for radiologists to travel between centres for film reading, and thus improve their work flexibility and efficiency.

DM might reduce the need to recall women for repeat screens, as fewer mammograms may need to be repeated for technical reasons, and acquired images can be processed and magnified (Legood & Alastair 2004). Clinical research is needed to quantify these savings before they can be considered in an economic evaluation.

Telemammography could substantially reduce the waiting time for women receiving their results, as well as the need for recall of women in rural and remote areas who have to

travel large distances to be reassessed. DM could thus reduce anxiety among the women screened and their families.

DM has the potential to reduce the radiation dose exposure to the patient. Radiographers also benefit through not being exposed to processing chemicals and not having to handle film cassettes (DR mammography only, not CR mammography).

Conclusions

Safety

The literature identified in the systematic review suggests that the radiation exposure from DM is equivalent to that from FM. However, phantom studies of more recent mammography units suggest that the radiation dose from DM is likely to be lower than that from FM. It can therefore be concluded that DM is as safe as, or safer than, FM in regards to radiation dose.

The best-quality evidence on false positive rates suggests that DM results in slightly more unnecessary further investigations than FM. The statistical significance or clinical relevance of this difference is unclear, and these results contrast with a large screening study that reported lower false positive rates from DM. This evidence, in conjunction with the diagnostic accuracy data supporting screening effectiveness, suggests that the false positive rates of the two methods are similar, so DM is unlikely to result in extra unnecessary intervention eg radiation exposure and biopsy.

Effectiveness

As a potential replacement for FM, DM is as accurate when used as a screening method in *asymptomatic* women. The case for replacing FM is not obvious overall, given the similar accuracy and cancer detection rates of the two methods. DM would, however, appear to be a reasonable alternative to FM on the basis of the population-based effectiveness data alone. In contrast, DM is *more accurate* in detecting breast cancer in women who are conventionally difficult to image with FM, specifically women aged under 50 years, those who are pre- or perimenopausal, and those with heterogeneously dense or extremely dense breasts. Thus, DM should replace or be used in addition to FM in these women (although the latter increases the cumulative radiation risk).

The relative impact of DM and FM on interval cancer rates has yet to be properly established.

Given the similar diagnostic accuracy of DM and FM in an asymptomatic population, it is likely that the health benefits from FM versus no screening would be transferable to a population screened with DM. For women aged over 50 years there appears to be a clear benefit from mammography screening in the reduction of relative risk of mortality ranging from 22 to 27 per cent between 7 and 14 years after diagnosis. On the other hand, in women less than 50 years old, there appears to be a reduction of 15 to 24 per cent relative risk of mortality, but only approximately 13 years after diagnosis. These health benefits should be greater for women who are at potentially high risk of breast cancer who receive mammography for surveillance.

If DM replaces FM for screening asymptomatic women who are aged under 50 years, pre- or perimenopausal, or have dense breasts, it is likely that the mortality reduction would be higher than currently seen in these subgroups.

For the diagnosis of women at potentially *high risk* or *symptomatic* women, it is unclear whether DM is as effective as FM. Although no significant differences in the diagnostic

properties of the two methods have been reported in the literature, it is difficult to make any strong conclusions on the available evidence. However, it appears likely that as women at potentially high risk present for surveillance at a younger age than the general population, and thus are more likely to be pre- or perimenopausal with dense breasts, the improved diagnostic accuracy found with DM in these population subgroups in the DMIST screening study would be applicable – leading to perhaps even further increased benefits over FM given that these women are starting at a higher baseline risk for breast cancer than the general population.

There is currently no evidence to suggest that DM should be used *in addition to* FM in a *symptomatic* or *surveillance* population.

Benefits from breast cancer treatment in symptomatic women are well known and should remain the same if DM replaces or is used as an alternative to FM for diagnosis – with the proviso that the similar diagnostic accuracy between the methods should be confirmed by more evidence.

Cost comparison and cost-effectiveness

The cost-comparison analysis suggested that DM is more expensive than FM. The incremental cost per examination of DM is significantly affected by the workload of mammography system.

The cost-effectiveness analysis of DM as a means of breast cancer screening in women younger than 50 years, pre- or perimenopausal women and women with heterogeneously or extremely dense breasts indicates that DM appears to represent good value for money in terms of the incremental cost per additional cancer detected. Sensitivity analyses suggest that the incremental cost per extra cancer detected varies widely as a result of wide confidence intervals of differences in cancer detection rates between DM and FM (due to small cancer numbers in the very large DMIST study), particularly in women with heterogeneously or extremely dense breasts.

Should DM replace FM, there would be a substantial financial impact on Australian society. However, with the wide use of DM, the prices of DM machines and associated PACS will continue to decrease, while the cost of maintaining film-screen technologies is likely to increase, and therefore in the future the actual cost borne by society may not be as great as estimated in this evaluation.

Recommendation

MSAC has considered the safety, effectiveness and cost-effectiveness of digital mammography when compared with conventional film mammography: as a screening test for breast cancer in asymptomatic women aged over 40 years or women at high risk, and in the investigation of women with symptoms of breast cancer.

MSAC finds that digital mammography is as safe and as effective as film mammography. There may be subgroups of patients in whom it is more effective.

Film mammography is being superseded by digital mammography and will lose technical support.

MSAC recommends that public funding for this procedure be supported under the arrangements that currently apply to film mammography.

> – The Minister for Health and Ageing endorsed this recommendation on 11 April 2008.

Appendix A MSAC terms of reference and membership

MSAC's terms of reference are to:

- advise the Minister for Health and Ageing on the strength of evidence pertaining to new and emerging medical technologies and procedures in relation to their safety, effectiveness and cost-effectiveness and under what circumstances public funding should be supported
- advise the Minister for Health and Ageing on which new medical technologies and procedures should be funded on an interim basis to allow data to be assembled to determine their safety, effectiveness and cost-effectiveness
- advise the Minister for Health and Ageing on references related to new or existing medical technologies and procedures
- undertake health technology assessment work referred by the Australian Health Ministers' Advisory Council and report its findings to the Council.

The membership of MSAC comprises a mix of clinical expertise covering pathology, nuclear medicine, surgery, specialist medicine, general practice, clinical epidemiology, clinical trials, health economics, consumer health and health administration.

Member	**Expertise or affiliation**
Dr Stephen Blamey (Chair)	general surgery
A/Prof John Atherton	cardiology
Professor Syd Bell	pathology
A/Prof Michael Cleary	emergency medicine
A/Prof Paul Craft	clinical epidemiology and oncology
Dr Kwun Fong	thoracic medicine
Dr David Gillespie	gastroenterology
Dr Debra Graves	medical administration
Professor Jane Hall	health economics
Professor John Horvath	Chief Medical Officer, Department of Health and Ageing
A/Prof Terri Jackson	health economics
Professor Brendon Kearney	health administration and planning
Professor Frederick Khafagi	nuclear medicine
Dr Ray Kirk	health research
Associate Professor Donald Perry-Keene	endocrinology
Dr Ewa Piejko	general practice
Mrs Sheila Rimmer	consumer health

Professor Ken Thomson	radiology
Dr Douglas Travis	urology
Dr Mary Turner	Australian Health Ministers' Advisory Council representative
Dr David Wood	orthopaedics
Assistant Secretary	Medical Benefits Schedule Policy Development Branch, Department of Health and Ageing

Appendix B Advisory panel and evaluators

Advisory Panel Reference 37 – Digital Mammography

Chair	member of MSAC
A/Prof Paul Craft	
Medical Oncology Unit, Canberra Hospital, ACT	
Evaluators (AHTA)	
Ms Tracy Merlin, Lead Researcher and Manager	
Ms Hedyeh Hedayati, Research Officer	
Dr Shuhong Wang, Health Economist	
Mr Thomas Sullivan, Research Officer	
Ms Skye Newton, Research Officer	
Mrs Liz Buckley, Research Officer	
Mr Florian Kreisz, Health Economist	
Ms Christina Zimprich, visiting intern	
Prof Janet Hiller, Director	
Panel Members	
Dr David Barton Medical Adviser, Diagnostics and Technology Branch, Department of Health and Ageing	Department of Health and Ageing, Diagnostics and Technology Branch
Associate Professor Richard Bell Director of Cancer Services, Barwon Health, Geelong Hospital Chairman, Board of Management, BreastScreen, Victoria Senior Clinical Consultant, Cancer Council Victoria	Medical Oncology Group of Australia nominee
Mr John Buckingham MB BS MS (Minn) FRACS FACS Surgeon, Calvary Hospital, Bruce ACT	Royal Australasian College of Surgeons nominee
Dr Debra Graves Medical Administrator, Royal College of Pathology of Australia, Surry Hills, NSW	MSAC member
Dr Bronwen Harvey Medical Adviser, Population Health Division, Targeted Prevention Programs, Department of Health and Ageing, Canberra, ACT	Department of Health and Ageing, Screening Division
Mr Ian Morris Chief Medical Imaging Technologist Department of Diagnostic Imaging Princess Margaret Hospital for Children and King Edward Memorial Hospital for Women, Subiaco, WA	Australian Institute of Radiography nominee
Clinical A/Prof Jonathan Osborne State Radiologist, BreastScreen Queensland, Brisbane, Qld	Royal Australian and New Zealand College of Radiology nominee
Ms Margaret Tassell Launceston, Tas	Consumer Health Forum nominee

Appendix C Staging system for breast cancer

American Joint Committee on Cancer (AJCC) staging system for breast cancer

Table 1

American Joint Committee on Cancer (AJCC)
TNM Staging System For Breast Cancer

Primary Tumor (T)

Definitions for classifying the primary tumor (T) are the same for clinical and for pathologic classification. If the measurement is made by the physical examination, the examiner will use the major headings (T1, T2, or T3). If other measurements, such as mammographic or pathologic measurements, are used, the subsets of T1 can be used. Tumors should be measured to the nearest 0.1 cm increment.

TX	Primary tumor cannot be assessed
T0	No evidence of primary tumor
Tis	Carcinoma in situ
Tis (DCIS)	Ductal carcinoma in situ
Tis (LCIS)	Lobular carcinoma in situ
Tis (Paget's)	Paget's disease of the nipple with no tumor

Note: Paget's disease associated with a tumor is classified according to the size of the tumor.

T1	Tumor 2 cm or less in greatest dimension
T1mic	Microinvasion 0.1 cm or less in greatest dimension
T1a	Tumor more than 0.1 cm but not more than 0.5 cm in greatest dimension
T1b	Tumor more than 0.5 cm but not more than 1 cm in greatest dimension
T1c	Tumor more than 1 cm but not more than 2 cm in greatest dimension
T2	Tumor more than 2 cm but not more than 5 cm in greatest dimension
T3	Tumor more than 5 cm in greatest dimension
T4	Tumor of any size with direct extension to (a) chest wall or (b) skin, only as described below
T4a	Extension to chest wall, not including pectoralis muscle
T4b	Edema (including peau d'orange) or ulceration of the skin of the breast, or satellite skin nodules confined to the same breast
T4c	Both T4a and T4b
T4d	Inflammatory carcinoma

Regional Lymph Nodes (N)

Clinical

NX	Regional lymph nodes cannot be assessed (e.g., previously removed)
N0	No regional lymph node metastasis
N1	Metastasis to movable ipsilateral axillary lymph node(s)
N2	Metastases in ipsilateral axillary lymph nodes fixed or matted, or in *clinically apparent** ipsilateral internal mammary nodes in the *absence* of clinically evident axillary lymph node metastasis
N2a	Metastases in ipsilateral axillary lymph nodes fixed to one another (matted) or to other structures
N2b	Metastasis only in *clinically apparent** ipsilateral internal mammary nodes and in the *absence* of clinically evident axillary lymph node metastasis
N3	Metastasis in ipsilateral infraclavicular lymph node(s) with or without axillary lymph node involvement, or in *clinically apparent** ipsilateral internal mammary lymph node(s) and in the *presence* of clinically evident axillary lymph node metastasis; or metastasis in ipsilateral supraclavicular lymph node(s) with or without axillary or internal mammary lymph node involvement
N3a	Metastasis in ipsilateral infraclavicular lymph node(s)
N3b	Metastasis in ipsilateral internal mammary lymph node(s) and axillary lymph node(s)
N3c	Metastasis in ipsilateral supraclavicular lymph node(s)

**Clinically apparent* is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination or grossly visible pathologically.

Staging continued on next page (ST-2)

Table 1 (continued)

Pathologic (pN)*

pNX	Regional lymph nodes cannot be assessed (e.g., previously removed, or not removed for pathologic study)
pN0	No regional lymph node metastasis histologically, no additional examination for isolated tumor cells (ITC)

Note: Isolated tumor cells (ITC) are defined as single tumor cells or small cell clusters not greater than 0.2 mm, usually detected only by immonohistochemical (IHC) or molecular methods but which may be verified on H&E stains. ITCs do not usually show evidence of malignant activity e.g., proliferation or stromal reaction.

pN0(i-)	No regional lymph node metastasis histologically, negative IHC
pN0(i+)	No regional lymph node metastasis histologically, positive IHC, no IHC cluster greater than 0.2 mm
pN0(mol-)	No regional lymph node metastasis histologically, negative molecular findings (RT-PCR)†
pN0(mol+)	No regional lymph node metastasis histologically, positive molecular findings (RT-PCR)†

*Classification is based on axillary lymph node dissection with or without sentinel lymph node dissection. Classification based solely on sentinel lymph node dissection without subsequent axillary node dissection is designated (sn) for "sentinel node," e.g., pN0(i+) (sn).

†RT-PCR: reverse transcriptase/polymerase chain reaction.

pN1	Metastasis in 1 to 3 axillary lymph nodes, and/or in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not *clinically apparent***
pN1mi	Micrometastasis (greater than 0.2 mm, none greater than 2.0 mm)
pN1a	Metastasis in 1 to 3 axillary lymph nodes
pN1b	Metastasis in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not *clinically apparent***
pN1c	Metastasis in 1 to 3 axillary lymph nodes and in internal mammary nodes with microscopic disease detected by sentinel lymph node dissection but not *clinically apparent.*** (If associated with greater than 3 positive axillary lymph nodes, the internal mammary nodes are classified as pN3b to reflect increased tumor burden)
pN2	Metastasis in 4 to 9 axillary lymph nodes, or in *clinically apparent** internal mammary lymph nodes in the *absence* of axillary lymph node metastasis
pN2a	Metastasis in 4 to 9 axillary lymph nodes (at least one tumor deposit greater than 2.0 mm)
pN2b	Metastasis in *clinically apparent** internal mammary lymph nodes in the *absence* of axillary lymph node metastasis
pN3	Metastasis in 10 or more axillary lymph nodes, or in infraclavicular lymph nodes, or in *clinically apparent** ipsilateral internal mammary lymph nodes in the *presence* of 1 or more positive axillary lymph nodes; or in more than 3 axillary lymph nodes with clinically negative microscopic metastasis in internal mammary lymph nodes; or in ipsilateral supraclavicular lymph nodes
pN3a	Metastasis in 10 or more axillary lymph nodes (at least one tumor deposit greater than 2.0 mm), or metastasis to the infraclavicular lymph nodes
pN3b	Metastasis in *clinically apparent** ipsilateral internal mammary lymph nodes in the *presence* of 1 or more positive axillary lymph nodes; or in more than 3 axillary lymph nodes and in internal mammary lymph nodes with microscopic disease detected by sentinel lymph node dissection but not *clinically apparent***
pN3c	Metastasis in ipsilateral supraclavicular lymph nodes

* *Clinically apparent* is defined as detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination.

** *Not clinically apparent* is defined as not detected by imaging studies (excluding lymphoscintigraphy) or by clinical examination.

Staging continued on next page (ST-3)

Table 1 (continued)

Distant Metastasis (M)

MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis

STAGE GROUPING

Stage 0	Tis	N0	M0	**Stage IIIB**	T4	N0	M0
Stage I	T1*	N0	M0		T4	N1	M0
Stage IIA	T0	N1	M0		T4	N2	M0
	T1*	N1	M0	**Stage IIIC**	Any T	N3	M0
	T2	N0	M0	**Stage IV**	Any T	Any N	M1
Stage IIB	T2	N1	M0				
	T3	N0	M0				
Stage IIIA	T0	N2	M0				
	T1*	N2	M0				
	T2	N2	M0				
	T3	N1	M0				
	T3	N2	M0				

* T1 includes T1mic

Note: Stage designation may be changed if post-surgical imaging studies reveal the presence of distant metastases, provided that the studies are carried out within 4 months of diagnosis in the absence of disease progression and provided that the patient has not received neoadjuvant therapy.

HISTOPATHOLOGIC TYPE

The histopathologic types are the following:

In situ Carcinomas

NOS (not otherwise specified)
Intraductal
Paget's disease and intraductal

Invasive Carcinomas

NOS
Ductal
Inflammatory
Medullary, NOS
Medullary with lymphoid stroma
Mucinous
Papillary (predominantly micropapillary pattern)
Tubular
Lobular
Paget's disease and infiltrating
Undifferentiated
Squamous cell
Adenoid cystic
Secretory
Cribriform

HISTOPATHOLOGIC GRADE (G)

All invasive breast carcinomas with the exception of medullary carcinoma should be graded. The Nottingham combined histologic grade (Elston-Ellis modification of Scarff-Bloom-Richardson grading system) is recommended.[1,2] The grade for a tumor is determined by assessing morphologic features (tubule formation, nuclear pleomorphism, and mitotic count), assigning a value of 1 (favorable) to 3 (unfavorable) for each feature, and adding together the scores for all three categories. A combined score of 3-5 points is grade 1; a combined score of 6-7 points is grade 2; a combined score of 8-9 points is grade 3.

[1] Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histologic grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19:403-410.

[2] Fitzgibbons PL, Page DL, Weaver D et al. Prognostic factors in breast cancer. College of American Pathologists consensus statement 1999. Arch Pathol Lab Med 2000;124:966-978.

HISTOLOGIC GRADE (NOTTINGHAM COMBINED HISTOLOGIC GRADE IS RECOMMENDED)

GX	Grade cannot be assessed
G1	Low combined histologic grade (favorable)
G2	Intermediate combined histologic grade (moderately favorable)
G3	High combined histologic grade (unfavorable)

Source: (National Comprehensive Cancer Network (NCCN) 2006)

Appendix D Search strategies

Literature sources

Electronic bibliographic databases were searched to find relevant studies (those meeting the inclusion criteria) addressing each of the research questions developed for this MSAC assessment. These databases are described in Table D.1. DM appears in the literature only since 1990, so the search period was restricted from 1990 (or, if inception of the database was later, from that date) until January or February 2007.

Table D.1 Bibliographic databases

Electronic database	Time period
Cochrane Library – including Cochrane Database of Systematic Reviews, Database of Abstracts of Reviews of Effects, Cochrane Central Register of Controlled Trials (CENTRAL), Health Technology Assessment Database, NHS Economic Evaluation Database	1990 – 1/2007 and 2/2007
Current Contents	1993 – 1/2007 and 2/2007
Embase.com (including EMBASE and MEDLINE)	1990 – 1/2007 and 2/2007
PreMEDLINE	1/2007 and 2/2007
ProceedingsFirst	1993 – 1/2007 and 2/2007
Web of Science – Science Citation Index Expanded	1995 – 1/2007 and 2/2007
EconLit	1990 – 1/2007 and 2/2007

Search terms for identifying literature within these bibliographic databases are given below. Table D.2 lists the search terms used to identify studies of DM for screening in an asymptomatic population. Table D.3 lists the search terms used to identify diagnostic studies (ie, asymptomatic, symptomatic and high-risk populations). Table D.4 lists the search terms used to identify any studies reporting change-in-management following DM. Table D.5 lists the search terms used to identify treatment effectiveness studies associated with breast cancer. All searches were complementary but there was some unavoidable overlap. Studies addressing one research question may have appeared in response to a different research question.

Table D.2 Suggested search terms for screening DM

Element of clinical question	Suggested search terms
Population	Female; Human
Intervention/test	Breast-Neoplasms/ra [radiography]; Breast-Neoplasms/di [diagnosis]; Breast-Neoplasms/px [psychology];Breast-Neoplasms/co [complications]; Breast-Neoplasms/ec [economics]; Mass-Screening; Mammography/mt [methods]; Radiographic-Image-Enhancement/mt [methods]; Software; X-Ray-Film AND (Digit* or full?field or direct or indirect or DR or CR) AND (mammogra*); FFDM; (digital or computed) AND radiogra* AND breast AND Screen*
Comparator (if applicable)	n/a
Outcomes (if applicable)	n/a

Table D.3 Search terms for diagnostic DM

Element of clinical question	Suggested search terms
Target population	Female; Human
Intervention/test	Breast-Neoplasms/ra [radiography]; Breast-Neoplasms/di [diagnosis]; Breast-Neoplasms/co [complications]; Breast-Neoplasms/ec [economics]; Breast-Neoplasms/px [psychology]; Mammography/mt [methods]; Radiographic-Image-Enhancement/mt [methods]; Image-Processing,-Computer-Assisted; Sensitivity-and-Specificity; Phantoms,-Imaging; Predictive-Value-of-Tests; Software; X-Ray-Film AND (Digit* or full?field or direct or indirect or DR or CR) AND (mammogra*); FFDM; (digital or computed) AND radiogra* AND breast AND Diagnos*
Comparator (if applicable)	n/a
Outcomes (if applicable)	n/a

Table D.4 Search terms to identify change in management studies

Element of clinical question	Suggested search terms
Target population	Female; Human
Intervention/test	(Digit* or full?field or direct or indirect or DR or CR) AND (mammogra*); FFDM; (digital or computed) AND radiogra* AND breast AND Breast-Neoplasms/ra [radiography]; Breast-Neoplasms/di [diagnosis]; Mammography/mt [methods]; Radiographic-Image-Enhancement/mt [methods]; Image-Processing,-Computer-Assisted; Software; X-Ray-Film
Comparator (if applicable)	n/a
Outcomes	Breast-Neoplasms/rt [radiotherapy]; Breast-Neoplasms/dt [drug therapy]; Breast-Neoplasms/ep [epidemiology]; Breast-Neoplasms/su [surgery]; Breast-Neoplasms/th [therapy]; breast AND cancer AND (treat* or manag* or therap*)

Table D.5 Search terms to identify treatment effectiveness studies

Element of clinical question	Suggested search terms
Target population	Female; Human
Intervention/test	Breast-Neoplasms/rt [radiotherapy]; Breast-Neoplasms/dt [drug therapy]; Breast-Neoplasms/rh [rehabilitation]; Breast-Neoplasms/su [surgery]; Breast-Neoplasms/th [therapy]; breast AND cancer AND (treat* or manag* or therap*)
Comparator (if applicable)	n/a
Outcomes	Breast-Neoplasms/mo [mortality]; Breast-Neoplasms/co [complications]; Breast-Neoplasms/px [psychology]; Therapeutics/ae [adverse effects]; Therapeutics/mo [mortality]; survival; death; morbidity; disease progression; metasta*; adverse events; quality of life

Additional sources of literature – peer-reviewed or grey literature – were sought from the sources listed in Table D.6 and from the HTA agency websites listed in Table D.7.

Table D.6 Additional sources of literature

Source	Location
Internet	
NHMRC – National Health and Medical Research Council (Australia)	http://www.health.gov.au/nhmrc/
US Department of Health and Human Services (reports and publications)	http://www.os.dhhs.gov/
New York Academy of Medicine Grey Literature Report	http://www.nyam.org/library/greylit/index.shtml
TRIP database	http://www.tripdatabase.com
Current Controlled Trials *meta*Register	http://controlled-trials.com/
National Library of Medicine Health Services/Technology Assessment Text	http://text.nlm.nih.gov/
UK National Research Register	http://www.update-software.com/National/
Google Scholar	http://scholar.google.com/
Hand-searching (journals from 2005–2006)	
Breast	Library or electronic access
Breast Cancer Online (BCO)	Library or electronic access
Breast Cancer Research and Treatment	Library or electronic access
Breast Cancer Research	Library or electronic access
Radiology	Library or electronic access
Canadian Association of Radiologists Journal	Library or electronic access
Journal of the National Cancer Institute	Library or electronic access
Clinical Radiology	Library or electronic access
Academic Radiology	Library or electronic access
European Radiology	Library or electronic access
Medical Journal of Australia	Library or electronic access
New England Journal of Medicine	Library or electronic access
Radiation Medicine	Library or electronic access
Specialty websites	
National Breast Cancer Centre (Australia)	http://www.nbccc.org.au
BreastScreen Australia	http://www.cancerscreening.gov.au/internet/screening/publishing.nsf/Content/breastscreen-1lp
Cancer Council Australia	http://www.cancer.org.au/
OncoLink	http://www.oncolink.com/
US National Cancer Institute	http://www.cancer.gov/
Expert clinicians	Library or electronic access
Studies other than those found in regular searches	MSAC advisory panel
Pearling	
Reference lists of all included articles were searched for additional relevant source material	

Table D.7 International Health Technology Assessment Agency websites

AUSTRALIA	
Australian Safety and Efficacy Register of New Interventional Procedures – Surgical (ASERNIP-S)	http://www.surgeons.org/open/asernip-s.htm
Centre for Clinical Effectiveness, Monash University	http://www.med.monash.edu.au/healthservices/cce/evidence/
Centre for Health Economics, Monash University	http://chpe.buseco.monash.edu.au
AUSTRIA	
Institute of Technology Assessment / HTA Unit	http://www.oeaw.ac.at/ita/e1–3.htm
CANADA	
Agence d'Evaluation des Technologies et des Modes d'Intervention en Santé (AETMIS)	http://www.aetmis.gouv.qc.ca/en/
Alberta Heritage Foundation for Medical Research (AHFMR)	http://www.ahfmr.ab.ca/publications.html
Canadian Agency for Drugs and Technologies in Health (CADTH)	http://www.cadth.ca/index.php/en/
Canadian Health Economics Research Association (CHERA/ACRES) – Cabot database	http://www.mycabot.ca
Centre for Health Economics and Policy Analysis (CHEPA), McMaster University	http://www.chepa.org
Centre for Health Services and Policy Research (CHSPR), University of British Columbia	http://www.chspr.ubc.ca
Health Utilities Index (HUI)	http://www.fhs.mcmaster.ca/hug/index.htm
Institute for Clinical and Evaluative Studies (ICES)	http://www.ices.on.ca
Saskatchewan Health Quality Council (Canada)	http://www.hqc.sk.ca
DENMARK	
Danish Centre for Evaluation and Health Technology Assessment (DACEHTA)	www.sst.dk/Planlaegning_og_behandling/Medicinsk_teknologivurdering.aspx?lang = en
Danish Institute for Health Technology Assessment (DIHTA)	http://www.dihta.dk/publikationer/index_uk.asp
Danish Institute for Health Services Research (DSI)	http://www.dsi.dk/engelsk.html
FINLAND	
Finnish Office for Health Technology Assessment (FINOHTA)	http://www.stakes.fi/finohta/e/
FRANCE	
L'Agence Nationale d'Accréditation et d'Evaluation en Santé (ANAES)	http://www.anaes.fr/
GERMANY	
German Institute for Medical Documentation and Information (DIMDI) / HTA	http://www.dimdi.de/en/hta/index.html
THE NETHERLANDS	
Health Council of the Netherlands – Gezondheidsraad	http://www.gr.nl/adviezen.php
Institute for Medical Technology Assessment (Netherlands)	http://www.imta.nl/
NEW ZEALAND	
New Zealand Health Technology Assessment (NZHTA)	http://nzhta.chmeds.ac.nz/
NORWAY	
Norwegian Centre for Health Technology Assessment (SMM)	http://www.oslo.sintef.no/smm/Publications/Engsmdrag/FramesetPublications.htm

Table D.7. (cont'd)

SPAIN	
Agencía de Evaluación de Tecnológias Sanitárias, Instituto de Salud 'Carlos III' I / Health Technology Assessment Agency (AETS)	http://www.isciii.es/aets/
Andalusian Agency for Health Technology Assessment (Spain)	http://www.juntadeandalucia.es/salud/orgdep/AETSA/default.asp?V = EN
Catalan Agency for Health Technology Assessment (CAHTA)	http://www.aatm.es/cgi-bin/frame.pl/ang/pu.html
SWEDEN	
Center for Medical Health Technology Assessment	http://www.cmt.liu.se/English/Engstartsida.html
Swedish Council on Technology Assessment in Health Care (SBU)	http://www.sbu.se/admin/index.asp
SWITZERLAND	
Swiss Network on Health Technology Assessment (SNHTA)	http://www.snhta.ch/
UNITED KINGDOM	
Health Technology Board for Scotland	http://www.htbs.org.uk/
National Health Service Health Technology Assessment (UK) / National Coordinating Centre for Health Technology Assessment (NCCHTA)	http://www.hta.nhsweb.nhs.uk/
NHS Quality Improvement Scotland	http://www.nhshealthquality.org/
National Institute for Clinical Excellence (NICE)	http://www.nice.org.uk/index.htm
European Information Network on New and Changing Health Technologies	http://www.euroscan.bham.ac.uk/
University of York NHS Centre for Reviews and Dissemination (NHS CRD)	http://www.york.ac.uk/inst/crd/
UNITED STATES	
Agency for Healthcare Research and Quality (AHRQ)	http://www.ahrq.gov/clinic/techix.htm
Harvard School of Public Health – Cost-Utility Analysis Registry	http://www.tufts-nemc.org/cearegistry/index.html
Institute for Clinical Systems Improvement (ICSI)	http://www.icsi.org
Minnesota Department of Health	http://www.health.state.mn.us/htac/index.htm
National Information Centre of Health Services Research and Health Care Technology	http://www.nlm.nih.gov/hsrph.html
Oregon Health Resources Commission	http://egov.oregon.gov/DAS/OHPPR/HRC/about_us.shtml
Office of Health Technology Assessment Archive	http://www.wws.princeton.edu/~ota
US Blue Cross / Blue Shield Association Technology Evaluation Center (TEC)	http://www.bcbs.com/consumertec/index.html
Veterans Affairs Research and Development Technology Assessment Program	http://research.va.gov/

Additional study selection criteria

Limited direct evidence for the screening of asymptomatic women was identified in one study. No direct evidence was identified in the literature searches to inform an assessment of DM for surveillance in asymptomatic women at potentially high risk, or in the diagnosis of breast cancer in symptomatic women. The criteria that were used to identify potentially relevant studies for these latter assessments are given in Tables D.8 and D.9.

Table D.8 Inclusion criteria for identification of direct evidence studies concerning the effectiveness and cost-effectiveness of DM in the diagnosis of breast cancer in symptomatic women

Characteristic	Criteria
Publication type	Effectiveness: Randomised or non-randomised controlled trials or cohort studies or systematic reviews of these study designs. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded Cost-effectiveness: Economic studies, modelling, economic analyses
Patient	Women with symptoms or a clinical abnormality of the breast suggesting malignancy Subgroups: (1) <50 y v. 50+ y [a] (2) very to extremely dense breast tissue v. less dense tissue (3) pre- to perimenopausal v. postmenopausal [a]
Intervention/test	Diagnostic DM, consisting of additional full-field x-ray views of affected breast and imaging of contralateral breast ± breast ultrasound scan ± breast MRI
Comparator	Diagnostic FM, consisting of additional x-ray views of affected breast and imaging of contralateral breast ± breast ultrasound scan ± breast MRI
Outcome	Effectiveness: Primary – *Patient-relevant:* survival, reduction in mortality (all-cause and breast cancer), quality of life; *Surrogate:* cancer (invasive or ductal carcinoma *in situ*) detection rates, tumour size or stage, recurrence rates, disease progression Secondary – biopsy rates, false positive rates, false negative rates, image quality or interpretation, duration of procedure, delayed intervention (due to false negative) Cost-effectiveness: Cost, incremental cost-effectiveness ratio (eg, cost/LYG, QALY, DALY), workforce issues
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

MRI = magnetic resonance imaging; LYG = life years gained; QALY = quality-adjusted life year; DALY = disability-adjusted life year. a: These subgroups are surrogate measures for breast tissue density, as density can be measured only via mammogram.

Table D.9 Inclusion criteria for identification of direct evidence studies concerning the effectiveness and cost-effectiveness of DM for surveillance of breast cancer in women at potentially high risk

Characteristic	Criteria
Publication type	Effectiveness: Randomised or non-randomised controlled trials or cohort studies or systematic reviews of these study designs. Non-systematic reviews, letters, editorials, and animal, *in vitro* and laboratory studies were excluded Cost-effectiveness: Economic studies, modelling, economic analyses
Patient	Women at potentially high risk of breast cancer due to: 1. Potentially high risk of ovarian cancer 2. two 1° or 2° relatives on one side of the family diagnosed with breast or ovarian cancer plus one or more of the following features on the same side of the family: - additional relative(s) with breast or ovarian cancer - breast cancer diagnosed before the age of 40 - bilateral breast cancer - breast and ovarian cancer in the same woman - Ashkenazi Jewish ancestry - breast cancer in a male relative 3. one 1° or 2° relative diagnosed with breast cancer at age 45 or younger plus another 1° or 2° relative on the same side of the family with sarcoma (bone or soft tissue) at age 45 or younger 4. being a member of a family in which the presence of a high-risk breast cancer gene mutation has been established, eg, *BRCA1*, *BRCA2*, *Tp53* 5. personal history of breast cancer 6. premalignant conditions, ie, lobal carcinoma *in situ* and atypical ductal hyperplasia Subgroups: (1) <50 y v. 50+ y [a] (2) very to extremely dense breast tissue v. less dense tissue (3) pre- to perimenopausal v. postmenopausal [a]
Intervention/test	Diagnostic DM, consisting of additional full-field x-ray views of both breasts ± breast ultrasound scan ± breast MRI
Comparator	Diagnostic FM, consisting of additional x-ray views of both breasts ± breast ultrasound scan ± breast MRI
Outcome	Effectiveness: Primary – *Patient-relevant:* survival, reduction in mortality (all-cause and breast cancer), quality of life; *Surrogate:* cancer (invasive or ductal carcinoma *in situ*) detection rates, tumour size or stage, recurrence rates, disease progression Secondary – biopsy rates, false positive rates, false negative rates, image quality or interpretation, duration of procedure, delayed intervention (due to false negative) Cost-effectiveness: Cost, incremental cost-effectiveness ratio (eg, cost/LYG, QALY, DALY), workforce issues
Language	Non-English-language articles were excluded unless they appeared to provide a higher level of evidence than the English-language articles identified

MRI = magnetic resonance imaging; LYG = life years gained; QALY = quality-adjusted life year; DALY = disability-adjusted life year. a: These subgroups are surrogate measures for breast tissue density, as density can be measured only via mammogram.

Appendix E Studies included in the review

Safety studies

Table 45 Safety study profiles

Study location	Study design quality	Study population	Intervention	Outcomes	Follow-up
(Fischmann et al 2005) Tübingen, Germany	Randomised controlled trial Level II interventional evidence 1/3 (NHMRC Quality Assessment Scale) Not double blinded Allocation not concealed 199/200 women included	*N* = 199 females DM = 199 FM = 199 Asymptomatic Age ≥40 years Breast density: not stated Pre/peri v. post menopausal: not stated	*Index test* DM (soft copy) *Product* – Senographe 2000D (GE Medical Systems, Milwaukee, WI, USA) *Comparator* FM (hard copy) *Product* – Senographe DMR+ (GE Medical Systems) *Views* – 2 views (CC, MLO); one method per breast *Reading* – 3 readers; independent reading; not blinded	Physical harms from mammography (radiation exposure)	n/a
(Gennaro & di Maggio 2006) Padua, Italy	Cohort study Level III-2 interventional evidence 2/6 (NHMRC Quality Assessment Scale)	*N* = 596 mammograms DM = 296 FM = 300 Population not described Age: not stated Mean breast density: FM 45.3 mm DM 48.5 mm	*Index test* DM *Product* – Senographe 2000D (GE Medical Systems) *Comparator* FM Product – Senographe DMR+ (GE Medical Systems) *Views* – 1 view (CC)	Physical harms from mammography (radiation exposure)	n/a

Study location	Study design quality	Study population	Intervention	Outcomes	Follow-up
(Gennaro et al 2004) Padua, Italy	Case series Level IV interventional evidence 3/6 (NHS CRD Quality Assessment Scale)	*N* = ~800 DM images (200 images for each of four sites) Population n/a Age: n/a Breast density: Site 1 = 50.69 ± 11.73 mm Site 2 = 48.61 ± 12.17 mm Site 3 = 46.13 ± 13.52 mm Site 4 = 46.48 ± 11.78 mm	*Index test* DM *Product* – Senographe 2000D (GE Medical Systems) *Comparator* None *Views* – 1 view (CC)	Physical harms from mammography (radiation exposure)	Study period 01–05/2002
(Hemdal et al 2005a) Malmö, Sweden	Case series Level IV interventional evidence 4/6 (NHS CRD Quality Assessment Scale)	*N* = 28 females DM = 56 images Asymptomatic population Age >50 years Mean compressed breast thickness 1st breast = 50 (21–71) mm 2nd breast = 49 (21–70) mm	*Index test* DM *Product* – Senographe 2000D (GE Medical Systems) *Comparator* None *Views* – Single view (MLO) First breast AOP/STD mode; second breast manual mode using same settings but only about half tube loading *Readings* – 3 reader/radiologists	Physical harms from mammography (radiation exposure)	n/a
(Hermann et al 2002) Göttingen, Germany	Case series Level IV interventional evidence 3/6 (NHS CRD Quality Assessment Scale)	*N* = 591 females DM = 1116 images Mixed population Age <50 years Mean compressed breast thickness 55.7 ± 13.21 (17–94) mm	*Index test* DM *Product* – Senographe 2000D (GE Medical Systems) *Comparator* None (literature-based comparison with FM) *Views* – 1 view (CC); AOP/STD mode	Physical harms from mammography (radiation exposure)	Study period 08–11/2000

Study location	Study design quality	Study population	Intervention	Outcomes	Follow-up
(Lewin et al 2001) (Lewin et al 2002) Denver, Colorado Colorado screening study	Prospective cohort, cross-classified CX P1 Level III-2 diagnostic evidence Q2 (QUADAS = 10/14)	*N* = 4489 (1665 women enrolled 2×, 291 women enrolled 3×); 6736 examinations DM = 6736 examinations FM = 6736 examinations Asymptomatic Age = mean 55.6 years (Colorado I and II) Breast density: 44.9% examinations with heterogeneous or extremely dense breasts with DM; 44.5% with FM (*n* = 4883, Colorado I) Pre/peri v. postmenopausal: not stated	*Index test* DM (soft-copy) *Type*-FFDM *Product* – Prototype indirect flat-panel detector (amorphous silicon) (GE Medical Systems). No automated processing algorithms *Comparator* FM (hard-copy) *Product* – Colorado I: Kodak Min-R 2000, Eastman Kodak, Rochester, NY, USA. Automated Optimisation Parameters feature; Colorado II: DMR, GE Medical Systems *Screening interval* – Annual. Follow-up of cohort was maintained to 12 months after study entry *Views* – 2 views (CC, MLO) of each breast with each method Single reading of DM or FM for each woman by independent radiologists. Therefore double reading per woman. Prior films available. • 2 scales: *probability* of malignancy (0%–100%); BIRADS [a] (0, 2, 3, 4, 5) • Work-up on basis of both DM and FM radiologist indicating abnormal mammogram. Discordant findings were assessed at recall meeting *Reference standard* Histopathology	Physical harms from false positives (unnecessary further intervention)	1 year
(Moran et al 2005) Madrid, Spain	Cohort study Level III-2 interventional evidence 2/6 (NHMRC Quality Assessment Scale) No description of method of subject selection for DM or FM No description of women who underwent FM	*N* = 27545 images DM = 20137 images (5034 females) Symptomatic Mean age = 56 ± 11 years Mean breast thickness = 52 ± 13 mm FM = 7401 images Screening population	*Index test* DM *Product* – Senographe 200D (GE Medical Systems) *Comparator* FM (hard copy) *Product* – Slow + fast screen type *Views* – 2 views (MLO, CC) of each breast	Physical harms from mammography (radiation exposure)	Study period: 3/2001 – 10/2003

Study location	Study design quality	Study population	Intervention	Outcomes	Follow-up
(Skaane et al 2005; Skaane et al 2003) Oslo, Norway Oslo I study	Prospective cohort, cross-classified CX, P1 Level III-2 diagnostic evidence Q1 (QUADAS = 12/14)	*N* = 3683 DM = 3683 FM = 3683 Asymptomatic Age = 58.2 y [50–68] 50–69 – 100% Breast density – not stated Pre/peri v. postmenopausal – not stated	*Index test* DM (soft copy) *Type* – FFDM *Product* – Senographe 2000D (GE Medical Systems), AOP mode *Comparator* FM (hard copy) *Product* – Mammomat 300 (Siemens, Erlangen, Germany), Mo:Mo, 29 kV *Screening interval* – 2 years *Views* – 2 views (CC, MLO) of each breast with each method *Reading* – Batch double reading each of DM and FM for each woman by independent radiologists • ≥1 positive [a] reading in each double reading = positive or assess for recall • Recall for DM and FM assessed independently. Recall on basis of ≥3 score by ≥1 reader; or consensus of 4 radiologists assessing DM or FM with 2 score. Previous films available *Reference standard* Histopathology + 2-year follow-up of interval cancers through linked registry data	Physical harms from false positive (unnecessary further intervention)	2 years
(Yamada et al 2004) Sendai, Japan	Cross-sectional study Level III-2 diagnostic evidence (false alarms) Q3 (QUADAS = 7/14) Level IV interventional evidence (radiation dose) 5/6 (NHS CRD Quality Assessment Scale)	*N* = 480 females DM = 480 FM = 480 Asymptomatic population supplemented with 2 diagnostic cases Age ≥50 years Breast density	*Index test* DM (hard copy) *Product* – Senographe 2000D (GE Medical Systems) *Comparator* FM (hard copy) Product – Mammomat 3000 Nova (Siemens Medical Systems, Solna, Sweden) Views – Single view (MLO) of each breast with each method *Reading* – 2 readers – Double radiologist; independent blinded review *Reference standard* Not stated	Physical harms from DM only (radiation exposure) Physical harms from false alarm (unnecessary further intervention)	Study period 07–11/2001

DM = digital mammography; FM = film-screen mammography; CR = computed radiography; FFDM = full-field digital mammography; CAD = computer-assisted diagnosis; n/a = not applicable; y = years; MLO = mediolateral oblique; CC = craniocaudal; AOP = automated optimisation of parameters; STD = standard; BIRADS = Breast Imaging Reporting and Data System.

Screening studies

Table 46 Diagnostic accuracy study profiles – screening of asymptomatic women

Study location	Study design quality	Study population	Inclusion/exclusion criteria	Diagnostic tests	Outcomes
(Lewin et al 2002; Lewin et al 2001) Denver, Colorado, USA Colorado screening study	Prospective cohort, cross-classified CX P1 Level III-2 diagnostic evidence Q2 (QUADAS = 10/14)	*N* = 4489 (1665 women enrolled 2×, 291 women enrolled 3×); 6736 examinations DM = 6736 examinations FM = 6736 examinations Asymptomatic Age = mean 55.6 years (Colorado I and II) Breast density: 44.9% examinations with heterogeneous or extremely dense breasts with DM; 44.5% with FM (*n* = 4883, Colorado I) Pre/peri v. postmenopausal: not stated	*Inclusion* All asymptomatic women presenting for screening mammography at either of two medical centres, aged at least 40 years old, who agreed to enrol *Exclusion* Women with breast implants; breasts that would not fit on a large film-screen image receptor (24 cm × 30 cm)	*Index test* DM (soft-copy) *Type* – FFDM *Product* – Prototype indirect flat-panel detector (amorphous silicon), GE Medical Systems. No automated processing algorithms. *Comparator* FM (hard-copy) *Product* – Colorado I:. Kodak Min-R 2000 (Eastman Kodak). Automated Optimisation Parameters feature; Colorado II: DMR, GE Medical Systems *Screening interval* – Annual. Follow-up of cohort was maintained to 12 months after study entry. *Views* – 2 views (CC, MLO) of each breast with each method Single reading of DM or FM for each woman by independent radiologists. Therefore double reading per woman. Prior films available. • 2 scales: probability of malignancy (0%–100%); BIRADS [a] (0, 2, 3, 4, 5) • Work-up on basis of both DM and FM radiologists indicating abnormal mammogram. Discordant findings assessed at recall meeting *Reference standard* Histopathology	*Diagnostic accuracy* Sensitivity PPV False negative AUC *Screening effectiveness* Recall rate Cancer detection rate Biopsy rate

Study location	Study design quality	Study population	Inclusion/exclusion criteria	Diagnostic tests	Outcomes
(Pisano et al 2005a; Pisano et al 2005b) Multicentre (33 sites) United States and Canada American College of Radiology Imaging Network DMIST	Prospective cohort, cross-classified CX P1 Level III-2 diagnostic evidence Q1 (QUADAS = 12/14) Randomised order of testing for index test and comparator	*N* = 49 528 (ITS) 13.3% lost to follow-up *N* = 42 760 (PP of verified breast cancer status) DM = 42 572 FM = 42 745 Asymptomatic Age at enrolment (*N* = 49 333) = 54.6 [IQR 47–61] age <50 y (PP): 33.5% Breast density (PP): 46.5% with heterogeneous or extremely dense breasts Menopausal status (PP): 37.0% pre or perimenopausal	*Inclusion* Women presenting for screening mammography at study sites *Exclusion* Women reporting symptoms, had breast implants, believed to be pregnant, received mammography in previous 11 months, history of breast cancer treated with both lumpectomy and radiation	*Index test* DM (hard or soft copy, depending on equipment used and site) *Type*-5 types including CR and FFDM *Products* – SenoScan (Fischer Medical, Denver, CO, USA), Computed Radiography System for Mammography (Fuji Medical), Senographe 2000D (GE Medical Systems), Digital Mammography System (Hologic, Danbury, CT, USA), Selenia FFDM System (Hologic) *Comparator* FM (hard-copy) *Products* – Bennett (Hologic), Instrumentarium (GE Medical Systems), Lorad (Hologic), Siemens (Berlin, Germany), Sophie (Planmed, Helsinki, Finland) *Screening interval* – Annual. Follow-up of cohort was maintained ~11–15 months after study entry. *Views* – 2 views (CC, MLO) of each breast with each method (or 1 breast if prior mastectomy) *Reading*– Single reading of DM or FM for each woman by independent radiologists. Therefore double reading per woman. Prior films or soft copy available. • 4 scales: probability of malignancy (0%–100%); BIRADS (0–5), ordinal malignancy scale (1–7); call-back scale (1–6). See paper for definitions • Work-up on basis of either DM or FM radiologist indicating abnormal mammogram *Reference standard* Histopathology	*Diagnostic accuracy* Sensitivity Specificity PPV AUC *Screening effectiveness* Cancer detection rate Biopsy rate

Study location	Study design quality	Study population	Inclusion/exclusion criteria	Diagnostic tests	Outcomes
(Skaane et al 2005; Skaane et al 2003) Oslo, Norway Oslo I study	Prospective cohort, cross-classified CX, P1 Level III-2 diagnostic evidence Q1 (QUADAS = 12/14)	*N* = 3683 DM = 3683 FM = 3683 Asymptomatic Age = 58.2 y [50–68] 50–69 y: 100% Breast density: not stated Pre/peri v. post menopausal: not stated	*Inclusion* Invitation and attendance at population-based breast cancer screening program *Exclusion* Women aged 45–49 years	*Index test* DM (soft copy) *Type* – FFDM *Product* – Senographe 2000D (GE Medical Systems), AOP mode *Comparator* FM (hard copy) *Product* – Mammomat 300 (Siemens), Mo:Mo, 29 kV *Screening interval* – 2 years *Views* – 2 views (CC, MLO) of each breast with each method *Reading* – Batch double reading each of DM and FM for each woman by independent radiologists. Therefore 4 readings per woman • ≥1 positive [b] reading in each double reading = positive or assess for recall • Recall for DM and FM assessed independently. Recall on basis of ≥3 score [b] by ≥1 reader; or consensus of 4 radiologists assessing DM or FM with 2 score [b]. Previous films available *Reference standard* Histopathology	*Diagnostic accuracy v. reference standard* PPV(1) – based on initial positive result PPV(2) – based on recalls *Full data set* (based on initial results + interval cancers / follow-up) Sensitivity Specificity PPV(3) NPV False negative False positive *Screening effectiveness* Recall rate Cancer detection rate (1) – initial positive result Cancer detection rate (2) – recalls Cancer detection rate (3) – full data set Interpretation time

Study location	Study design quality	Study population	Inclusion/exclusion criteria	Diagnostic tests	Outcomes
(Skaane & Skjennald 2004) Oslo, Norway Oslo II Study	Randomised controlled trial C1, P1 Level II screening evidence Q3 (NHMRC = not double blinded, no allocation concealment, no ITS analysis – *this was done in the evaluation*)	*N* = 25 263 DM = 7209 (ITS) DM = 6998 (PP) Age: 50–69 y (mean 59.1 y); 45–49 y (mean 47.4 y) FM = 18 054 (ITS) FM = 17 913 (PP) Age: 50–69 y (mean 59.2 y); 45–49 y (mean 47.4 y) Asymptomatic Breast density: not stated Pre/peri v. postmenopausal: not stated	<u>*Inclusion*</u> Invitation and attendance at population-based breast cancer screening program <u>*Exclusion*</u> Post-randomisation exclusions (excluded from PP analysis) of women who underwent mammography with a method other than that assigned (*n* = 352, FM: 141, DM: 211)	<u>*Index test*</u> DM (soft copy) *Type* – FFDM *Product* – Senographe 2000D (GE Medical Systems), AOP mode <u>*Comparator*</u> FM (hard copy) *Product* – Mammomat 300 (Siemens), Mo:Mo, 29 kV *Screening interval* – 2 years for women outside Oslo; all women aged 45–49 (only screened in Oslo) had 1-year screening interval; women aged 50–69 living in Oslo had 1-year screening interval *Views* – 2 views (CC, MLO) of each breast *Reading* – Batch double reading each of DM or FM by independent radiologists. Therefore double reading per woman. In first 4 months, 4 readings of DM undertaken to ensure cancer detection rate no lower than FM (stopping rule). Only first 2 readings for each DM case used in the analysis during this period • ≥1 positive [b] reading in each double reading = positive or assess for recall • Recall on basis of ≥3 score [b] by ≥1 reader; or consensus of ~2 radiologists assessing DM or FM with 2 score [b]. Previous films available <u>*Reference standard*</u> Histopathology	<u>*Diagnostic accuracy*</u> PPV NPV <u>*Screening effectiveness*</u> Recall rate Cancer detection rate Tumour size

Age is presented as mean ± standard deviation where possible. PPV = positive predictive value; NPV = negative predictive value; DM = digital mammography; FM = film-screen mammography; AUC = area under the curve; CR = computed radiography; FFDM = full-field digital mammography; CC = craniocaudal; MLO = mediolateral oblique; Mo = molybdenum filter. a: BIRADS = Breast Imaging Reporting and Data System – 0 = need for additional imaging evaluation, 1 = negative finding, 2 = benign finding, 3 = probably benign finding, 4 = suspicious abnormality, 5 = highly suggestive of malignancy. Positive finding given as BIRADS score 0, 2, 3, 4 or 5. b: Positive = score of 2–5 on 5-point scale, 1 = normal or definitely benign, 2 = probably benign, 3 = indeterminate finding, 4 = probably malignant, 5 = malignant; ITS = intention-to-screen; PP = per-protocol analysis by screening method actually received; IQR = interquartile range.

Diagnostic and surveillance studies

Table 47 Diagnostic accuracy study profiles: women at potentially high risk and symptomatic women

Study location	Study design quality	Study population	Inclusion/exclusion criteria	Diagnostic tests	Outcomes
(Seo et al 2006) North Carolina, USA	Cross-sectional CX, P1 Level III-2 diagnostic evidence Q1 (QUADAS = 12/14)	*N* = 11 621 DM = Not stated FM = Not stated Ref. standard = 1121 (incl. 2 males) Symptomatic women, women at high risk and women with abnormal screening mammograms Age: mean 56.5 y (range 28–85 y) Breast density: high v. low	*Inclusion* Undergoing diagnostic mammography *Exclusion* Not stated	*Index test* Diagnostic DM FFDM Senographe 2000D (GE Medical Systems) *Comparator* Diagnostic FM Views: not stated Single radiologist *Reference standard* Histopathology	*Diagnostic accuracy* PPV
(Venta et al 2001) Chicago, IL, USA	Cross-sectional CX, P1 Level III-2 diagnostic evidence Q1 (QUADAS = 12/14)	*N* = 692 DM = 692 FM = 692 Ref. standard = 50 Symptomatic women, women at high risk, women with abnormal screening mammograms, women with breast augmentation and women referred for diagnostic mammography due to anxiety or proximity to clinic Age: mean 53.0 y (range 40–88 y)	*Inclusion* Undergoing diagnostic mammography *Exclusion* Less than 40 y, known or suspected pregnancy, immobile, unable to have breasts completely imaged on an 18 × 24 cm image receptor, unable to provide informed consent	*Index test* Diagnostic DM FFDM Prototype Senographe 2000D *Comparator* Diagnostic FM ± breast ultrasound Views: 3 for DM, 3 or more for FM (with views chosen to best evaluate each patient) Single radiologist *Reference standard* Histopathology	*Diagnostic accuracy* Sensitivity Specificity PPV NPV

Study location	Study design quality	Study population	Inclusion/exclusion criteria	Diagnostic tests	Outcomes
(Cole et al 2004) Multicentre study, North America	Cross-sectional CX, P1 Level III-2 diagnostic evidence Q1 (QUADAS = 12/14)	*N* = 676 DM = 247 FM = 247 Ref. standard not stated Symptomatic women, women with abnormal screening mammograms and women recommended or scheduled to undergo histopathology Age: not stated	*Inclusion* Undergoing diagnostic mammography or scheduled to undergo histopathology *Exclusion* Less than 21 y of age, known or suspected pregnancy, unable to provide informed consent	*Index test* Diagnostic DM FFDM Fischer SenoScan (Fischer Imaging) *Comparator* Diagnostic FM Views: 2 (CC and MLO) Single radiologist *Reference standard* Histopathology	*Diagnostic accuracy* Sensitivity Specificity AUC
(Hendrick et al 2001) Multicentre study, USA	Cross-sectional CX, P1 Level III-2 diagnostic evidence Q3 (QUADAS = 7/14)	*N* = 625 DM = 625 FM = 625 Ref. standard not stated Symptomatic women, women with abnormal screening mammograms and 20 consecutive cancer cases generated from a screening population Age: median 55.0 y (range 40–86 y)	*Inclusion* Undergoing diagnostic mammography *Exclusion* Less than 40 y of age, known or suspected pregnancy, breast implants, non-focal or bilateral breast pain, unable to have breasts completely imaged on a 24 × 30 cm image receptor, unable to provide informed consent	*Index test* Diagnostic DM FFDM Senographe 2000D (GE Medical Systems, Milwaukee, WI, USA) *Comparator* Diagnostic FM Views: 2 (type not stated) Single radiologist *Reference standard* Not stated	*Diagnostic accuracy* Sensitivity Specificity NPV PPV AUC

PPV = positive predictive value; NPV = negative predictive value; DM = digital mammography; FM = film-screen mammography; AUC = area under the curve; FFDM = full-field digital mammography; CC = craniocaudal; MLO = mediolateral oblique.

Appendix F Excluded studies

Screening studies

Data not presented numerically

Hildell, J., Hofer, B. & Zynamon, A. (1992). Storage phosphor digital mammography vs. screen-film mammography. Preliminary results of comparison in a phantom model and initial clinical experience. *Radiol Diagnost,* 33, 312–319.

Systematic review no longer relevant owing to availability of recent information

Elmore, J. G., Armstrong, K. et al. (2005). Screening for breast cancer. *J Am Med Assoc,* 293, 1245–1256.

Irwig, L., Houssami, N. & van Vliet, C. (2004). New technologies in screening for breast cancer: a systematic review of their accuracy. *Br J Cancer,* 90, 2118–2122.

Diagnostic case-control study design

Skaane, P., Balleyguier, C. et al. (2005). Breast lesion detection and classification: comparison of screen-film mammography and full-held digital mammography with soft-copy reading – observer performance study. *Radiology*, 237, 37–44.

Diagnosis and surveillance studies

Systematic review no longer relevant owing to availability of recent information

Blue Cross Blue Shield Association (2006). *Computer Aided Detection with Full-Field Digital Mammography.* Blue Cross Blue Shield Technology Evaluation Center. Available from: http://www.bcbs.com/blueresources/tec/vols/21/21_03.html [Accessed July 2007]

References

AIHW (1997). *National Health Priority Areas Report: Cancer control* [internet]. Australian Institute of Health and Welfare. Available from: http://www.health.gov.au/internet/wcms/publishing.nsf/Content/79565B6D9AFE0544CA257082001DEB53/$File/nhpaall.pdf [Accessed September, 2007].

AIHW (2006). *BreastScreen Australia Monitoring Report 2002–2003,* Australian Institute of Health and Welfare, Canberra.

AIHW (2007a). *Breast Screen Australia Monitoring Report 2003–2004,* Australian Institute of Health and Welfare and the Australian Government Department of Health and Ageing for the Breast Screen Australia Program, Canberra.

AIHW (2007b). *Cancer in Australia: an overview, 2006,* AIHW (Australian Institute of Health and Welfare) & AACR (Australasian Association of Cancer Registries), Canberra.

AIHW & NBCC (2006). *Breast Cancer in Australia: an overview, 2006,* Australian Institute of Health and Welfare and National Breast Cancer Centre, Canberra.

American College of Radiology (2003). *Breast Imaging Reporting and Data System (BI-RADS) Mammography*, 4th ed. American College of Radiology, Reston, VA, USA.

Armitage, P., Berry, G. & Matthews, J. N. S. (2002). *Statistical Methods in Medical Research,* 4th ed. Blackwell Science, Oxford.

Australian Bureau of Statistics (2007). 6401.0 – Consumer Price Index, Australia, March 2007. Available from: http://www.abs.gov.au/ausstats/abs@.nsf/mf/6401.0/. [Accessed 5 July 2007].

Berg, W. A., D'Orsi, C. J. et al. (2002). Does training in the breast imaging reporting and data system (BI-RADS) improve biopsy recommendations or feature analysis agreement with experienced breast imagers at mammography? *Radiology,* 224(3), 871–880.

Berns, E. A., Hendrick, R. E. et al. (2006). Digital and screen-film mammography: comparison of image acquisition and interpretation times, *AJR Am J Roentgenol,* 187(1), 38–41.

Blue Cross Blue Shield Association (2006). *Computer Aided Detection with Full-Field Digital Mammography.* Blue Cross Blue Shield Technology Evaluation Center. Available from: http://www.bcbs.com/blueresources/tec/vols/21/21_03.html [Accessed July 2007]

Bradley, G. & Mozjerin, C. (2000). *Occupancy chargeback: an overview of different models and a glimpse of "one" future.* ATEM-AAPPA Conference, Melbourne, September 25, 2000. Available from: http://mams.rmit.edu.au/n6kmum2aewlb.doc [Accessed 5 July 2007].

BreastScreen South Australia (2005). *BreastScreen SA: 2001 and 2002 Statistical Report.* Adelaide.

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

Ciatto, S., Brancato, B. et al. (2006). A methodology to evaluate differential costs of full field digital as compared to conventional screen film mammography in a clinical setting. *Eur J Radiol,* 57(1), 69–75.

Cole, E., Pisano, E. D. et al. (2004). Diagnostic accuracy of Fischer Senoscan Digital Mammography versus screen-film mammography in a diagnostic mammography population. *Acad Radiol,* 11(8), 879–886.

Comité d'Evaluation et de Diffusion des Innovations. 2002. *Digital mammography* . Available from: http://cedit.ap-hop-paris.fr/servlet/siteCeditGB?Destination=reco&numArticle=99.01. [Accessed July 2007]

Crawford, M. D., Biankin, A. V. et al. (2000). The operative management of screen-detected breast cancers. *Aust NZ J Surg,* 70(3), 168–173.

Deeks, J. J. (2001). Systematic reviews of evaluations of diagnostic and screening tests. *Br Med J,* 323 (21 July), 157–162.

Dershaw, D. (2006). Status of mammography after the Digital Mammography Imaging Screen Trial: digital versus film. *Breast J,* 12(2), 99–102.

Early Breast Cancer Trialists' Collaborative Group (1996). Ovarian ablation in early breast cancer: overview of the randomised trials. *Lancet,* 348 (9036), 1189–1197.

Early Breast Cancer Trialists' Collaborative Group (2001a). Multi-agent chemotherapy for early breast cancer. *Cochrane Database of Systematic Reviews 2001, Issue 4,* Art. No. CD000487. DOI: 10.1002/14651858.CD000487.

Early Breast Cancer Trialists' Collaborative Group (2001b). Tamoxifen for early breast cancer. *Cochrane Database of Systematic Reviews 2001, Issue 1.,* Art. No. CD000486. DOI: 10.1002/14651858.CD000486.

Early Breast Cancer Trialists' Collaborative Group (2002). Radiotherapy for early breast cancer. *Cochrane Database of Systematic Reviews 2002, Issue 2,* Art. No. CD003647. DOI: 10.1002/14651858.CD003647.

Egger, M., Juni, P. et al. (2003). How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess,* 7(1), 76.

Elmore, J. G., Nakano, C. Y. et al. (2003). International variation in screening mammography interpretations in community-based programs. *J Natl Cancer Inst,* 95(18), 1384–1393.

Evans, S. & Lozano, G. (1997). The Li-Fraumeni syndrome: an inherited susceptibility to cancer. *Mol Med Today,* 3(9), 390–395.

Faulkner, K. (2005). Mammographic screening: is the benefit worth the risk? *Radiat Prot Dosim,* 117(1–3), 318–320.

Feig, S. A. & Yaffe, M. J. (1998). Digital mammography. *Radiographics,* 18(4), 893–901.

Fischmann, A., Siegmann, K. C. et al. (2005). Comparison of full-field digital mammography and film-screen mammography: image quality and lesion detection. *Br J Radiol,* 78(928), 312–315.

Forrest, A. & Anderson, E. (1999). Breast cancer screening and management. *Med J Aust,* 171(9), 479–484.

Frank, E., Rimer, B. K. et al. (2000). US women physicians' personal and clinical breast cancer screening practices. *J Women's Health & Gender-Based Med,* 9(7), 791.

Gennaro, G., Baldelli, P. et al. (2004). Patient dose in full-field digital mammography: an Italian survey. *Eur Radiol,* 14(4), 645–652.

Gennaro, G. & di Maggio, C. (2006). Dose comparison between screen/film and full-field digital mammography. *Eur Radiol,* 16(11), 2559–2566.

Gitlin, J. N., White, D. M. et al. (2003). Field trial of mobile digital telemammography Phase I – installation and testing of equipment. In: *CARS 2003: Computer Assisted Radiology and Surgery, Proceedings,* 1256, pp. 305–310. International Congress Series.

Glasziou, P. P., Woodward, A. J. & Mahon, C. M. (1995). Mammographic screening trials for women aged under 50: a quality assessment and meta-analysis. *Med J Aust,* 162(12), 625–629.

Gøtzsche, P. C. & Nielsen, M. (2006). Screening for breast cancer with mammography. *Cochrane Database of Systematic Reviews: Reviews 2006, Issue 4*. John Wiley & Sons, Chichester, UK. DOI: 10.1002/14651858.CD001877.pub2, (4).

Gur, D. & Sumkin, J. H. (2006). CAD in screening mammography. *AJR Am J Roentgenol,* 187(6), 1474.

Hayes, T. E., Peterson, L. E. & Weinberg, A. D. (1998). Current guidelines for the diagnosis and treatment of breast cancer. *Disease Manage Health Outcomes,* 3(5), 239–250.

Hemdal, B., Andersson, I. et al. (2005a). Can the average glandular dose in routine digital mammography screening be reduced? A pilot study using revised image quality criteria. *Radiat Prot Dosim,* 114(1–3), 383–388.

Hemdal, B., Herrnsdorf, L. et al. (2005b). Average glandular dose in routine mammography screening using a Sectra MicroDose Mammography unit. *Radiat Prot Dosim,* 114(1–3), 436–443.

Hemminger, B. M. (2003). Soft copy display requirements for digital mammography. *J Dig Imaging,* 16(3), 292–305.

Hendrick, R. E., Smith, R. A. et al. (1997). Benefit of screening mammography in women aged 40–49: a new meta-analysis of randomized controlled trials, *J Natl Cancer Inst Monogr,* 22, 87–92.

Hendrick, R. E., Lewin, J. M. et al. (2001). Non-inferiority study of FFDM in an enriched diagnostic cohort: comparison with screen-film mammography in 625 women. In: *IWDM 2000, 5th Int Workshop on Digital Mammography*, 11–14 June 2000, Toronto. 475–481.

Hermann, K. P., Obenauer, S. et al. (2002). Average glandular dose with amorphous silicon full-field digital mammography – clinical results. *Rofo,* 174(6), 696–699.

Hopper, J. L., Southey, M. C. et al. (1999). Population-based estimate of the average age-specific cumulative risk of breast cancer for a defined set of protein truncating mutations in *BRCA1* and *BRCA2*. *Cancer Epidemiol Biomarkers Prevent,* 8, 741–747.

Humphrey, L., Helfand, M. et al. (2002). Breast cancer screening: a summary of the evidence for the US Preventive Services Task Force. *Ann Intern Med,* 137(5 part 1), 347–360.

Ingal, V. N., Beliaevskaya, E. A. et al. (1998). Phase mammography – a new technique for breast investigation. *Phys Med Biol,* 43(9), 2555–2567.

Irwig, L., Houssami, N. & van Vliet, C. (2004). New technologies in screening for breast cancer: a systematic review of their accuracy. *Br J Cancer,* 90(11), 2118–2122.

Jackson, V. P., Reynolds, H. E. & Hawes, D. R. (1996). Sonography of the breast. *Seminars in Ultrasound CT and MRI,* 17(5), 460–475.

James, J. J. (2004). The current status of digital mammography. *Clin Radiol,* 59(1), 1–10.

Jong, R. & Yaffe, M. (2005). Digital mammography: 2005. *Can Assoc Radiol J,* 56(5), 319–323.

Kerlikowske K., Grady D. & Rubin S. M. (1995). Efficacy of screening mammography: a meta-analysis. *J Am Med Assoc,* 273, 149–154.

Khan, K. S., Ter Riet, G. et al. (2001). *Undertaking Systematic Reviews of Research on Effectiveness. CRD's guidance for those carrying out or commissioning reviews*. NHS Centre for Reviews and Dissemination, University of York, York.

Kimme-Smith, C. (1999). New digital mammography systems may require different X-ray spectra and, therefore, more general normalized glandular dose values. *Radiology,* 213(1), 7–10.

Kimme-Smith, C., Yee T. Y., Bassett, L. W. (2000). Decreased Breast Compression Using Digital Mammography at Screen/Film Mean Glandular Dose. In Yaffe, M. J. P. O. (ed.) *5th International Workshop on Digital Mammography 2000*. Toronto, Canada, June 2000, Springer, New York.

Konica Minolta (2006). Konica Minolta's Regius PureView mammography system [internet]. Available from: http://kmmi.konicaminolta.us/eprise/main/kmmi/content/news/detail/press/PureViewMamSystem [Accessed 17 May 2007].

Kricker, A. & Jelfs, P. (1996). *Breast Cancer in Australian Women 1921–1994* [internet]. Australian Institute of Health and Welfare, Canberra, ACT. Available from: www.aihw.gov.au. [Accessed September 2006].

Leader, J. K., Hakim, C. M. et al. (2006). A multisite telemammography system for remote management of screening mammography: an assessment of technical, operational, and clinical issues. *J Dig Imag,* 19(3), 216–225.

Legood, R. & Alastair, G. (2004). *A Cost Comparison of Full Field Digital Mammography (FFDM) with Film-Screen Mammography in Breast Cancer Screening*. Health Economics Research Centre, University of Oxford, Oxford.

Lewin, J. M., D'Orsi, C. J. et al. (2002). Clinical comparison of full-field digital mammography and screen-film mammography of breast cancer. *Am J Roentgenol,* 179(3), 671–677.

Lewin, J. M., Hendrick, R. E. et al. (2001). Comparison of full-field digital mammography with screen-film mammography for cancer detection: results of 4,945 paired examinations. *Radiology,* 218(3), 873–880.

Lloyd, S., Jirotka, M. et al. (2005). Digital mammography: a world without film? *Methods Inf Med,* 44(2), 168–171.

Lostumbo, L., Carbine, N. et al. (2004). Prophylactic mastectomy for the prevention of breast cancer. *Cochrane Database of Systematic Reviews: Reviews 2004, Issue 4*. John Wiley & Sons, Chichester, UK. DOI: 10.1002/14651858.CD002748.pub2, (4).

MBS (2007). *Medicare Benefits Schedule (MBS)* [internet]. Available from: http://www9.health.gov.au/mbs/search.cfm. Commonwealth of Australia [Accessed 13 March 2007].

McLean, I. D., Heggie, J. C. P. et al. (2007). ACPSEM position paper: Interim recommendations for a digital mammography quality assurance program. *Australasian Physical & Engineering Sciences in Medicine* 30(2), 65-100.

Medicare Australia (2006). *Medicare Benefits Schedule (MBS) Item Statistics Reports* [internet]. Australian Government. Available from: http://www.medicareaustralia.gov.au/statistics/dyn_mbs/forms/mbs_tab4.shtml [Accessed 28 November 2006].

Moher, D., Cook, D. J. et al. (1999). Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement: Quality of Reporting of Meta-analyses. *Lancet,* 354(9193), 1896–1900.

Moran, P., Chevalier, M. et al. (2005). A survey of patient dose and clinical factors in a full-field digital mammography system. *Radiat Prot Dosim,* 114(1–3), 375–379.

MSAC (2005). *Guidelines for the Assessment of Diagnostic Technologies* [internet]. Medical Services Advisory Committee, Canberra. Available from: http://www.msac.gov.au/ [Accessed August, 2005].

MSAC (2007). *Breast Magnetic Resonance Imaging. MSAC Application 1098* [internet]. Commonwealth of Australia. Available from: http://www.msac.gov.au/internet/msac/publishing.nsf/Content/AD35ED216E990FC7CA2571420004A192/$File/1098%20final%20report.pdf [Accessed August 2007].

NBCC (National Breast Cancer Centre) (2004). *Breasthealth* [internet]. Available from: http://www.breasthealth.com.au/index.html [Accessed 18th September 2007].

NBCC (National Breast Cancer Centre) (2006). *Advice About Familial Aspects of Breast and Epithelial Ovarian Cancer: A guide for health professionals*. National Breast Cancer Centre, Sydney.

NCCN (National Comprehensive Cancer Network) (2006). *Breast Cancer. Clinical Practice Guidelines in Oncology, Version 2.2006* [internet]. Available from: http://www.nccn.org/professionals/physician_gls/PDF/breast.pdf [Accessed December 2006].

NHMRC (1999). *Familial Aspects of Cancer: A guide to clinical practice*. National Health and Medical Research Council, Canberra.

NHMRC (2000). *How to Use the Evidence: Assessment and application of scientific evidence*. National Health and Medical Research Council, Canberra.

NHMRC (2001). *Clinical Practice Guidelines for the Management of Early Breast Cancer*, 2nd ed. National Health and Medical Research Council, Canberra.

NHMRC (2005). *NHMRC additional levels of evidence and grades for recommendations for developers of guidelines, Pilot Program 2005*, National Health and Medical Research Council. [internet]. Available from: www.nhmrc.gov.au/consult/index.htm [Accessed June 2005]

Peer, S., Faulkner, K. et al. (2005). Relevant training issues for introduction of digital radiology: results of a survey. *Radiat Prot Dosim,* 117(1–3), 154–161.

Pisano, E. D., Cole, E. B. et al. (2002). Interpretation of digital mammograms: comparison of speed and accuracy of soft-copy versus printed-film display. *Radiology,* 223(2), 483–488.

Pisano, E. D. & Yaffe, M. J. (2005). Digital mammography. *Radiology,* 234(2), 353–362.

Pisano, E., Gatsonis, C. et al. (2005a). American College of Radiology Imaging Network digital mammographic imaging screening trial: objectives and methodology. *Radiology* 236(2), 404–412.

Pisano, E. D., Gatsonis, C. et al. (2005b). Diagnostic performance of digital versus film mammography for breast-cancer screening. *N Engl J Med,* 353(17), 1773–1783.

Samei, E., Saunders, R. S., Jr et al. (2007). Digital mammography: effects of reduced radiation dose on diagnostic performance. *Radiology,* 243(2), 396–404.

Sapir, R., Patlas, M. et al. (2003). Does mammography hurt? *J Pain Symptom Manag,* 25(1), 53–63.

Scott, C. L., Jenkins, M. A. et al. (2003). Average age-specific cumulative risk of breast cancer according to type and site of germline mutations in *BRCA1* and *BRCA2* estimated from multiple-case breast cancer families attending Australian family cancer clinics. *Human Genet,* 112(5–6), 542–551.

Seo, B. K., Pisano, E. D. et al. (2006). The positive predictive value for diagnosis of breast cancer full-field digital mammography versus film-screen mammography in the diagnostic mammographic population. *Acad Radiol,* 13(10), 1229–1235.

Skaane, P., Young, K. & Skjennald, A. (2003). Population-based mammography screening: comparison of screen-film and full-field digital mammography with soft-copy reading – Oslo I study. *Radiology,* 229(3), 877–884.

Skaane, P. & Skjennald, A. (2004). Screen-film mammography versus full-field digital mammography with soft-copy reading: randomized trial in a population-based screening program – the Oslo II Study. *Radiology,* 232(1), 197–204.

Skaane, P., Skjennald, A. et al. (2005). Follow-up and final results of the Oslo I Study comparing screen-film mammography and full-field digital mammography with soft-copy reading. *Acta Radiol,* 46(7), 679–689.

Smart, C. R. (1994). Highlights of the evidence of benefit for women aged 40–49 years from the 14-year follow-up of the Breast Cancer Detection Demonstration Project. *Cancer,* 74(1 suppl), 296–300.

Smart, C. R., Hendrick, R. E. et al. (1995). Benefit of mammography screening in women ages 40 to 49 years: current evidence from randomized controlled trials. *Cancer,* 75(7), 1619–1626.

Stata Corporation (2005). *Intercooled Stata 9.0 for Windows*. College City, TX, USA.

Tabar, L., Vitak, B. et al. (2000). The Swedish two-county trial twenty years later: updated mortality results and new insights from long-term follow-up. *Radiol Clinics N Am,* 38(4), 625–651.

Venta, L. A., Hendrick, R. E. et al. (2001). Rates and causes of disagreement in interpretation of full-field digital mammography and film-screen mammography in a diagnostic setting. *AJR Am J Roentgenol,* 176(5), 1241–1248.

Whiting, P. R. A., Reitsma, J. B. et al. (2003). The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol,* 3(1), 25.

Woolf, S. H. (2001). The accuracy and effectiveness of routine population screening with mammography, prostate-specific antigen, and prenatal ultrasound: a review of published scientific evidence. *Int J Technol Assess Health Care,* 17(3), 275–304.

Wright, C. J. & Mueller, C. B. (1995). Screening mammography and public health policy: the need for perspective. *Lancet,* 346(8966), 29–32.

Yamada, T., Saito, M. et al. (2004). Comparison of screen-film and full-field digital mammography in Japanese population-based screening. *Radiat Med,* 22(6), 408–412.