Chapter 2
Screening techniques

Screening mammography

Mammography is an X-ray technique that was developed specifically for soft tissue radiography of the breast. It is based on the differential absorption of X-rays between the various tissue components of the breast such as fat, fibroglandular tissue, tumour tissue and calcifications. Mammography is used both as a clinical tool to examine symptomatic patients and as a screening examination. The goal of screening mammography is to detect breast cancer early (Figure 22). To reach this goal, mammographs of consistently high quality must be produced with minimal exposure of the women to radiation.

X-ray equipment

The physics of modern screen–film mammography techniques have been reviewed elsewhere (Säbel & Aichinger, 1996; Barnes, 1999; Haus, 1999; Hendrick & Berns, 1999; Haus & Yaffe, 2000).

Image contrast and spatial resolution are important determinants of image quality in mammography. The contrast depends on many factors, such as beam quality, screen–film combination, film processing and scattering of radiation.

Beam quality

The image contrast depends on the energy distribution of the radiation used. The attenuation coefficients of fat, fibroglandular tissue and tumour tissue differ more at lower energy (below 20 keV) than at higher energy. However, compromises have to be made to keep the exposure within acceptable limits. This is accomplished by using various target–filter combinations. Most current mammography systems have molybdenum targets combined with molybdenum filtration; many also have rhodium filtration. Dual-target tubes, such as molybdenum–wolfram or molybdenum–rhodium, are also available.

The combination of a molybdenum target with molybdenum filtration results in an energy distribution that is ideal for imaging small-to-medium-sized breasts (energy, 15–20 keV) at tube voltages of 25–30 kVp. Increasing the voltage increases the penetration of the beam and thus decreases the dose; however, it also decreases the image contrast by decreasing the attenuation differences. Switching to rhodium filtration and rhodium or wolfram targets has the same effect (Thilander-Klang, 1997; Thilander-Klang *et al.,* 1997; Figures 23 and 24). Many modern mammography units have programmes that choose automatically, on the basis of breast thickness and composition, a target–filter combination, that represents a reasonable compromise between image contrast and dose.

Tube current

The tube current must be high enough to produce adequate film density with short exposure. Exposure times longer than about 1 s imply a risk of added dose and also lack of sharpness due to motion. Rhodium targets cannot be operated at

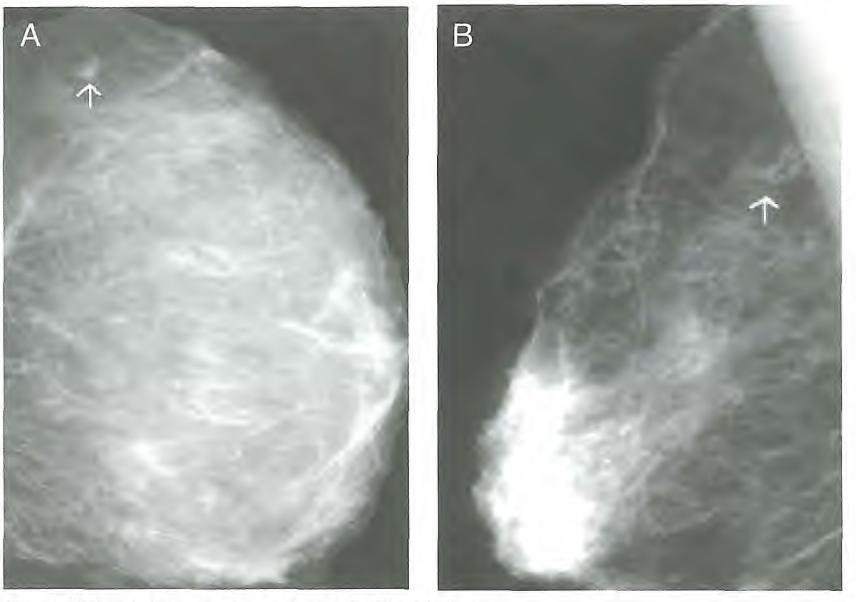


Figure 22 Screening mammograms of early breast lesions
A, slightly spiculated tumour measuring 0.5 cm (arrow); a 0.6-cm invasive ductal carcinoma grade 2 was found on microscopy; **B**, cluster of calcifications (arrow); a ductal carcinoma *in situ* grade 2 measuring 1.4 cm was found on microscopy

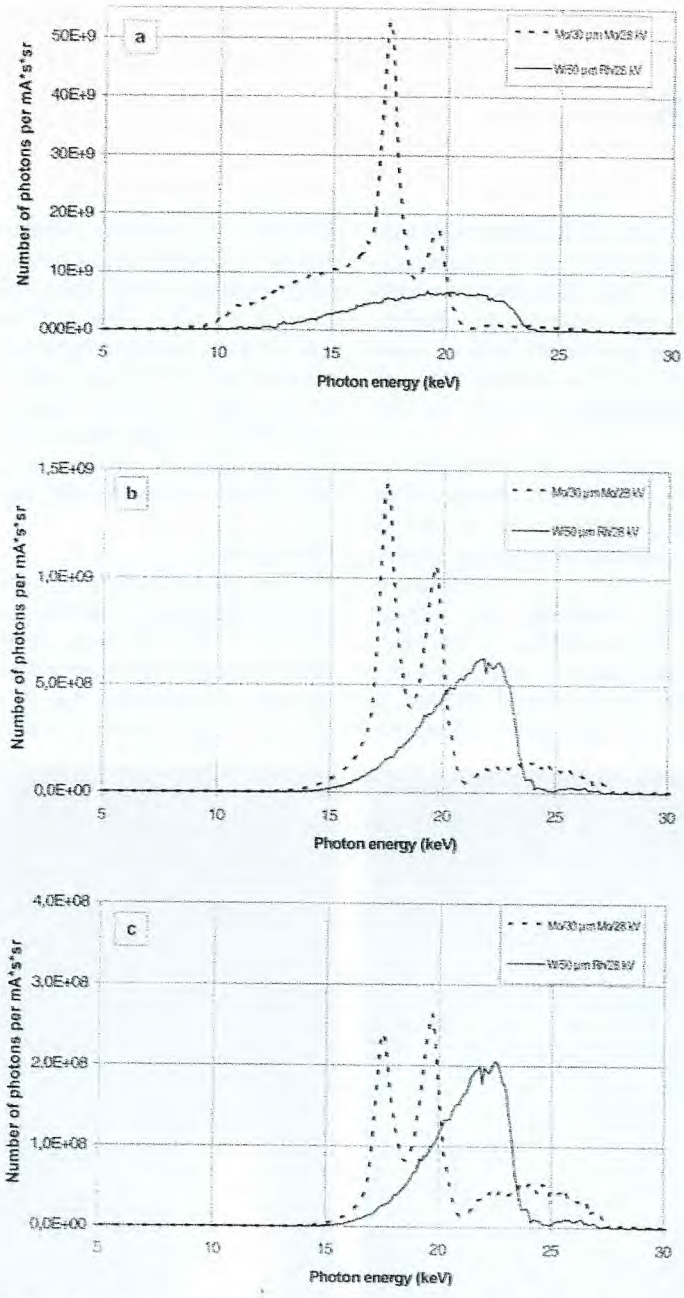


Figure 23 Absolute X-ray spectra, with (a) measured entrance, (b) calculated transmitted energy with 40-mm polymethyl methylacrylate phantom and (c) calculated transmitted energy with 60-mm polymethyl methylacrylate phantom, obtained with various anode–filter combinations and a tube voltage of 28 kV

Dotted lines, molybdenum–molybdenum; solid lines, wolfram–rhodium

as high a tube current as molybdenum and wolfram targets because of a lower melting-point.

Scatter control

One important factor that degrades image contrast is scattering of radiation (Friedrich, 1975; Barnes & Brezovich, 1978). The amount of scatter depends heavily on breast thickness and to some extent also on breast area. This is one of the reasons why breast compression is necessary to obtain a good mammogram. Further control of scatter can be obtained locally by spot compression, which can be combined with magnification.

The most important means of scatter reduction is the anti-scatter grid, which consists of thin lead lamellae separated by a radiolucent spacer. The grid ratio (height of lead lamellae divided by inter-space thickness) is usually 4:1 or 5:1. The grid usually starts moving during exposure and acts to absorb most of the scattered radiation (75–85%) while transmitting most of the primary radiation (60–75%) (Yester *et al.*, 1981). As a result, image contrast is improved at the price of increased dose. Improvement in contrast is related to breast thickness, being greater with increasing thickness. The grid technique was introduced in the late 1970s. It was not used in most of the mammography screening trials, the exception being that in Göteberg, where it was used throughout (Bjurstam *et al.*, 1997).

An even more efficient way of reducing scatter is the slot scanning technique (Barnes *et al.*, 1989a,b), in which the X-ray beam is collimated to a thin fan beam which is scanned across the breast. Conventional linear grids reduce scatter only perpendicular to the grid septa, and there is little reduction in the direction parallel to the grid lines. Another solution is the cellular grid (Rezentes *et al.*, 1999), which has a square pattern and therefore controls scatter in two dimensions. These grids have been shown to reduce scattered radiation further and

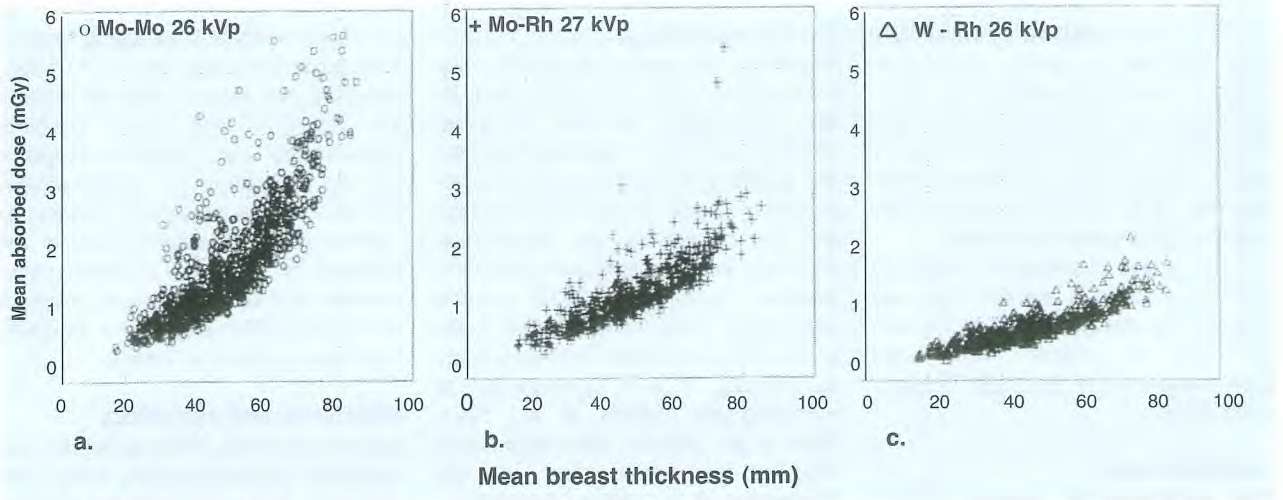


Figure 24 Mean absorbed dose to glandular tissue versus mean breast thickness with (a) molybdenum–molybdenum, (b) molybdenum–rhodium and (c) wolfram–rhodium target–filter combinations

thus to improve image contrast over conventional grids.

Still another way of reducing scatter is geometric magnification with an air gap. The magnification factor is usually 1.7–2.0. For magnification work, a focus of 0.10–0.15 is necessary. Modern mammography machines are equipped with magnification capabilities.

Automatic exposure control

Adequate automatic exposure control makes it possible to achieve optimal, reproducible density of images, independently of breast thickness and the beam quality used. Automatic exposure control devices have been refined substantially over the past few years. Most can be operated either manually or automatically. In the automatic mode, the instrument can choose both voltage and filter and, in some cases, also the target, depending on the thickness and density of the breast. The sensor should be of sufficient size and location to cover various components of breast tissue, and it should be moveable away from the chest wall, as the exposure should be determined by the densest part of the breast.

An automatic exposure control device should meet certain standards. A minimum requirement is that it should maintain a uniform film density to ± 0.10 or 0.15 optical density unit when the thickness of the phantom varies from 2 to 7 cm, for all techniques used (Socialstyrelsen, 1998). Optical density is one determinant of the sensitivity of mammography (Young *et al.*, 1997). According to the European guidelines (Commission of the European Communities, 2001), the optical density should be between 1.4 and 1.8 (National Health Service Breast Screening Programme, 1998).

Compression

Optimal compression is an important part of the mammography procedure. Compression improves contrast by reducing scatter and hardening the X-ray beam and also reduces the dose to the breast. Furthermore, patient motion is reduced, and the density of the image becomes more uniform. With proper compression, the structures of the breast are spread apart, facilitating image interpretation.

Screen film, processing and viewing

The introduction of rare-earth intensifying screens in the mid-1970s represented a major step forward, as these screens could be combined with fast-speed films, thereby reducing the dose. The screens are virtually always used as back-screens combined with a single emulsion film in order to achieve optimal spatial resolution. Various phosphors have been used; one that is commonly used is gadolinium oxysulfide (Gd_2O_2S:Tb), which emits visible light in the green spectral region (wavelength, about 500 nm).

Film processing is critical for obtaining a high-quality mammogram, and suboptimal image quality is frequently due to suboptimal processing. Processing is one of the key determinants of film contrast, as reflected in the so-called characteristic curve of the film. Critical factors in processing are temperature, processing time and replenishment rate.

Extended cycle processing is sometimes used for single-emulsion films (Kimme-Smith *et al.*, 1989), thereby increasing film contrast and speed. Today, a 90-s processing cycle is recommended by almost all film manufacturers.

Once the key processing parameters have been set, a quality control programme should be implemented. Base parameters like film speed, contrast and base plus fog can be determined easily by sensitometry, either by manual measurement or by running a sensitometry strip through an automatic reader.

Light boxes of adequate luminance and a low level of ambient light are important for viewing, as well as masking of films to reduce stray light (Commission of the European Communities, 2001).

Radiation dose

The mammographic imaging system must be optimized in order to keep the radiation dose as low as possible. Such a requirement has been included in national legislation in some countries and in international guidelines.

The mean absorbed dose in the breast gland per mammographic film is in the order of 1.0–1.5 mGy for the average breast examined with modern equipment. Surveys have shown considerable variation in dose among centres. In southern Sweden, the dose varied by a factor of 4.3, depending on the radiologists' preference in terms of optical density of the films, variation in film processing parameters and other factors (Socialstyrelsen, 1997). In Sweden, the mean absorbed dose to the breast gland per film must not exceed 1.5 mGy at the optical density setting used and should not exceed 1.0 mGy at net density 1.0 as measured with a 4.5-cm polymethyl methylacrylate phantom according to the European protocol (Zoetelief *et al.*, 1996). According to the European guidelines (Commission of the European Communities, 2001), the dose should be < 2.0 mGy. A woman being screened every 2 years between the ages of 40 and 70, with two views of each breast, would thus receive an absorbed dose of 64 mGy or less from the screening. For a discussion of the possible hazard of mammographic radiation, see Chapter 5.

Quality control

Guidelines for quality assurance have been issued by several bodies, such as the Commission of the European Communities (2001). Many factors affect the accuracy of mammography, including those related to the X-ray machine and film processing, the examination technique including positioning and compression (Taplin *et al.*, 2002) and the radiologist's performance. It has been shown that radiologists vary, sometimes substantially, in their interpretation of mammograms (Elmore *et al.*, 1994; Beam *et al., 1*996a,b). One determinant may be the volume read per day (Esserman *et al.*, 2002). Recommendations vary regarding the minimum number of mammograms that should be read yearly, from 480 to 5000 (Food & Drug Administration, 1997; National Health Service Breast Screening Programme, 1998). Another factor is training, which has been shown to improve sensitivity with no change in specificity (Linver *et al*., 1992).

Continuous correlation of radiographic findings with cytology and pathology is another essential component, with training and continuing education. Furthermore, a database should be established that contains basic information such as patient identification, date of mammographic examination, mammographic diagnosis, results of needle biopsy and surgical procedures, including microscopic diagnosis. If cancer is present, the tumour size, lymph node status, malignancy grading and the presence or absence of distant metastasis should be recorded.

High, consistent image quality is mandatory to achieve the objectives of mammography. To maintain the image quality at an acceptable level, regular tests must be carried out. The day-to-day consistency of the procedure should be based on sensitometry and phantom exposure. While sensitometry specifically monitors the performance of the processor, phantom exposure provides an overall check of the imaging system. If the process is stable, as shown by sensitometry, the phantom film will indicate the status of the X-ray machine. Sensitometry and phantom exposure can be performed by radiographers, whereas several parameters relating to mammography machines should be checked by a medical physicist semi-annually or at least annually. An example of a quality control programme for mammography is shown in Table 4.

Sensitivity and specificity

Several estimates of the sensitivity and specificity of mammography have been published. In most of them, the cancers detected at screening, expressed as the proportion of all these cancers and those occurring during the first 12 months after screening ('interval cancers') were used as a proxy for sensitivity. This was called the 'detection' method by Fletcher *et al*. (1993). The preferred expression for sensitivity is 1 minus the incidence of interval cancers expressed as a proportion of the estimated underlying incidence of breast cancer in the population. This was called the 'incidence' method by Fletcher *et al*. (1993; see Chapter 1).

Table 5 summarizes estimates of the sensitivity, specificity and positive predictive value of mammography, with or without clinical breast examination, as reported in breast screening trials (described above in Chapter 1 and more fully in Chapter 4) and some population-based screening programmes, which covered screening from as early as 1963 to as late as 1997. In all instances in this table, the estimates of sensitivity are based on 1-year interval cancer rates and are calculated by the detection method, the incidence method or both (Fletcher *et al*., 1993). The estimates of specificity and predictive positive value take into consideration all women referred for further investigation after a positive result at screening. Unless otherwise specified, the estimates are based on invasive cancers only in the

Table 4. Technical quality control programme for mammography used in southern Sweden

Function to be checked	Method, test and/or responsibility	Frequency
Film processing	Sensitometry (technologist or radiologist)	Daily
Entire imaging process	Phantom exposure (technologist or radiologist) Visual comparison with reference	Daily
Phototimer	PMMA phantom exposure with recording of milliamperes (technologist or radiologist)	Daily if batch processing
Beam quality: filtration, tube potential, half-value layer	Service or physicist	Annually or semi-annually
Phototimer: reproducibility, dependence on object thickness and tube potential	Service or physicist	Annually or semi-annually
Output: millampere accuracy and linearity	Service or physicist	Annually or semi-annually
Beam geometry: radiation field extension	Service or physicist	Annually or semi-annually
Compression device	Service or physicist	Annually or semi-annually
Film cassettes: sensitivity, film screen contact, spatial resolution	Service or physicist	Annually or semi-annually
Anti-scatter grid	Service or physicist	Annually or semi-annually
Absorbed dose	Physicist	Annually or semi-annually

Modified from Commission of the European Communities (2001)
Milliamperes are the product of tube current x length of exposure
PMMA, polymethyl methacrylate

first screening round or the combination of second and subsequent screening rounds.

The estimates of sensitivity derived with the detection method, available from almost all the programmes, varied from low values of 68% (Stockholm trial of one-view mammography) and 74% (Health Insurance Plan trial of early two-view mammography and clinical breast examination) to high values of over 90% in several populations. There is no strong evidence that the sensitivity of these programmes increased over time. As expected from differences in the way in which they are computed, the estimates of sensitivity derived with the preferred incidence method were generally smaller than those computed with the detection method and varied by 52–82%, again with little evidence of a trend over time. The estimates of sensitivity were generally higher by a small margin in first than in subsequent screening rounds. The estimates of specificity were derived mainly from the screening trials and exceeded 90%, with few exceptions; many exceeded 95%. The corresponding values for positive predictive value ranged from 2% to 22%; most were 12% or less.

Estimates of sensitivity for women in different age groups have been reported from a number of studies, and some are shown in Table 5. In addition, Tabár *et al.* (1987b) reported estimates obtained by

Table 5. Summary of estimates of sensitivity and specificity of breast cancer screening by mammography with or without clinical breast examination reported from screening trials and some population-based screening programmes

Trial or programme	Screening method, period, age group	Sensitivity (%)		Specificity (%)	PPV (%)
		Detection method	Incidence method		
Health Insurance Plan (Fletcher *et al.*, 1993; Shapiro, 1997)	Two-view mammmography and CBE 1963–66, 40–64	74	77	98.5	12
Breast Cancer Detection Demonstration Project (Seidman *et al.*, 1987)	Two-view mammography, CBE and thermography 1972–81, 40–59	88[1], 84[2]			
Utrecht, Netherlands (de Waard *et al.*, 1984a)	Two-view mammography and CBE 1974–80, 50–67	91			
Nijmegen, Netherlands (Verbeek *et al.*, 1988)	One-view mammography 1975–85, 35–64	89	82		
Malmö, Sweden (Fletcher *et al.*, 1993)	Two-view mammography 1976–90, 43–70	79	68	96[1], 97[2]	10[1], 22[2]
Two-county trial, Sweden (Fletcher *et al.*, 1993)	One-view mammography 1977–81, 40–75	76	60[1],70[2]	95[1], 98[2]	12
Edinburgh, Scotland (Fletcher *et al.*, 1993)	Two-view mammography and CBE 1979–86, 45–64	88	79	96[1], 97[2]	15[1], 4[2]
Edinburgh, Scotland (Chamberlain *et al.*, 1991)	Two-view mammography and CBE 1979–86, 45–64	92[1],932	73[1,3],78[1,3]	96[1], 97[2]	15[1], 4[2]
Guildford, England (Chamberlain *et al.*, 1991)	Two-view mammography and CBE 1979–86, 45–64	94[1], 90[2]	73[1,3],78[1,3]	92[1], 94[2]	6[1], 2[2]
Canada 1 (Fletcher *et al.*, 1993)	Two-view mammography and CBE 1980–85, 40–49	81	58	82[1], 93[2]	2
Canada 2 (Fletcher *et al.*, 1993)	Two-view mammography and CBE 1980–85, 50–59	88	72	83[1], 96[2]	4[1],6[2]
Stockholm, Sweden (Fletcher *et al.*, 1993)	One-view mammography 1981–83, 40–64	68[1]	75	95[1], 97[2]	8[1],10[2]
Stockholm, Sweden (Fletcher *et al.*, 1993)	One-view mammography 1981–83, 40–49	53[1]	39		

Table 5. (contd)

Trial or programme	Screening method, period, age group	Sensitivity (%)		Specificity (%)	PPV (%)
		Detection method	Incidence method		
Göteberg, Sweden (Bjurstam *et al.*, 1997)	Two-view mammography 1982–84, 39–49		82		
Sydney, Australia (Rickard *et al.*, 1998)	Two-view, double reader film–screen mammography 1988–92, 40–69		71		
Ontario, Canada (Libstug *et al.*, 1998)	Two-view mammography and CBE 1990–95, 50–69	90[1], 81[2]		87[1], 92[2]	6.7[1], 6.2[2]
British Columbia, Canada (Olivotto *et al.*, 2000)	Two-view mammography 1988–97, ≥ 40	86 86		94	6
East Anglia, England (Day *et al.*, 1995)	One- and two-view mammography 1990–93, 52–64		76[4]		
Netherlands (Fracheboud *et al.*, 1999)	One- and two-view mammography 1990–93, 50–69	92[1], 85[2]	73[1], 74[2]		
Victoria, Australia (BreastScreen Victoria, 2001)	Two-view, double reader film–screen mammography 1996, 50–69	91[1], 82[2]			

PPV, positive predictive value; CBE, clinical breast examination
[1] First round
[2] Subsequent rounds
[3] Edinburgh and Guildford combined

the incidence method for the Two-county trial of 62%, 88% and 85% for women aged 40–49, 50–59 and 60–69, respectively. In a study in Utrecht, the Netherlands, in which mammography and clinical breast examination were used, Day *et al*. (1988) reported a sensitivity of screening of 83% for women aged 50–59 and 86% for women aged 60–64. In a study in Nijmegen, in which mammography alone was used, Verbeek *et al*. (1988) reported a sensitivity of 44% for women aged 35–49 and 75% for women aged 50–64. Peer *et al*. (1996) later replicated this age difference in screening rounds four through eight, with an estimate of 64% for women under 50 and 85% for those above 50. Chamberlain *et al*. (1991) evaluated sensitivity by age for all screenings in the combined programmes in Edinburgh and Guildford (United Kingdom); the sensitivity was 70% for women aged 45–54 at entry and 84% for those aged 55–64 at entry. In the screening programme in British Columbia, Canada, the sensitivity (with the detection method) was 76% for women aged 40–49, 85% for those aged 50–59, 90% for those aged 60–69, 91% for those aged 70–79 and 91% for those aged 80 or more (Olivotto *et al*., 2000). Thus, there was a consistent trend for increasing sensitivity with increasing age.

In the Canadian trials, review by the reference radiologist allowed identification of the cancers missed by the radiologists in the screening centre and suspected by the reference radiologist. This process included both the interval cancers and the cancers detected at the second screening (Baines *et al*., 1986a). These, together with the cancers identified by physical examination but missed on mammography, allowed identification

Mammographic density

- Breast parenchymal 'density' as seen on a mammogram is a determinant of the sensitivity of mammography.
- Breast parenchymal density decreases with age.
- Hormone replacement therapy of the combination type may result in increased breast density.
- Tamoxifen may decrease breast density.

of the false-negative findings. This is in practice a refinement of the detection method. On this basis, two reports of the sensitivity of mammography were made for the first screening for both components of the trial together (i.e. for women aged 40–59 on entry). The first related to the first five centres in the trial, which were entered in 1980 and 1981 (Baines *et al.*, 1986b), with an overall sensitivity of 69%, a specificity of 94%, and a positive predictive value of 8.6%. Baines *et al*. (1988a) subsequently reported the sensitivity of the first screening in all 15 centres to be 75%, a specificity of 94% and a positive predictive value of 7%. The authors postulated that the differences in sensitivity between the first and second reports were a consequence of a general improvement in mammography with time since the trial was initiated, and the benefit the later centres derived from entering the trial with mammography quality control procedures fully in place.

Chamberlain *et al.* (1991) determined what they called the 'relative' sensitivities of mammography and clinical breast examination as the proportion detected by each of all cancers found at each round in the Trial of Early Detection of Breast Cancer in the United Kingdom. The relative sensitivity of mammography was 94% at the prevalence screen and 90% at the incidence screens. For comparison, the relative sensitivities were 72% and 45% for clinical breast examination.

Host factors that affect sensitivity

Mammography is based on the principle of differential absorption of X-rays between fat, fibroglandular tissue, tumour tissue and calcifications, fat being more radiolucent (blacker on the film) than the other tissues (which are 'denser' or whiter on the film). Thus, the density of a mammogram is determined by the relationship between fat and fibroglandular tissue or tumour tissue, the mammogram being 'denser' the more of the latter tissue components are present. The mammographic pattern of the breast thus varies between individuals (Figure 25). Furthermore, breast cancer is more readily detected in a fatty breast than in a dense breast (Mandelson *et al.*, 2000).

In addition to age, several other factors seem to be related to the amount of fibroglandular tissue in the breast,

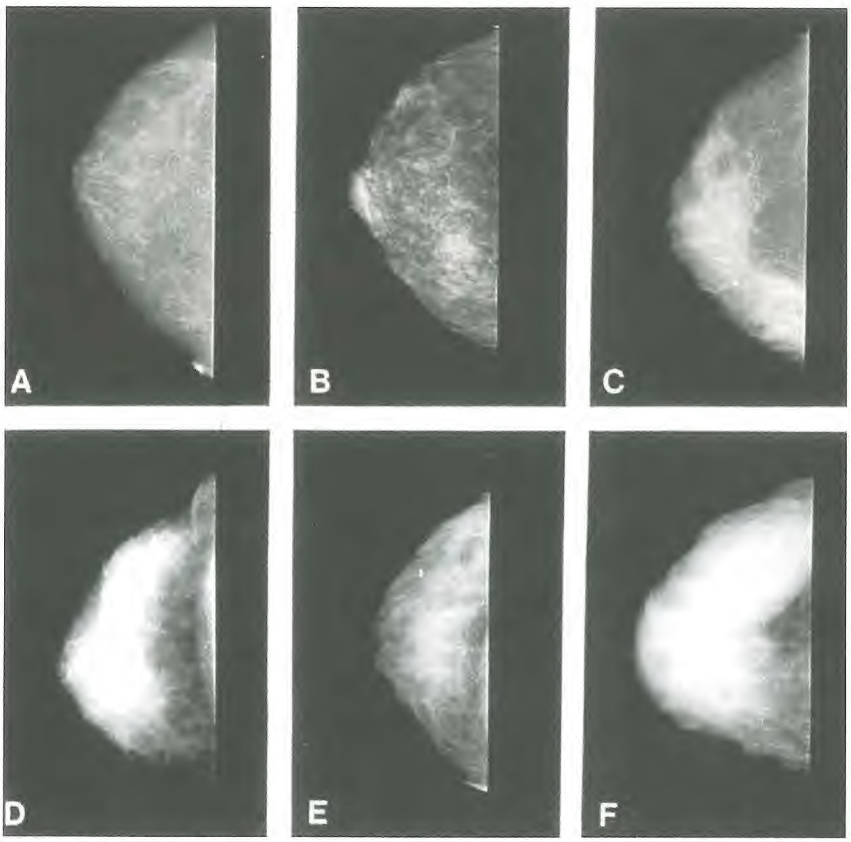


Figure 25 Six categories of mammographic density. A = 0%; B = 0 < 10%; C = 10 < 25%; D = 25 < 50%; E = 50 < 75%; F = > 75%
From Boyd *et al.* (2001)

including parity and age at birth of first child (Andersson *et al.*, 1981; de Waard *et al.*, 1984b); greater age at birth of first child and nulliparity are associated with denser breasts. There is also suggestive evidence that density may vary with the phase in the menstrual cycle, being on average greater in the luteal phase (White *et al.*, 1998). This might explain the lower sensitivity and specificity of mammography in women in the luteal phase than in the follicular phase seen in one study (Baines *et al.*, 1997).

It has been demonstrated fairly consistently that breast density increases in a certain proportion of women undergoing hormone replacement therapy, especially if they are treated with combinations of estrogen and progestin (Figure 26; Sala *et al.*, 2000). In one study, greater density was seen in 3.5–23.5% of women, depending on the preparation used (Greendale *et al.*, 1999). The increase in density usually appears within months after the start of treatment and appears to subside within a few months of termination of treatment.

Increased density after hormone replacement therapy can be assumed to decrease the sensitivity of mammography, and this has been demonstrated (Laya *et al.*, 1996). Kavanagh *et al.* (2000) reported a sensitivity of 80% for non-users of hormone replacement therapy and 64% for users in a large screening programme. Furthermore, the specificity was marginally lower for users. However, Thurfjell *et al.*, (1997) found no decrease in sensitivity of mammography in women on hormone replacement therapy.

Tamoxifen, which has mainly antiestrogenic effects, has been reported to decrease the density of the breast parenchyma in some women (Atkinson *et al.,* 1999; Chow *et al.*, 2000).

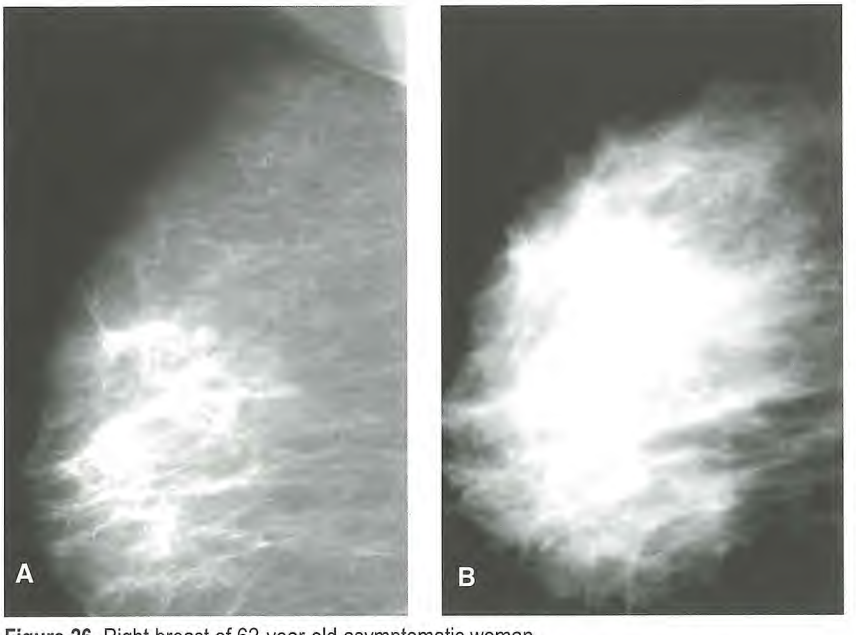


Figure 26 Right breast of 62-year-old asymptomatic woman
A, before hormone replacement therapy: the breast is predominantly fatty; **B,** after hormone replacement therapy (combined estrogen–progestin preparation): the density of the breast parenchyma has increased substantially

One *versus* two views

Screening with a single view (the mediolateral oblique) was suggested by a pioneer of mammographic screening, Lundgren (1977), on the presumption that virtually all breast cancers could be detected with one view. However, it was soon demonstrated that addition of a second view (the cranio-caudal) could improve sensitivity. The results of the Malmö mammographic screening trial suggested that 10–20% of invasive carcinomas < 10 mm in diameter would have been overlooked if only one projection had been used at screening. This applied mainly to mass lesions, while calcifications were consistently observed in both projections (Andersson, 1981).

Ample evidence in the same direction came from the screening programme in the United Kingdom (Wald *et al.,* 1995), which changed from using one to two views in the mid-1990s. A 25–42% increase in detection of invasive cancers < 15 mm in diameter was seen in incidence screens (Blanks *et al.*, 1997). Furthermore, the increase in sensitivity with two views was greatest for small cancers and cancers of low grade (Given-Wilson & Blanks, 1999).

The results of studies of the effect of two views on specificity varied. No significant change was noted in several, while a decrease was found in one study (Thurfjell *et al.*, 1994a). The results also indicated that the rate of false-positive findings was higher with one view only (Andersson, 1981).

The strategy used in several Swedish programmes is to classify the parenchyma as either 'dense' or 'not dense' at baseline, representing breasts with more than and less than approximately 25% fibroglandular tissue, respectively, as assessed visually. In subsequent screening rounds, 'not dense' breasts are examined with the oblique view only and at a 2-year interval. Women with 'dense' breasts are examined with two views, the cranio-caudal and oblique, at intervals of 18 months (Socialstyrelsen, 1998).

There are currently no published comparisons of this strategy with two-view mammography.

Double reading

An increase in sensitivity of 10–15% has been reported as the result of double reading compared with single reading (Anttinen *et al.*, 1993; Anderson *et al.*, 1994; Thurfjell *et al.*, 1994b). However, Ciatto and collaborators (1995) found only a 5% increase. Most authors report decreased specificity with independent double reading, although consensus decisions or arbitration on selected cases improves specificity (Anttinen *et al.*, 1993; Brown *et al.*, 1996). There are two main reasons for not detecting a significant lesion at screening: overlooking it or misinterpreting it. Of all interval cancers, 15–30% were found retrospectively to have been overlooked and about 15% misinterpreted (Ikeda *et al.*, 1992). Double reading can reduce these proportions and can detect some of the cancers that pass unnoticed until a subsequent screening. Furthermore, the wide variability in radiologists' interpretations of screening mammograms (Beam *et al.*, 1996a) can be partly offset by double reading (Beam *et al.*, 1996b).

Most authors recommend double reading, although Ciatto *et al.* (1995) questioned the cost–effectiveness of the procedure. In a Finnish programme, the incremental cost per cancer with double reading was not drastically higher than with single reading (39%) (Leivo *et al.*, 1999). Several factors have to be taken into consideration, such as the experience of readers (Warren & Duffy, 1995). With very experienced readers, the advantage of double reading is probably smaller (Ciatto *et al.*, 1995). In the European guidelines (Commission of the European Communities, 2001), double reading is mandatory in decentralized programmes and in programmes in which the radiologists are not yet sufficiently experienced. In centralized programmes with radiologists experienced in screening and diagnosis, double reading is not mandatory. Double reading practically doubles the resources required in terms of radiologists in a screening programme. Good results have been reported with suitably trained radiographers as second readers (Pauli *et al.*, 1996). Computer-aided detection systems may replace a second reader in the future (Warren Burhenne *et al.*, 2000).

Other and emerging imaging techniques

X-ray mammography is the only imaging method for breast cancer screening that has received serious evaluation. More recently, alternatives and adjuncts have begun to be evaluated, primarily for their potential in breast cancer diagnosis. This section deals with their potential application to breast cancer screening. An overview of the techniques described below is given in Table 6.

The evidence for the accuracy of recently proposed methods of screening is reviewed below. To avoid bias, the literature was reviewed systematically to ensure that all relevant studies had been located, and their quality and applicability were examined before their results were assessed (Glasziou *et al.*, 1999). To ensure the applicability of the results to screening, the studies had to have been done on women eligible for screening. Studies on women presenting clinically cannot be used to infer the accuracy of a new technique for screening, because the objective of testing is different. In clinical settings, the objective is to determine whether a previously detected abnormality is cancer. In screening, it is to identify abnormalities that may be found on further testing to be early cancers. Furthermore, the spectrum of disease is different, as the clinical abnormalities are larger and more advanced. Papers were therefore included only if they referred to new tests done in asymptomatic women, including populations at higher risk for breast cancer because of genetic predisposition or those in whom mammography is less accurate because they are younger or have radiologically dense breast tissue. Very few studies fulfilled these criteria. The remainder of the papers were review articles, were concerned with the development of tests or referred to use of tests in individual cases or as a diagnostic tool in women with a clinically or mammographically detected breast abnormality. Papers on screening were excluded if important technological changes made them no longer relevant. On these grounds, articles on thermography before 1988 were excluded, as were papers on ultrasonography with water baths or frequency probes with a resolution < 7.5 mHz.

No eligible papers were found for computed tomography scanning, magnetic resonance spectroscopy, scintimammography, electrical impedance or infrared spectroscopy. Light scanning and thermography have been suggested for screening but hold little promise. Light scanning was evaluated in two studies conducted over a decade ago (Alveryd *et al.*, 1990; Braddick, 1991), and thermography was evaluated in one study (Williams *et al.*, 1990); all suggested that these techniques are of insufficient accuracy, and no further eligible studies were identified. The results for the remaining techniques are described below. The relative sensitivities are presented for those studies in which interval cancers were not counted. Relative sensitivities allow comparison of tests but overestimate true sensitivity.

Digital mammography

In digital mammography, the image receptor (screen–film) used in conventional mammography is replaced by a digital receptor; in all other respects, the imaging techniques are the same. From the woman's point of view, receiving a digital mammogram is similar to having a conventional mammogram, as breast

Table 6. Other and emerging imaging techniques. Description and potential strenghts and limitations

Screening technique	Description	Potential strengths	Current limitations
Digital mammography	Electronic detectors capture X-rays in a matrix of square picture elements. Computer generates image.	Image processing Easy display, transmission and storage Lower radiation dose Computer-aided detection	Higher cost than mammography for low-volume operations
Ultrasonography	High-frequency ultrasound waves generate images based on the acoustic–mechanical properties of breast tissue	Increased sensitivity for mammographically dense breasts. No X-irradiation	Operator-dependent More expensive than mammography Less specific than mammography
Magnetic resonance imaging	Based on radiofrequency signals generated by exciting hydrogen nuclei (protons) in a strong magnetic field. Dynamic study of spatial and temporal distribution of intravenous contrast medium	More sensitive than mammography No X-irradiation	Less specific than mammography More expensive than mammography Claustrophobic
Positron emission tomography (PET)	Tomographic nuclear imaging procedure with positron-emitting tracers (usually fluorodeoxyglucose)	Staging of breast cancer	Expensive Limited access Low sensitivity
Scintimammography	Nuclear imaging technique usually technetium-99m isonitrile (Sestamibi)	May be more sensitive for detection of certain histological types of breast cancer e.g. lobular invasive carcinoma	Poor spatial resolution Expensive
Electrical impedance imaging	Technique involving low-level bioelectric currents to map electrical impedance properties of the breast	No harmful radiation	
Infrared thermography	Measurement of heat emissions	No harmful radiation	Less sensitive and specific than mammography
Transillumination (near-infrared spectroscopy, light scanning)	Technique for scanning the breast with red or near-infrared light and recording the light image on infrared-sensitive film or with a television camera	No harmful radiation	Less sensitive and specific than mammography
Laser transillumination	Refinement of the above with extremely short laser pulses and time-resolved detection	Better resolution than infra-red transillumination	Still experimental

compression and positioning are unchanged.

The digital receptor consists of a matrix of square picture elements (pixels), usually measuring 50–100 mm. In most current receptors, the signal is created in a two-step procedure. In the first step, the X-ray energy is converted to light in a structure that is similar to a conventional intensifying screen. In a second step, the light is converted to an electrical signal, which is digitalized. In other detectors, the light step is omitted, and the X-rays interact directly with the detector, creating electrical charges that are digitalized. Still other detectors count the X-ray photos directly.

The signal value of pixels is usually digitalized into 12–16 bytes, which creates a grey scale of 4096–65 536 levels. This wide dynamic range is a major

advantage over conventional techniques and represents the basis for higher contrast resolution and various image processing and display techniques. One practical advantage is that areas that are very dark or bright on screen–film mammograms can be displayed to better advantage. The images can be printed on paper, but, to obtain full advantage of the technique, a monitor is required.

For high-volume screening, special work stations have been developed in order to handle large data sets and to display the images in a rational, customized way. One digital image may comprise 8–32 megabytes. Other advantages of digital mammography are related to storage and communication. Digital images can be transmitted electronically for centralized reading or consultation.

Digital mammography has the potential to provide images with lower doses of radiation than screen–film mammography. This may apply even more to the photon counting detectors, but no data have so far been published to support this contention. The cost of acquiring a complete digital system is several times that of a conventional system; however, in a high-volume screening setting, this cost may be offset by more rational working procedures and the elimination of fibre and developing chemicals.

Full-field digital mammography has been evaluated as a screening modality in one study (Lewin *et al*., 2001), which showed it to have similar sensitivity to screen–film mammography and greater specificity.

Computer-aided detection can be incorporated into the work station and the results of the computer analysis added onto the image, thereby assisting the radiologist in detecting suspect lesions. Computer-aided detection has been assessed in several studies (te Brake *et al*., 1998; Warren Burhenne *et al*., 2000; Birdwell *et al*., 2001; Freer & Ulissey, 2001), which suggest an incremental value in terms of sensitivity. The evidence on specificity is conflicting. Some data suggest that computer-aided detection could replace a second reader (Warren Burhenne *et al*., 2000).

Ultrasonography

Ultrasound images are produced from reflected high-frequency sound waves, without exposure to ionizing radiation. The technique is currently used mainly as an adjunct to mammography to characterize suspected lesions further and to guide needle biopsy. Breast ultrasound examination of asymptomatic women has some potential limitations:

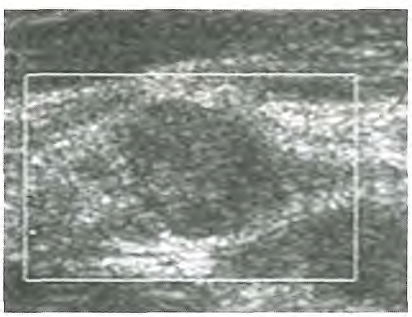
- The sensitivity is highly dependent on the operator (Teh & Wilson, 1998).
- The field of view is limited to a few centimeters, which makes a full breast examination difficult and time-consuming (Nass *et al*., 2001) as well as more expensive than mammography.
- Creation of hard copies of ultrasound images is costly; recording the entire examination is impractical (Teh & Wilson, 1998).
- It is relatively ineffective for detecting microcalcifications (Nass *et al.,* 2001; National Alliance of Breast Cancer Organizations, 2001)

In a type of ultrasonography called elastography, the firmness of tissue is imaged. Softer tissues, such as fat, appear brighter on the images than do firmer tissues—including tumours. The technique involves combining two ultrasound images of the same tissue: a compressed view and an uncompressed view. While elastography may become a useful adjunct for distinguishing between benign and malignant lesions, its potential has not yet been clarified.

Ultrasonography has been assessed in several studies, primarily in women who had mammographically dense breasts of who were at high risk for breast cancer (Kolb *et al.*, 1998; Buchberger *et al.*, 2000; Warner *et al.,* 2001). The results suggest that ultrasonography may increase the sensitivity of screening if used as an adjunct to mammography for mammographically dense breasts. Combined testing is likely to decrease specificity. It is not clear whether ultrasonography on its own is better than mammography in an unselected population.

Magnetic resonance imaging

Magnetic resonance imaging (MRI) involves use of rapidly fluctuating, high magnetic fields to excite the protons of the hydrogen atoms within the water molecule. Weak electromagnetic signals produced within the body are detected by antenna coils and used to generate planar and three-dimensional images of internal structures. Planar images can be created from virtually any viewing angle, at a resolution of approximately 1 mm^3, and without ionizing radiation. The magnetic field presents minimal hazards.



Ultrasonography image

Magnetic resonance imaging

As used in breast cancer detection, MRI must be performed as a dynamic study of contrast enhancement of breast tissue after intravenous administration of contrast medium (Heywang-Köbrunner *et al.*, 1988; Kaiser, 1989). The pathophysiological basis for the contrast enhancement of breast cancer is the presence of newly formed vascular structures, which have increased permeability and, furthermore, increased extravascular space. In typical cases, the rapid contrast enhancement is followed by an immediate decrease (so called wash-out). Benign lesions tend to enhance more slowly. Most invasive cancers show the typical enhancement pattern, but there are exceptions, especially lobular invasive carcinoma, which may resemble a benign lesion. This is also true for many uninvasive carcinomas. In contrast, some benign lesions, such as some fibroadenomas and papillomas, show rapid enhancement, similar to carcinomas.

MRI is considerably more expensive than ultrasound and mammography. Other drawbacks to MRI include the following:

- It is more time-consuming than mammography; an MRI examination takes approximately 30 min to complete, during which time the woman must remain motionless within the cramped quarters of the MRI device.
- The restricted MRI machinery conditions might discourage women with claustrophobic tendencies from undergoing the examination.
- The current technique requires an intravenous infusion of a contrast agent.
- The image obtained in MRI is affected by the phase of a woman's menstrual cycle. It is best done during the second or third week of the cycle, to minimize hormonal effects (Stoutjesdijk *et al.*, 2001).

There are several well-established indications for use of MRI in the clinical setting, such as in investigation of possible multifocality or multicentricity of breast cancer in patients who cannot be fully evaluated with conventional techniques. Other indications are breast prostheses or extensive scarring, which may be difficult to evaluate with conventional techniques, and cases of axillary metastases and an unknown primary (Heywang-Köbrunner *et al.*, 1988).

MRI has not been evaluated as a screening modality in unselected populations. Four studies on the sensitivity of MRI in high-risk women have been published within the past 2 years and are summarized in Table 7. The combined studies covered fewer than 40 cancers. The results suggest that MRI is more sensitive than mammography but may be less specific. A study is under way on a larger number of women (UK MRI Breast Cancer Screening Advisory Group, 2000).

Positron emission tomography

Positron emission tomography (PET) scans create computerized cross-sectional images of metabolic changes within a tissue. A radiolabelled tracer (usually a glucose analogue) is used to highlight differences in metabolic activity. The usefulness of PET in screening for breast cancer has not been demonstrated. Small studies have indicated

Table 7. Sensitivity and specificity of magnetic resonance imaging

Reference	Population	Age	No. of invasive cancers	Total sample size	Sensitivity (%)	Specificity (%) (% requiring biopsy)
Warner *et al.* (2001)	High risk (*BRCA* or several family members)	Mean, 43 Range 26–59	6	196	M, 33 CBE, 33 Ultrasound, 60 MRI, 100	M, 99.5 CBE, 99.5 Ultrasound, 93 MRI, 91
Stoutjesdijik *et al.* (2001)	*BRCA* lifetime risk > 15%	21–71	13	262 exams on 179 women	M, 42 MRI, 100	M, 96 MRI, 93
Tilanus-Linthorst *et al.* (2000a,b)	High risk (> 25%) and > 50% breast density	Mean, 41.5	3	109		MRI, 95
Kuhl *et al.* (2000)	High familial risk	Mean, 39 Range, 18–65	9	105	M, 33 Ultrasound, 33 MRI, 100	M, 93 Ultrasound, 80 MRI, 95

M, mammography; CBE, clinical breast examination; MRI, magnetic resonance imaging

fairly good sensitivity and more limited specificity, with poor sensitivity for tumours smaller than 1 cm (Nass *et al.*, 2001). PET might be useful in screening women with implants, scarring or dense breast tissue (Cole & Coleman, 1999; National Cancer Institute, 2001a). In addition, some lesions are seen on PET scans but not on mammograms, which makes biopsy difficult. Finally, the technique is costly, and PET scanners are relatively scarce because they must be located near particle accelerators that produce the short-lived radioisotopes used as tracers (National Cancer Institute, 2001a).

PET scanning is also time-consuming. After receiving the radioactive tracer, the woman must lie still for about 45 min while the tracer circulates, and the scanning takes another 45 min (National Cancer Institute, 2001a).

The National Cancer Institute (2001a) is sponsoring a clinical trial to evaluate PET and other imaging techniques in women with a diagnosis of breast cancer, but there is little likelihood of any use for PET in breast cancer screening in the near future.

PET has been evaluated as a screening tool in only one study of consecutive screenees (Yasuda *et al.*, 2000). There were only five breast cancers, of which one was detectable with PET only.

Scintimammography

In scintimammography—also called mammoscintigraphy—a radioactive tracer is introduced into the body and may accumulate at higher levels in tumour and some other tissues. A camera that detects γ-rays is then used to produce images. Newer cameras specifically designed for breast imaging are being evaluated clinically. The images can be two- or three-dimensional. One radioactive tracer (technetium-99m) has been approved by the Food and Drug Administration, and others are being studied (Nass *et al.*, 2001).

The current role of scintimammography is as an adjunct to mammography to identify metastatic cells distant to the breast and to localize tumours. The technique does not appear to be affected by implants, scarring or dense breast tissue. The health risks are minimal and similar to those from mammography, although the entire body is exposed to radiation. Scintimammography is more expensive than mammography or ultrasound, but less expensive than MRI or PET.

Electrical impedance imaging

Some cancerous tissue may conduct electricity much better than normal tissue does. Electrical impedance scanning is done with a hand-held probe connected to an electrode patch placed on a woman's arm. The probe measures the current passing through the skin covering the breast, and this information is used to reconstruct parametric images of the breast (National Cancer Institute, 2001a).

Electrical impedance imaging is painless and requires no exposure to ionizing radiation (Nass *et al.*, 2001). The technique may give false-positive results because of problems such as poor contact of the device on the skin, air bubbles and superficial skin lesions. The images reflect the superficial tissues, limited to about 35 mm deep, and cancerous lesions directly behind the nipple were difficult to detect (Malich *et al.*, 2001). At present, electrical impedance may have promise as an adjunct, but the high false-positive rates and other limitations compromise its use as a primary screening tool.

Other techniques

Radioactive antibodies

This technique involves radiolabelling antibodies to proteins that are selectively produced by cancer cells. Some have shown promise, but there have been no large-scale studies to determine a role for this technique in screening.

Infrared thermography

Changes in blood flow cause temperature changes, and some breast tumours can raise skin temperature, which can be detected by thermography (Sudharsan *et al.*, 1999). Infrared thermography was tested several decades ago, then essentially abandoned after the 1970s until recently. The technique is uninvasive and does not require compression of the breast or exposing women to radiation. The sensitivity and specificity of thermography are poor, and its application to screening is unlikely.

Near-infrared spectroscopy

Near-infrared techniques involve use of light sources at 700–900 nm to image the breast. Some differences between oxygenated and unoxygenated haemoglobin can be detected, with imaging of excessive oxygen consumption in some tumours. As with all imaging methods based on sources prone to problems such as scatter and diffraction, the sensitivity of this method for imaging deep lesions will remain limited.

Electrical potential measurement

As rapid cell proliferation disrupts the tissue's normal polarization, tools that measure electrical potential might allow identification of this disruption. Trials of the use of this technique in diagnosis, rather than screening, showed a specificity of only 55–60% (Fukuda *et al.*, 1996; Cuzick *et al.*, 1998).

Electronic palpation

This technique, also called 'tactile imaging', is essentially an objective method for specifying the parameters of a clinical examination. A company in Massachusetts (USA) is seeking approval of their hand-held device containing a group of sensors which is pressed against the breast and moved around to image the tissue. It has undergone only limited evaluation (Wellman *et al.*, 2001) and has not been assessed for screening.

Other techniques at an early stage of development

Other techniques—including magnetic resonance spectroscopy, magnetomammography, Hall effect imaging, thermoacoustic computed tomography, microwaves and three-dimensional interactive visualization—are in early stages of development.

Conclusions

None of the tests evaluated showed sufficient accuracy to support their use in general screening. However, the conduct and reporting of the studies were limited, and the populations were generally too small for adequate precision in critical measures, such as test sensitivity. Future studies should have adequate sample sizes, for example as is being done in a trial of digital mammographic imaging screening, which aims to enroll 49 500 asymptomatic women presenting for screening (http://www.acrin.org/protocols/6652/-6652abstract.html). Studies should conform to high standards of conduct and reporting (http://www.consort-statement.org/stard-statement.htm).

The design of cross-sectional studies to assess the accuracy of new techniques depends on how they are to be used. If a new technique is to replace an old one, the assessments (e.g. reading of images) should be performed independently and with similar information (e.g. clinical history) available to the readers of both tests. As the objective is usually to compare the accuracy of the tests under set conditions, procedures to deal with reader inaccuracy, e.g. selection of well-trained, experienced readers or the number of readers, should be similar for the two tests.

A new technique might be meant to complement an older one. For example, in a study by Lewin *et al*. (2001), the sensitivity of conventional mammography was 63% and that of full-field digital mammography was 60%. As the two modalities detected different cancers, however, doing both tests increased the sensitivity of mammography by 26%.

Larger, better studies of new techniques should be started soon after their introduction. As new techniques often change rapidly, it might be argued that evaluation should be left until the new technique has become 'stable'. Unfortunately, evaluations are often problematic and a technique may come into common use before the evaluation is finalized. It is therefore wise to start evaluation early, using the technique in order to assess how changes and developments can be incorporated (Lilford *et al*., 2000).

Clinical breast examination

Clinical breast examination long pre-dates imaging for evaluating mammary health and disease. While it depends on the eyes and fingers and subjective assessment of any abnormality found, it may still have a place in modern breast cancer screening programmes.

Technique

Procedure

No one technique for screening breasts for cancer has been shown to be better than any other in comparative studies against an assumed 'gold standard' or combination of methods. A systematic technique described by Pennypacker and Pilgrim (1993) was developed after extensive research with silicone breast models. This system illustrates the rigour that may be required to maximize the accuracy of clinical breast examination.

In the protocol of the Canadian National Breast Screening trials, described in Chapter 1, a 'spoke of the wheel' search pattern was used, with no explicit recognition of three levels of palpation (Bassett, 1985). In other respects, the technique was similar to that described by Pennypacker and Pilgrim (1993). Use of a vertical strip pattern was subsequently shown to result in more complete coverage of breast tissue than either a 'spoke of the wheel' pattern or a search in concentric circles (Saunders *et al.*, 1986).

The duration of a clinical breast examination depends on the skill of the examiner, the size and lumpiness of the breast and how many components of the examination are included. Visual examination is often cursory or omitted; applying three levels of pressure at each site is relatively uncommon. A study of periodic health examinations in an ambulatory care setting showed that the average duration of a complete clinical examination of both breasts and counselling on self-examination was 1.8 min (Kahn & Goldberg, 1984). Pennypacker *et al.* (1999) suggest a minimum of 5 min per breast for an experienced examiner using their programme.

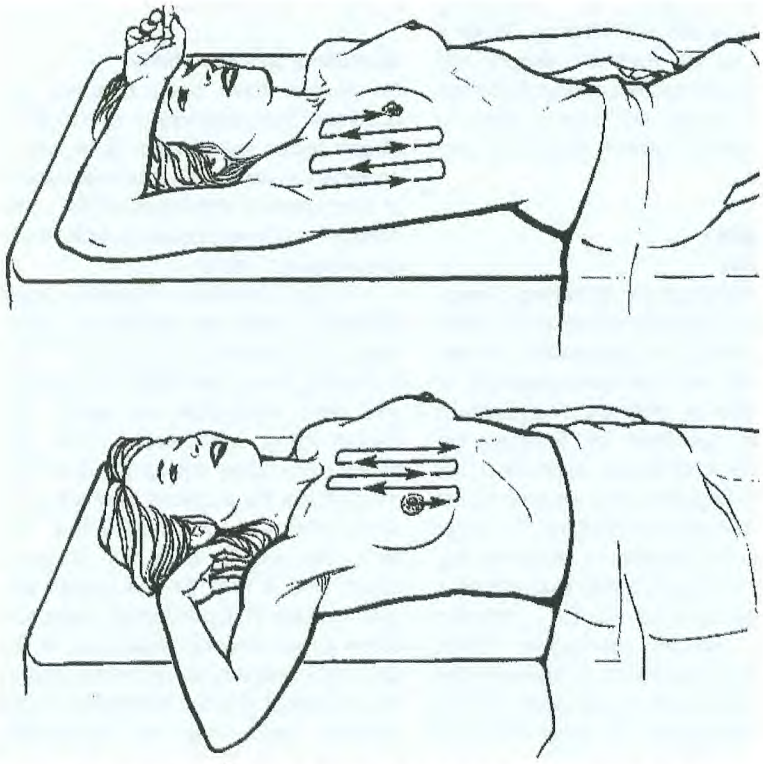
Sensitivity and specificity

No studies have been reported that document the sensitivity or specificity of clinical breast examination done fully in accordance with the recommendations of Pennypacker and Pilgrim (1993). The studies mentioned below include those described in Chapter 1.

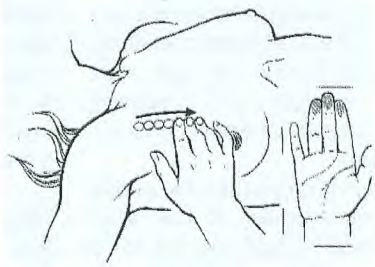
In the Canadian National Breast Screening trials, the sensitivity, specificity and predictive values of a first screening were estimated for women who were randomized to receive only clinical breast examination. Three estimates were made: one for the examiner, a second for the surgeons involved in the study (who saw only participants who were deemed to have an abnormal result) and a third for the overall programme (which depended on implementation by community physicians of the diagnostic procedures recommended by the surgeons). For the examiners, a true positive result was an abnormality reported at the first screen by clinical breast examination, during the 12-month interval after the first screen or at the

Clinical breast examination

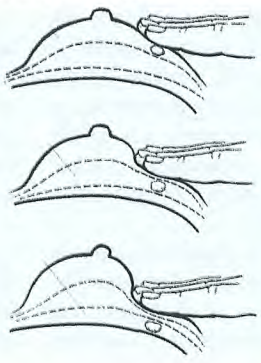
- visual examination of the woman in three different standing positions: arms relaxed at her sides; hands pressed firmly on her waist and leaning forward; and arms over her head. The examiner seeks subtle asymmetries in the appearance of the breasts;
- palpation of the supraclavicular and axillary nodes with the woman seated and re-palpation of the axillary nodes with the woman supine;
- vertical-strip search of the breasts over an area extending from the mid-axillary line to the mid-sternum and from above the sub-costal margin (fifth rib) to the clavicle, including palpation of the nipple and areola;
- application in this search of three levels of pressure, superficial, medium and deep, at each palpation site. Palpation is done with the finger pads of the three middle fingers, and pressure is applied with circular motions at each site. For the lateral half of the breast, the torso is rotated in the medial direction; for the medial half of the breast, the torso is rotated laterally in order to spread out the breast tissue;
- when an abnormality is detected, the corresponding area of the other breast is examined. If the finding is not bilateral, further investigation is required.



Position of patient and direction of palpation for clinical breast examination
The figure shows the lateral portion of the breast and, bottom, the medial portion of the breast. Arrows indicate vertical strip pattern of examination.



Palpation technique
Pads of the index, third, and fourth fingers (inset) make small circular motions.



Levels of pressure for palpation of breast tissue shown in a cross-sectional view of the right breast

second screen. At the first screen, they achieved 71% sensitivity, 84% specificity and a positive predictive value of 1.5% in women aged 40–49 years at entry and 83%, 88% and 3%, respectively, in women aged 50–59 at entry. With repeated screening of the older women, the sensitivity of screening by the examiners decreased and the specificity increased. These estimates represent detection, in that cancers detected up to 12 months after screening were taken into account (Baines *et al*., 1989).

Barton *et al.* (1999) pooled data from the study of the Health Insurance Plan of New York, USA, the trial in the United Kingdom, the Canadian national breast screening trials, the Breast Cancer Detection Demonstration Project in the USA and a study in West London, United Kingdom. Sensitivity was defined as the number of women with cancer detected by clinical breast examination divided by the sum of cancers detected at screening plus cancers detected within 12 months of screening. This yielded an overall estimate for the sensitivity of clinical breast examination of 54% (95% confidence interval [CI], 48–60%) and a specificity of 94% (95% CI, 90–97%). These estimates are remarkably close to similarly derived values reported by Bobo *et al.* (2000) in an analysis of 752 081 clinical breast examinations performed in the USA in the National Breast and Cervical Cancer Early Detection Program between 1995 and 1998. They found an overall sensitivity of 59%, a specificity of 93% and a predictive value of 4%; however, their ascertainment of interval cancers was limited to women who had undergone more than one screening. Ohuchi *et al*. (1995) reported 85% sensitivity and 97% specificity for clinical breast examination in Miyagi Prefecture, Japan, and Morimoto *et al*. (1997) reported 73% sensitivity in Tokushima Prefecture (see Chapter 4 for a description of these studies).

Studies with silicone models have shown that the sensitivity of clinical breast examination increases with increasing lump size and with increasing firmness of lumps, while greater depth is associated with decreased sensitivity (McDermott *et al.,* 1996). Barton *et al.* (1999) reported a sensitivity of 14% for 3-mm lumps and 79% for 1-cm lumps. Others have shown that the duration of the examination is positively correlated with sensitivity (Fletcher *et al.,* 1985; Campbell *et al.*, 1991). Comparisons of the results obtained with models with pre- and post-menopausal characteristics indicated that the sensitivity of clinical breast examination increases with age (McDermott *et al*., 1996; Barton *et al.*, 1999).

Seven screening studies afford the opportunity to document whether offering clinical breast examination and mammography simultaneously in a screening programme increases sensitivity. There is clear variation among the studies in the mode of detection (by mammography alone, by clinical breast examination alone or by combined mammography and clinical breast examination) for cancers detected at screening. Some of the variation is due to study design; for example, in the Edinburgh trial, women were screened annually by clinical breast examination alone alternately with combined breast examination and mammography. Another explanation for the variation would lie with the adequacy of the protocol for clinical breast examination and its monitoring (Baines, 1992a). Only the Canadian national breast screening trials incorporated a protocol for clinical breast examination with evaluation and feedback (Baines *et al.,* 1989). As shown in Table 8, the detection rate with mammography alone is increasing over time and that with clinical breast examination decreasing. It has been shown in experimental situations that sensitivity decreases with lump size and with other factors such as duration of examination (Fletcher *et al*., 1985).

More recently, Bobo *et al.* (2000) analysed the results of the National Breast and Cervical Cancer Early Detection Program in the United States for 1995–98 and found that at least 5.1% of breast cancers were not detected by mammography but were detected by clinical breast examination alone. A further 11% for which the mammography results were not reported were detected by breast examination. The procedure for clinical breast examination was not standardized.

Training

Over the past two decades, training in clinical breast examination has been conducted increasingly with manufactured silicone models of the breast (Pennypacker & Pilgrim, 1993). Models can be designed to display the characteristics of pre- or postmenopausal breasts and lumps of varying size, depth and firmness. The models are placed horizontally for palpation, corresponding to a patient in the supine position. They have proved acceptable to health professionals for evaluation of their competence in clinical breast examination (Fletcher *et al.*, 1985). Furthermore, the rate of lump detection on such models correlated with that in in actual breast tissue (Hall *et al*., 1980).

A randomized controlled trial was conducted with silicone breast models to evaluate the effect of training on the accuracy of lump detection by physicians and nurses (Campbell *et al.*, 1991). The mean sensitivity increased in the intervention group from 57 to 63% but decreased in the control group from 57 to 56%. Mean specificity declined in the experimental group from 56 to 41% and increased in the control group from 56 to 68%, indicating that the number of false-positive results increased with training. The recommended technique had six components: use of pads of the middle three fingers, circular motion, vertical-strip search pattern, three levels of pressure and total coverage. Four months after training, 80% of the intervention group were still using the correct tech-

Table 8. Rates of cancer detection in programmes with combined mammography and clinical breast examination

Study (date of start)[a]	Age (years)	Frequency	Percentage detected at screening		
			Mammography only	Clinical breast examination only	Both
Health Insurance Plan, USA (1963)	40–49	Annual	19	61	19
Breast Cancer Detection Demonstration Project, USA (1972)	40–49	Annual	45	8	46
Canadian National Breast Screening Study-1 (1980)	40–49	Annual	40	23	36
Health Insurance Plan, USA (1963)	50–59	Annual	41	41	18
Breast Cancer Detection Demonstration Project USA (1972)	50–59	Annual	47	7	45
Utrecht, Netherlands (1975)	50–64	Variable	56	10	35
Canadian National Breast Screening Study-2 (1980)	50–59	Annual	53	12	35

Adapted from Baines and Miller (1997)

[a] For descriptions of these studies, see Chapters 1 and 4.

nique. Training thus achieved increased sensitivity at the cost of decreased specificity.

Campbell *et al.* (1994) compared standardized with unstandardized training of medical students in clinical breast examination. The group with standardized teaching achieved improved accuracy of lump detection and technique, accompanied by decreased specificity. Interestingly, women with no previous medical experience were found to be able, after training, to teach a standardized technique as well as medical personnel. Another study showed that practice on silicone breast models and volunteers in medical schools was more effective than lectures alone for teaching clinical breast examination (Pilgrim *et al.*, 1993).

A recent controlled study (Lane *et al.*, 2001) showed that continuing medical education for community-based primary care physicians effectively improved their communication and counselling skills with respect to screening, clinical breast examination and administrative strategies to enhance routine screening. Of the two continuing medical education strategies used—face-to-face teaching and self-study—the former was more effective. However, use of such strategies on a national basis would be difficult. An alternative to physicians is nurse-examiners, as shown in the Canadian national breast screening trials (Miller *et al.*, 1991a). Acceptance of screening with clinical breast examination was increased by sending an invitation to women who were at high risk for breast cancer, although, again, nationwide implementation of such a strategy is unlikely (Richardson *et al.*, 1996).

Maintainance of standards

No programme for evaluating clinical breast examination in large screening programmes has been published. However, the 15 centres of the Canadian national breast screening trials were provided with a protocol. Furthermore, at each centre, the examiners benefitted from regular feedback from study surgeons when participants with abnormal results from either clinical breast examination or mammography attended the review clinic. Annual (semi-annual when required) site visits allowed the deputy-director of the trial to observe examiner–participant interactions and to identify violations of the protocol for clinical breast examination; however, the consequences of these interventions were not evaluated (Baines *et al.*, 1989).

Detection of lumps in silicone breast models may be a useful way of

evaluating performance of clinical breast examination and could be used to identify practical standards and to monitor and improve performance against them. This technique has been used in assessing physicians' performance in breast examination and in evaluating the effectiveness of standardized training in breast examination (Fletcher *et al.*, 1985).

Costs and potential harms

The costs of clinical breast examination include the cost of training, the cost of delivery, the cost of enhancing delivery and acceptance and the cost of diagnostic follow-up when abnormalities are found. Substantial costs are associated with training for the method of Pennypacker and Pilgrim (1993). The cost of delivery depends on the professional status of the examiner, being highest for physicians and lowest for 'supporting personnel'. The cost of enhancing implementation and acceptance of clinical breast examination depends on the intervention. Diagnostic follow-up may include fine-needle aspiration, fine-needle aspiration biopsy, core-needle biopsy, open biopsy, ultrasonography and diagnostic mammography. The procedures implemented, their frequency and the associated costs vary.

There are also direct and indirect costs to the women being examined. The palpation procedure itself is associated with no physical hazards other than minor discomfort. However, Elmore *et al.* (1998) calculated the 10-year risk associated with false-positive results in 10 905 clinical breast examinations among 2400 women in the Boston area (USA). After 10 annual examinations, the estimated cumulative risk for a false-positive result was 22%, and these all required further diagnostic follow-up, with the attendant expenses and anxiety. It is important that clinically suspect masses be evaluated even if a mammogram is normal (Pruthi, 2001).

Other issues

One survey of 2800 participants in the Canadian national breast screening trials (82% response rate) revealed that women found clinical breast examination more acceptable than mammography, in that there was less associated discomfort. Furthermore, only 20% expressed a preference for clinical breast examination performed by physicians rather than nurses. Attendance at screening is enhanced by convenient site location, punctual appointments and courteous and supportive staff (Baines *et al.*, 1990).

Breast self-examination

Systematic breast self-examination has been recommended for almost 70 years (Adair, 1933), in the absence of compelling evidence of its efficacy. Initially, self-examination was justified because a substantial proportion of breast cancers were discovered by women themselves (Hislop *et al.*, 1984; Joensuu *et al.*, 1992); more recently, the practice has been seen to empower women, allowing them to take responsibility for their own health.

Technique

Procedure

Mamon and Zapka (1983) outlined one of many techniques that have been described for breast self-examination. Eight steps were to be performed lying down, first for the left and then for the right breast. They included: placing one hand behind the head and a prop under the shoulder; using the hand opposite to the breast being examined; pressing with the finger pads; covering the entire breast area; squeezing the nipple; examining the armpit; and using a circular or 'ladder' search pattern. Seven steps were outlined for a similar process in the upright position, including squeezing the nipple. Finally, there were four steps for conducting a visual examination in front of a mirror. Expecting women to comply with 34 steps may be unrealistic, and such complexity may lead to lack of confidence (Eggertsen *et al.*, 1983; Baines, 1988).

Thus, Baines (1992b) argued for a simplified technique based on the paedagogical principle that 'less is more' in terms of remembering what has been taught (Russell *et al.*, 1984). Baines (1992b) also urged that the nipple squeeze, likely to be a deterrent to self-examination, be eliminated, because it is a spontaneous discharge, not a manually expressed discharge, that is pathognomic (Pilnik & Leis, 1978; Haagensen *et al.*, 1981). Another disincentive to women may be the requirement that the practice be done in two positions, lying down and standing up. This led to a proposal that women with large breasts might choose to do self-examination lying down, while women with smaller breasts might prefer to do it while standing (Baines, 1992b). The proposal is consistent with a 21-step procedure in the upright position described by others (Carter *et al.*, 1985).

The crucial components of breast self-examination appear to be visual examination and palpation of the entire breast with the finger pads in an effective search pattern. Hislop *et al.* (1984) showed that visual inspection was associated with smaller tumours and that careful palpation was associated with the absence of palpable nodes. Harvey *et al.* (1997), in a case–control study, identified three important components for the efficacy of breast self-examination, namely visual examination, palpation with the finger pads and using the three middle fingers. The proposed search patterns are of three types: concentric circles, radial spokes and vertical strips. The last has been shown to provide the of breast tissue (Saunders *et al.*, 1986; Murali & Crabtree, 1992). Frequency of breast self-examination has been reported not to be a proxy for competence in practising it (Howe, 1980; Assaf *et al.*, 1983; Fletcher *et al.*, 1989; Janz *et*

Breast self-examination

- Is any visual examination done ?
- Is most of the breast examined ?
- Are the armpits examined ?
- Is there a systematic search pattern?
- Are three fingers used?
- Are finger pads used ?
- Is a rotatory palpation applied?
- Is breast self-examination performed 12 times a year?

al., 1989), although contrary conclusions were drawn from the Canadian national breast screening trials (Baines & To, 1990).

Sensitivity and specificity

Many published measures of the sensitivity of breast self-examination were based on detection of lumps in a silicone model of a breast or by a health professional in vivo. Assaf *et al.* (1983) concluded that the number of lumps that women detect in a silicone model is positively related to the number of components of breast self-examination that are performed correctly. Another study showed that increased accuracy (sensitivity) of detection of lumps in breast tissue, increased duration of examination and increased confidence were associated with training; however, training also increased the rate of false-positive findings and thus diminished specificity (Hall *et al.*, 1980).

In the Canadian national breast screening trials, a proxy for the sensitivity of breast self-examination was estimated for 18 242 women who received five screening examinations (Baines, 1989). A report of a positive finding from breast self-examination was considered a 'true' positive if it agreed with the subsequent findings of the examiner. On the basis of their self-examination scores, participants were divided into good, medium and poor performers. The scores improved over time. Higher scores were associated with higher sensitivity (never higher than 17%), and the positive predictive value improved from 39% at the third screen to 45% at the fifth. There was no difference between women who entered the programme when in their 40s and those who entered when in their 50s with regard to competence in breast self-examination (Baines *et al*., 1986c).

Training

A frequently used system for training in breast self-examination is the MammaCare programme, which includes approximately 45 min of instruction from a nurse. The programme stresses tactile skills (lump detection and discrimination) and examination techniques. Silicone breast models are used both during teaching and in private sessions at home (Pennypacker *et al.*, 1982).

A randomized controlled trial involving 300 women aged 40–68 was conducted to compare three methods for teaching breast self-examination: MammaCare, traditional instruction from a nurse and no instruction, half of each group being encouraged by their physicians to do self-examination (Fletcher *et al.*, 1990). The follow-up evaluation 1 year later was completed by 260 women. The group taught by MammaCare achieved more long-term improvement in lump detection in silicone models and in breast self-examination than those given traditional instruction or encouragement by a physician. Other investigators showed that female university students found significantly more lumps in breast models after MammaCare training than health professionals not taught with the MammaCare system, and the two groups had similar false-positive rates (Jacob *et al.*, 1994).

The MammaCare system is not often used for training in breast self-examination. Other approaches have been shown to be most effective when done on a one-to-one basis, even though one study showed that competence in breast self-examination can be improved and the frequency increased after one session (Dorsay *et al*., 1988). In a study in which women were randomized to one of four approaches to training in self-examination, individual instruction was more successful in terms of proficiency and frequency than group teaching, and individual teaching plus reminders was even more successful (Bennett *et al*., 1990). Coleman *et al*. (1991) found that individual instruction resulted in greater proficiency than did group teaching. Ferro *et al.* (1996) concluded that instruction in breast self-examination based on theoretical and practical discussions significantly improved the quality of examination when compared with instruction based only on mailed material.

In a cohort of almost 90 000 women in the Canadian national breast screening trials, the scores for breast self-examination improved over time when it was taught annually and was reinforced on an individual basis in the context of a clinical breast examination (Baines & To, 1990).

Maintenance of standards

Reinforcement

Reinforcement was shown to be necessary in order to maintain skills in breast self-examination in a research setting involving 29 women trained in MammaCare (Pinto, 1993), in a programme involving almost 90 000 women in the Canadian national breast screening trials (Baines & To, 1990) and in four communities in Vermont, USA (Worden *et al.*, 1990). Pinto (1993) showed that women whose skills were evaluated 2 months after MammaCare training and who had received re-training as needed had greater proficiency at 4 months and 1 year follow-up than women who were not evaluated or re-trained at 2 months. The results of the Canadian national breast screening trials showed that annual evaluation and re-training consistently improved breast self-examination scores over time (Baines & To, 1990), and Worden *et al.* (1990), comparing four communities, concluded that maintenance measures improved competence in breast self-examination over and above that achieved with training alone. The maintenance measures were designed to overcome barriers to self-examination: forgetfulness, by prompts and rewards; decreased confidence, by supportive messages in the media; and anxiety, by more media messages. Such interventions are unlikely to be widely generalizable.

Thomas *et al*. (2002) also reported improved performance on silicone breast implants in terms of technique and lump detection after reinforcement.

Regular observation (with feedback) of all examiners in the Canadian national breast screening trials to evaluate their performance with respect to instruction and evaluation of breast self-examination (Baines, 1987) may also have enhanced instruction in this practice.

Performance indicators

A study with silicone breast models involving 126 women showed that three indicators were strongly associated with accurate detection of lumps: pressing firmly and deeply, examining all regions and adequate duration of examination (Haughey *et al.*, 1984). The indicators used in another study were frequency, knowledge about when to do breast self-examination, technique and number of lumps detected in a silicone model (Carter *et al.*, 1985). A more complex set of performance indicators was based on a combination of three scores: one for technique with four components, one for completeness based on nine components and one for lump detection based on the number of lumps detected in a silicone model (Dorsay *et al.*, 1988). Such an approach is useful in a research setting.

In contrast, Baines (1988) proposed eight indicators appropriate for evaluation in a clinical setting (see box). The weakness of these indicators is that they are equally weighted, and it is extremely unlikely that they are equivalent.

The performance indicators used by Celentano and Holtzman (1983) were also equally weighted. They concluded that most women do not do breast self-examination correctly and that their competence can be evaluated from a self-report. The indicators they used were the components described by Mamon and Zapka (1983), listed above. However, when 81 women were asked to report their usual breast self-examination practice and were assigned a score on the basis of the number of components mentioned, the score was not associated with performance on a silicone model, indicating that what women say they do is not a reliable indicator of performance (Newcomb *et al.*, 1995).

Researchers studying the MammaCare method developed a weighted scoring system for performance of breast self-examination that could be used in a clinical setting (Coleman & Pennypacker, 1991). The components, in descending order of weight, were: area examined, pressure used, motion while applying pressure, part of fingers used, search pattern, number of fingers used, number of motions and duration of examination.

Mechanisms for improving breast self-examination

Encouragement or instruction by a physician is related to the frequency of breast self-examination (Senie *et al.*, 1981; Bennett *et al.*, 1983; Celentano & Holtzman, 1983; Amsel *et al.*, 1984; Champion, 1987). However, achievement of both competence and adequate performance probably requires more than encouragement.

Cue enhancement was investigated by providing calendars with reminders and sending monthly reminders on postcards (Grady, 1984). These interventions were effective in achieving high rates of breast self-examination but only by menstruating women, and the frequency of practice declined after the experimental period. In contrast, distribution within the Canadian national breast screening trials of 1166 calendars on which women were asked to enter their findings from breast self-examination, analogous to the Finnish Mama Programme (Gästrin, 1981), had no effect on performance or the competence of breast self-examination when compared with that of 1027 women who did not receive the calendars (Baines *et al.*, 1988b). Craun and Deffenbacher (1987) evaluated the efficacy of three approaches to increasing the frequency of breast self-examination and found that sending women monthly reminders was successful, while educational and demonstration programmes were not.

A 12-month public education campaign aimed at 40% of the Australian population was conducted through the mass media, with the support of local doctors, to teach women how to practise breast self-examination (Hill *et al.*, 1982). Surveys of the general public, of patients in general practitioners' practices and of patients with newly diagnosed breast

cancer before and after the campaign showed that 13% more of the general public and 6% more of breast cancer patients reported practising breast self-examination than at baseline. Performance was self-reported, and the competence of practice was not evaluated.

Costs and potential harms

The costs associated with use of breast self-examination as a screening intervention are easy to conceptualize. The direct monetary costs include those for supportive health education and for training trainers, that to trainers in terms of professional time expended and that involved in evaluating the outcome. The indirect monetary costs include those for visits to health professionals triggered by findings at breast self-examination and any diagnostic and therapeutic procedures arising from such visits.

From the woman's perspective, the costs are the time it takes to acquire skill in breast self-examination, the associated monetary costs in terms of lost time from work and that of the instruction programme, the time it takes to do breast self-examination on a regular basis and the anxiety associated with lack of confidence or with problems in interpreting findings.

Only one well-designed study of the benefit of breast self-examination was identified (O'Malley, 1993). Benefit was defined as the increase in the number of women performing competent, frequent self-examination after training by nurses in MammaCare or traditional methods. Interestingly, the medical costs after teaching were not increased. Nevertheless, the costs associated with breast self-examination are considerable.

The potential for harm from the practice of breast self-examination resides in over-confidence, which might lead to delayed presentation with symptoms of cancer, false reassurance by health professionals when cancer is present and unnecessary investigation of benign lesions with subsequent morbidity and scarring. These harms may be most relevant to women under 30 who practise breast self-examination (Frank & Mai, 1985).

Other issues

A study based on self-administered questionnaires of women's attitudes to screening after participation in the Canadian national breast screening trials achieved an 82% response rate (Baines *et al.*, 1990). Analysis of 2299 questionnaires revealed a strong commitment to continuing breast self-examination. It also revealed that women found it difficult to do so, and almost 50% rated their competence in breast self-examination as only adequate or poor; only 7% considered it excellent. Self-reported impediments to breast self-examination were laziness, forgetfulness, being too busy and lack of confidence in both skills and interpretation. These attitudes are surprising, given that these women had annual instruction and reinforcement in breast self-examination. Janz *et al.* (1989) noted that, because breast self-examination is done in private, it excludes social approval and regular critical feedback. Also, breast self-examination does not alleviate symptoms or make women feel 'better' for doing it.

Some women practising breast self-examination may experience fear of cancer, pain and death (Moore, 1978). It has also been suggested that breast self-examination might arouse fear of mutilation and loss of desirability and be a threat to sexual identity (Bernay *et al.*, 1982). Whatever the factors are that influence women's practice of breast self-examination, it is clear that, after years of research and encouragement, compliance with breast self-examination is less than impressive.

Examination of the breast by the surgeon Teodorico Borgognoni (1275)

Given the date of the painting (1275), breast cancer has probably been common in nuns for many centuries.
The painting is from Leiden University.