Chapter 1

Breast Cancer and Screening

The world-wide burden of breast cancer

Of the 10 million new cases of invasive cancer world-wide each year in males and females combined, 10% arise in the breast, which makes it the second most common site of malignant neoplasms after the lung (Parkin, 2001). In 2000, breast cancer accounted for 22% of all new cancers in women, making it by far the most common cancer in females (Figure 1). In high-income countries, the proportion rises to 27%, more than twice as common as any other cancer in women. In 2000, cancer of the breast was also the commonest tumour among women in low-income regions, with 470 000 new cases per year, whereas invasive cervical cancer had been the leading cancer during the previous two decades. More than half of the 1.05 million cases occur in high-income countries in North America and western Europe and in Australia and New Zealand (Figure 2), where an average of 6% of women develop invasive breast cancer before the age of 75. Incidence rates of a similar magnitude are observed in Argentina and Uruguay. The risk for breast cancer is low in the low-income regions of sub-Saharan Africa and Southern and Eastern Asia, including Japan, where the probability of developing breast cancer by the age of 75 is one-third that of other high-income countries. The rates are intermediate elsewhere. Japan is the only affluent country where in 2000 the incidence rate was low.

Clear increases in the incidence of and mortality from breast cancer were

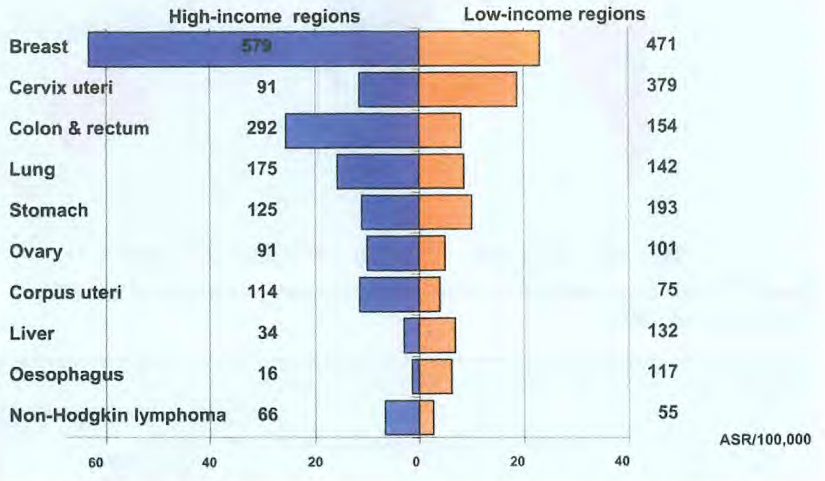


Figure 1 The 10 commonest sites of cancer in women world-wide, with incidence rates for 2000. World age-standardized rates per 100 000 population and total numbers of cases (thousands)
From Ferlay *et al.* (2001)

Breast cancer incidence

- **Breast cancers accounted for 22% of all cancers in women worldwide (1 million new cases) in 2000.**
- **The incidence of breast cancer in women in high-income countries in 2000 was at least twice that of any other cancer in women, and was similar to the incidence of cancer of the cervix in low-income countries (see Figure 1).**
- **More than half the breast cancers that occurred throughout the world in 2000 were estimated to have been in high-income countries (see Figure 2).**

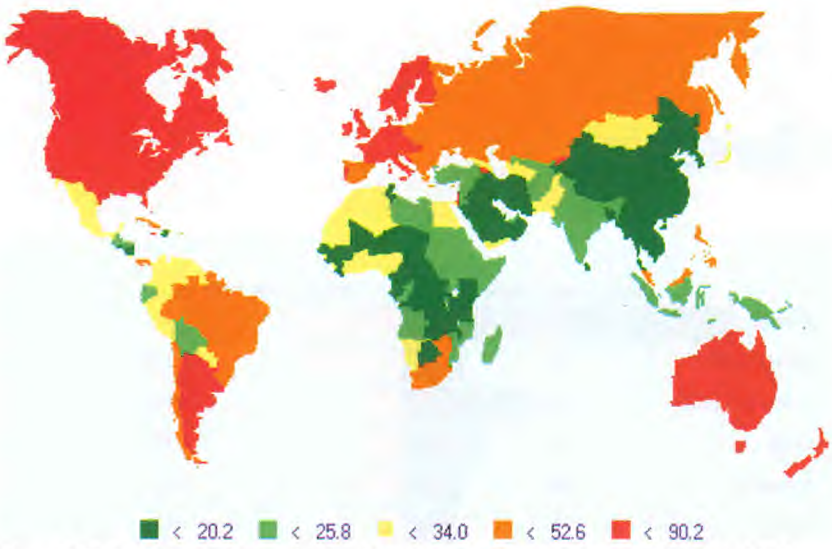


Figure 2 Estimated age-standardized incidence rates of breast cancer world-wide in 2000. From Ferlay *et al.* (2001).

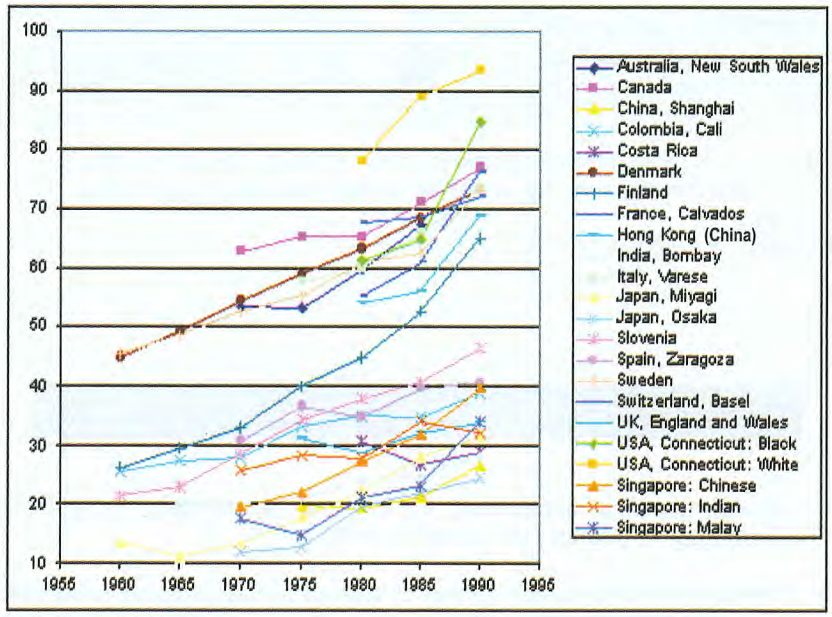


Figure 3 Trends in age-standardized incidence rates of breast cancer among women in selected populations.
From Doll *et al.* (1966, 1970); Waterhouse *et al.* (1976, 1982, 1987), Parkin *et al.* (1992, 1997); http://www.depdb.iarc.fr/who/menu.htm

observed until the early 1980s in both high- and low-income countries (Figure 3). The subsequent advent of mammography and improvements in prognosis in high-income countries altered the reported rates of both incidence and mortality, masking trends in the underlying risk for the disease. The risk continues to increase in eastern Europe and Latin America (Figure 3), as seen mainly from trends in mortality, and in some urban populations of Asia, as indicated by population-based incidence rates in, e.g., Japan, Singapore, Shanghai and Hong Kong (China) and Mumbai (India).

Around 1990, the incidence of breast cancer varied eightfold world-wide, indicating large differences in the distribution of the underlying causes (Parkin *et al.*, 1997). Studies of geographical variation, time trends and populations migrating from low- to high-risk areas (Geddes *et al.*, 1993; Ziegler *et al.*, 1993; Kliewer & Smith, 1995) suggest an important role of environmental factors in the etiology of the disease. Low parity, late age at first pregnancy, early menarche and late menopause are all factors that are consistently associated with an increased risk for breast cancer. Trends towards lower reproductive rates in western populations therefore explain part of the observed increase and may predict similar increases in populations where the reproduction rates are declining (Lopez-Carrillo *et al.*, 1997; dos Santos Silva & Beral, 1997; Gao *et al.*, 2000). As for most epithelial tumours, the risk for the disease increases steadily with age (Figure 4A).

Substantial improvements in survival have been recorded in western countries since the late 1970s (Adami *et al.*, 1989; Chu *et al.*, 1996; Quinn *et al.*, 1998), and an increasing number of women live with the consequences of the disease and its treatment. In the USA, survivors of breast cancer were estimated to constitute 1.5% of the female population (Hewitt *et al.*, 1999), which is about 10 times the annual incidence. The mortality rate, which had been increasing until the 1980s, levelled off or declined in several high-risk countries (Hermon & Beral, 1996; La Vecchia *et al.*, 1998; Howe *et al.*, 2001; Figures 4B, 4C, 4D). Despite these positive achievements,

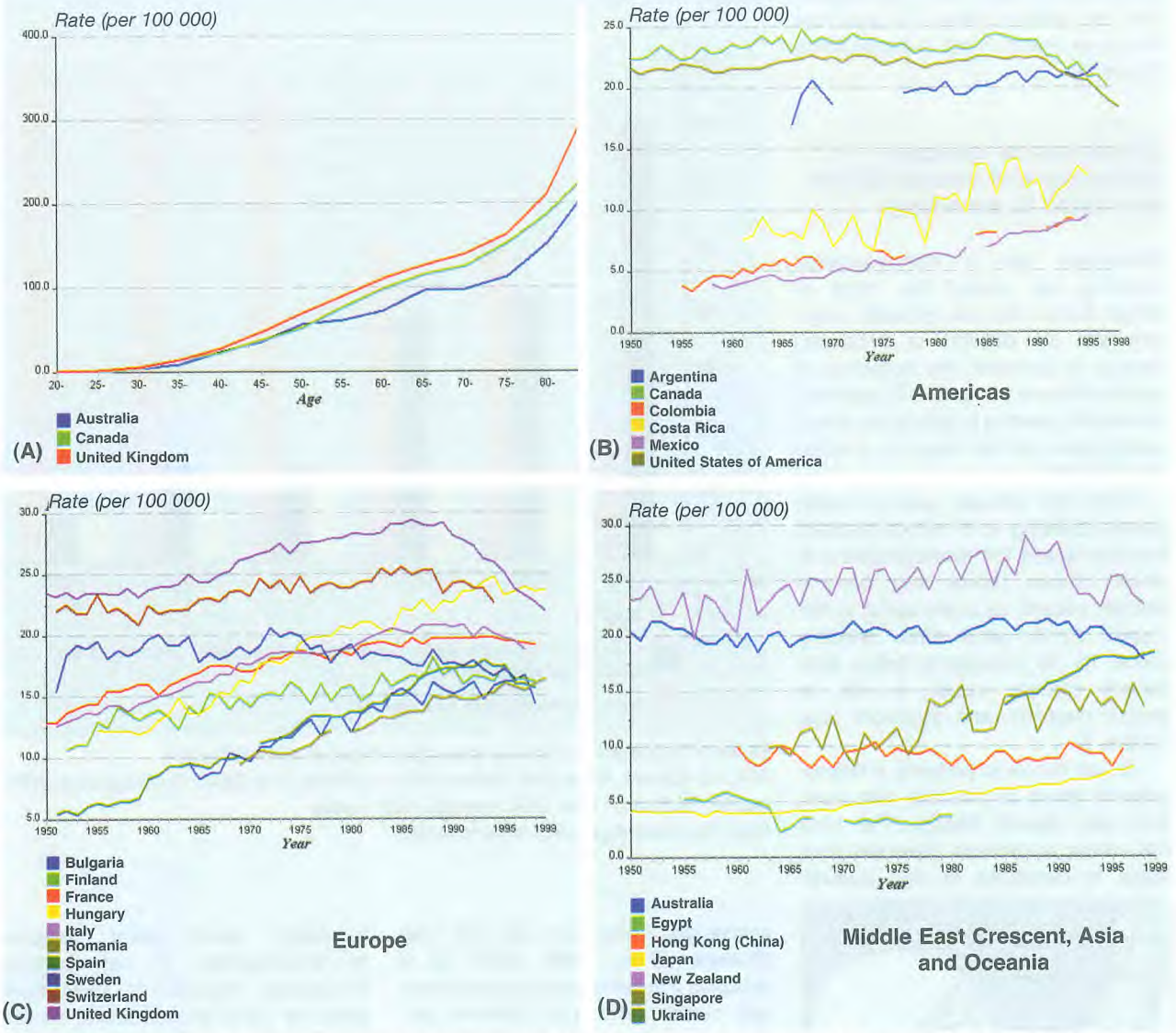


Figure 4 Age-specific mortality rates from breast cancer in women in Australia, Canada and the United Kingdom, 1990 (A). Mortality from breast cancer, time trends of age-standardized rates per 100 000 female population in the Americas (B), Europe (C), Middle East Crescent, Asia and Oceania (D)
From http://www-depdb.iarc.fr/who/menu.htm

breast cancer remains the malignancy that causes the most deaths from cancer among women in high-income countries. The only exceptions are Canada and the USA, where mortality from lung cancer is still rising and is characterized by a poor prognosis (http://www-depdb.iarc.fr/who/menu.htm).

Survival from breast cancer in low-income countries is generally poorer than that in high-income regions, reflecting late presentation of cases (Sankaranarayanan *et al.*, 1998). According to WHO, in 2000, noncommunicable diseases, including cancer, accounted for 75% of all deaths in the Americas, Europe and the Western Pacific region including China, half of all deaths in

southern and middle Asia, the Middle East and northern Africa and less than 25% of all deaths in sub-Saharan Africa (Figure 5).

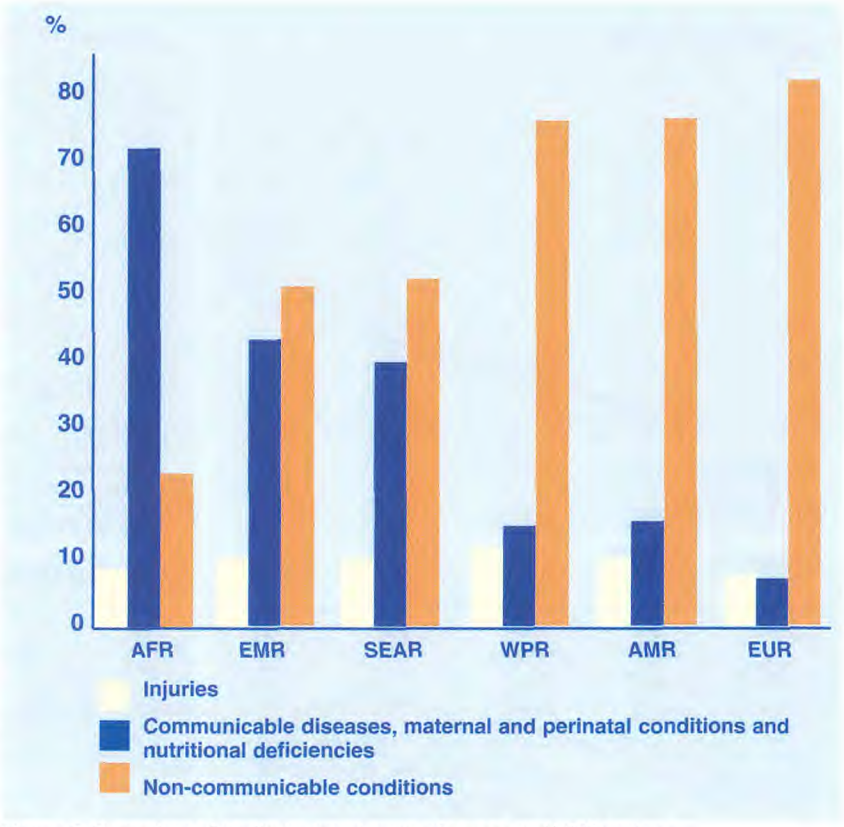


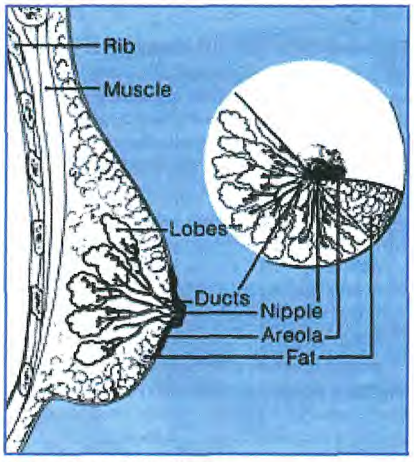
Figure 5 Proportions of deaths by broad group of causes and WHO region.
AFR, sub-Saharan Africa; EMR, northern Africa and Middle East; SEAR, South-East Asia; WPR, western Pacific and China; AMR, Americas; EUR, Europe
From http://www-depdb.iarc.fr/who/menu.htm

Breast cancer biology, pathology and natural history as related to screening

Widespread use of mammographic screening has altered the range of benign lesions that are removed surgically and the patterns of neoplastic disease. In particular, the frequency of ductal carcinoma *in situ* (DCIS) has risen dramatically, leading to debate on clinical management and the meaning of small in-situ lesions.

While the ultimate goal of breast cancer screening is to reduce mortality from the disease, the immediate goal is to detect cancers before they become clinically evident, as noted earlier in this chapter. At the same time, detecting cancer (or its precursors) before they present clinically raises a risk of excess diagnosis and treatment (see Chapter 5).

Breast cancer is probably a heterogeneous group of diseases with more than one natural history. The view that cancer progresses inexorably from atypia to carcinoma *in situ*, invasive cancer and metastasis may not hold (Buerger *et al.*, 1999, 2001). It is accepted that benign disease associated with ductal and lobular epithelial proliferation and hyperplasia, especially with atypia, confers an increased risk for developing breast cancer, and these lesions may form part of a spectrum of neoplastic breast disease or an interface between some benign and malignant breast conditions. However, these lesions may not be the explanation or the basis for development of all forms of breast cancer.



Normal breast structure

As screening mammography and other techniques, in contrast to symptoms, allow earlier detection of abnormalities, it has become increasingly important to know more about the risk for progression of the various lesions identified. Understanding the progression rates is crucial for answering questions relevant to screening programmes, including how such abnormalities should be treated and how intensively they should be sought.

This section reviews what is known about the progression of breast cancer at three points in the disease course: benign disease, in-situ cancer and invasive cancer.

Barriers to understanding of early cancers

At the outset, it is important to emphasize that there are two barriers to better understanding of progression. First, the nomenclature used for the microscopic appearance of lesions has been inconsistent, making comparisons across studies and time difficult. Secondly, there is concern about the reproducibility of pathological observations in early forms of breast cancer, some of which may reflect the problem of nomenclature and some the diagnostic threshold of a pathologist. In both cases, variation in diagnosis can confuse assessment of the risks for progression.

Most of the long-term data on progression of in-situ cancers and their precursors refer to lesions that presented clinically (e.g. a mass or nipple discharge) and not to those currently detected mammographically. Because disease detected at screening is asymptomatic and the tumours are generally smaller than those detected clinically, early detection could influence natural history. Thus, the progression rates reported here may be overestimates of the natural history of those detected by mammography.

Benign breast disease

Significance in breast screening

Many benign conditions can be seen mammographically, but those that lead to surgical biopsy are of particular concern. Woman may be recalled for assessment after primary mammographic screening because of benign disease or involutional changes, which can be seen mammographically as ill-defined masses (fibroadenoma and cysts), parenchymal deformity (radial scar, sclerosing adenosis) and calcification. A variety of benign calcified lesions are seen (Table 1; Spencer *et al.*, 1994). The commonest abnormalities leading to benign surgical biopsy are non-comedo-type suspect calcification (29%) (Figure 6), a poorly defined mass (21%), architectural distortion (19%) and a well-defined mass (18%) (Spencer *et al.*, 1994) (Figure 7) (American College of Radiology, 1995; Liberman *et al.*, 1997).

The positive predictive value for malignancy by type of mammographic abnormality is shown in Table 2 (Burrell *et al.*, 1996). The sensitivity of mammography in cancer detection must be high, but it is also important to achieve high diagnostic specificity to avoid morbidity associated with unnecessary surgical biopsy. The aim of assessments after screening should be both accurate diagnosis of breast cancer with prompt referral for treatment and accurate diagnosis of benign and involutional changes, if possible without surgical biopsy.

Association with an increased risk for breast cancer

Many studies have shown an increased risk for cancer among patients with usual epithelial hyperplasia, which is 1.5–2.0 times greater than that of a reference population, and a 2.5–4-fold increase in risk for patients with atypical ductal hyperplasia (Dupont & Page, 1985; Dupont *et al.*, 1993; Marshall *et al.*, 1997) (Figure 8). Atypical lobular hyperplasia increases the relative risk by four to five times (Page *et al.*, 1991; Marshall *et al.*, 1997). Other forms of benign breast disease, such as sclerosing adenosis, fibroadenoma and papillary apocrine change, appear not to alter the risk or to be associated with a 1.5–2-fold increase (Jensen *et al.*, 1989; Dupont *et al.*, 1994). The invasive cancers occurring after diagnosis of these types of epithelial proliferation occur at roughly equal frequency in the ipsilateral and contralateral breast. All these epithelial proliferative lesions may be found coincidentally in a lesion found as a result of breast screening.

Figure 6 Punctate calcification identified at mammographic screening. The resulting biopsy revealed benign stromal calcification

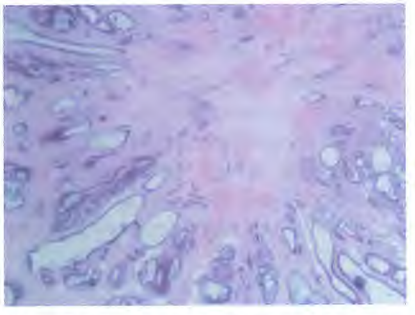


Figure 7 A benign radial scar which has a stellate configuration similar to some forms of breast carcinoma and can produce a parenchymal deformity mimicking carcinoma mammographically

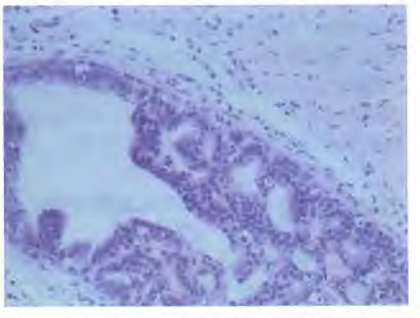


Figure 8 An example of atypical ductal hyperplasia with a single duct space, part of which contains uniform, small-cell epithelial proliferation

Table 1. Causes of indeterminate calcification diagnosed by core biopsy alone in 151 samples

Lesion	(%)
Fibrocystic change	33
Fibroadenoma	18
Stromal calcification	15
Fibroadenomatoid hyperplasia	15
Involutional change	11
Sclerosing adenosis	7
Duct ectasia	4
Apocrine change	4
Blunt duct adenosis	3
Mucocoele	2
Vascular	1
Fat necrosis	1
Radiation change	0.6
Foreign body reaction	0.6

From Spencer *et al.* (1994)

Table 2. Positive predictive value (PPV) for malignancy of various mammographic abnormalities

Abnormality	PPV (%)
Microcalcifications	
All	45
Comedo	83
Non-comedo	35
Masses	
Spiculate	94
Ill defined	54
Well defined	4
Parenchymal deformity	37
Density with calcification	44

From Burrell *et al.* (1996)

Carcinoma *in situ*

Definition

Two non-invasive forms of breast carcinoma *in situ* are recognized: DCIS and lobular carcinoma *in situ* (LCIS). Each arises from its respective epithelial cell population in the lobule or duct of the normal breast. However, the neoplastic cell population is confined within the parenchymal site of origin, and the cells do not infiltrate beyond the limiting basement membrane. DCIS may harbour calcifications that make it mammographically apparent, but LCIS rarely gives rise to mammographic abnormalities (Goldschmidt & Victor, 1996).

Association of LCIS with invasive carcinoma

Lobular neoplasia includes LCIS and atypical lobular hyperplasia and is typically found incidentally in other benign and malignant breast lesions on histological examination (Figure 9). The relative risk for subsequent development of invasive carcinoma among patients with lobular neoplasia ranges from 4- (atypical lobular hyperplasia) to about 10-fold in women with LCIS (Page *et al.*, 1991; Dupont *et al.*, 1993; Marshall *et al.*, 1997), higher risks being associated with more extensive lesions (Page *et al.*, 1991; Fisher *et al.*, 1996). The invasive cancers seen after diagnosis of lobular neoplasia occur at roughly equal frequency in the ipsilateral and contralateral breast. Management of lobular neoplasia has evolved (Gump, 1993; Schnitt & Morrow, 1999) with better understanding of the disease. The current consensus is that both LCIS and atypical lobular hyperplasia are risk factors for subsequent development of invasive carcinoma in either breast. The value of routine mastectomy with or without contralateral breast biopsy has been questioned, and the majority of patients are managed by careful follow-up (Gump *et al.*, 1998).

Pathological classification of ductal carcinoma in situ (DCIS)

The classification of DCIS is evolving, and several groups have described systems for subdividing the lesions. The traditional classification, which is based on both architectural growth pattern and cytological features, is poorly reproducible, with up to 30% of cases in multicentre trials requiring reclassification (van Dongen *et al.*, 1992a). The lack of agreement among pathologists may be due largely to the architectural heterogeneity of DCIS. There is less heterogeneity in nuclear grade characteristics, and most of the contemporary histological classification systems are based on a three-tier grading or differentiation system with nuclear grade (National Coordinating Group for Breast Screening Pathology, 1995; Sneige *et al.*, 1998), grade and polarity (Holland *et al.*, 1994) or grade in the presence or absence of necrosis (Poller *et al.*, 1994: Silverstein *et al.*, 1995). Silverstein and colleagues have been particularly innovative in using histological grade, lesion size and distance of the excision margin in making a prognostic index (Silverstein *et al.*, 1996), and this has shown significant predictive power for local recurrence.

Although many of the histological classification systems appear to have

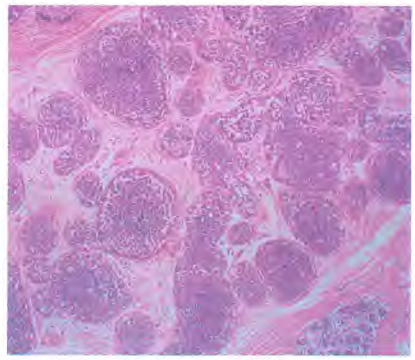


Figure 9 An example of lobular carcinoma *in situ*, showing filling of a distention of the acini of a breast lobule by a uniform population of epithelial cells. There are no associated features, such as calcification, inflammation or fibrosis, which would allow mammographic detection, and LCIS is typically a chance finding in breast biopsies resulting from breast screening.

been predictive, questions remain about diagnostic reproducibility among pathologists (Douglas-Jones *et al*., 1996; Scott *et al*., 1997; Badve *et al*., 1998; Sneige *et al*., 1998). Pathologists appear to have little difficulty in separating the entities at either end of the spectrum: problems of concordance of classification are generally found in the middle group and its boundaries and also at the boundary between low-grade DCIS and atypical ductal hyperplasia (Rosai, 1991; Schnitt *et al*., 1992; Sloane *et al*., 1994, 1999). Three recent consensus meetings came to similar conclusions and recommended that, until better data on clinical relevance and agreement among pathologists emerge, the morphological features present in DCIS and their nuclear grade should be recorded (Recht *et al*., 1994, Australia–New Zealand Breast Cancer Trials Group, 1996; Consensus Conference Committee, 1997). Nuclear grade should be assigned according to internationally accepted guidelines (Commission of the European Communities, 1996; Tavassoli & Stratton, 2002).

Calcification can be seen in both high- and low-grade DCIS (Figure 10) (Elston & Ellis, 1998; Evans *et al.,* 1994a; Tavassoli & Stratton, 2002). The mammographic calcification found in high-grade DCIS is more predictive of malignancy and generally more obvious, often showing coarse rod and branching shapes (Burrell *et al.*, 1996). This profile of subtypes of screen-detected DCIS suggests that radiologists might be able to distinguish subtypes of DCIS with different risks of progression to high-grade invasive disease (Evans *et al.*, 1994a).

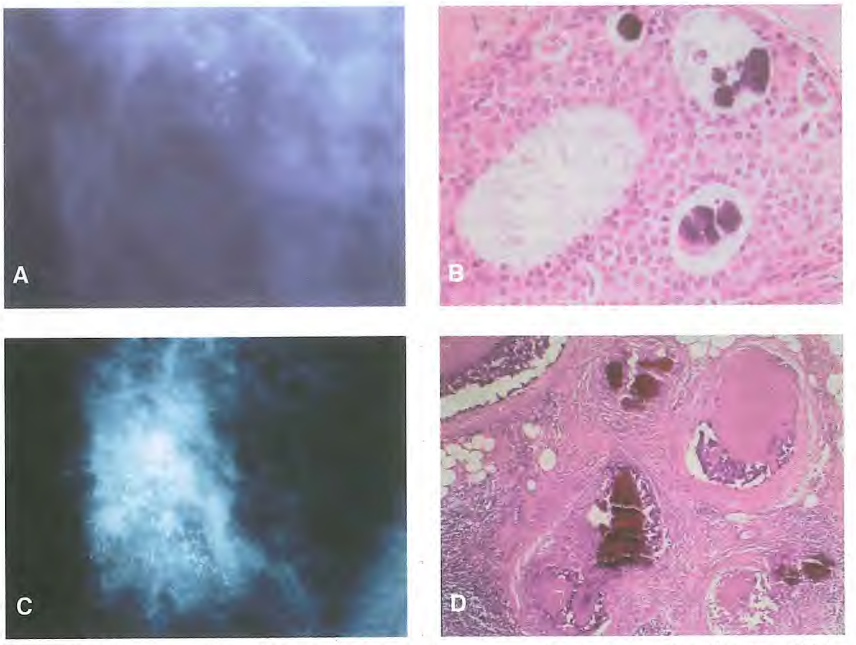


Figure 10 Mammograms and histological photomicrographs of typical examples of low- and high-grade DCIS. **A** Mammogram showing fine punctate calcification corresponding to the calcifications seen in secondary luminal spaces in the example of low-grade DCIS seen in **B**. The mammogram in **C** shows an extensive area of coarse calcification arising in luminal necrotic debris formed in the centre of ducts involved by high-grade DCIS, illustrated in **D**.

Association of DCIS with invasive carcinoma

For ethical reasons, there are limited data on the natural history of untreated DCIS. The available studies are from the 1930s to 1950s and relate to symptomatic, extensive, high-grade comedo DCIS. At that time, DCIS was rare in clinical practice, and patients typically presented with a mass lesion, nipple discharge or Paget disease of the nipple. This form of DCIS was defined at the time as aggressive. One very small but widely quoted series showed a 75% rate of progression to invasive disease, with a mean time to progression of 4 years (Dean & Geshchicter, 1938). This type of experience led to the prevailing effective use of mastectomy as the treatment of choice for symptomatic DCIS.

More recent studies reflect the opposite end of the spectrum of DCIS and are based on lesions originally classified as benign. Virtually all are examples of low-grade DCIS. In the studies with the longest follow-up, about 40% progressed to invasive disease after 30 years. In contrast to epithelial hyperplasia, atypical hyperplasia and LCIS, invasive tumours tend to occur in the same quadrant of the breast as the intitial lesion (Page *et al*., 1995).

Evidence from studies of recurrence after breast-conserving surgery for DCIS indicates that about 50% of recurrences are as invasive cancer and that high-grade DCIS and DCIS with necrosis represent a biologically aggressive subset of DCIS with higher rates of invasive and in-situ recurrence than low-grade DCIS lesions without necrosis (Solin *et al*., 1993; Silverstein *et al*., 1995, 1996; Fisher *et al*., 1999). One large randomized trial (Bijker *et al*., 2001a) showed that margin status is the most important factor in the success of breast-conserving therapy for DCIS. In this trial, the risk for subsequent development of distant metastasis after invasive local recurrence was significantly higher in patients with poorly differentiated DCIS

than in those with well-differentiated DCIS. Analysis of recurrences in this trial also showed that most primary DCIS lesions and their local recurrences were similar histologically or in marker expression, suggesting that local recurrence usually reflects outgrowth of residual DCIS; progression of well-differentiated DCIS to poorly differentiated DCIS or grade III invasive carcinoma is unusual (Bijker *et al.*, 2001b).

Invasive lesions with an extensive intraductal component also show a predisposition to local recurrence after breast-conserving therapy (van Dongen *et al.*, 1989). The grade of DCIS associated with invasive cancers has been shown to correlate with both disease-free interval and survival (Lampejo *et al.*, 1994). Strong associations also exist between the grade of invasive cancer and the grade of coexisting DCIS. High-grade DCIS is associated with high-grade invasive cancer and low-grade DCIS with low-grade invasive cancer (Lampejo *et al.*, 1994; Douglas-Jones *et al.*, 1996; Cadman *et al.*, 1997). An association between grade 3 invasive cancer and poorly differentiated DCIS is seen whatever the grading system used (Douglas-Jones *et al.*, 1996).

Genetic changes seen in in-situ carcinoma and atypical ductal and lobular hyperplasia

Molecular genetic studies of low-grade DCIS and atypical ductal hyperplasia with loss of heterozygosity techniques have demonstrated similar genetic lesions, providing, in informative cases, confirmatory evidence that these lesions are clonal and therefore fulfil the basic criterion of neoplastic transformation (Lakhani *et al.,* 1995). In addition, it has been shown that in-situ and invasive elements of breast cancers have identical molecular alterations, implying that they are stages in the same pathway (Stratton *et al.*, 1995). These findings are consistent with the observation that the two components have similar morphological characteristics (Lampejo *et al.*, 1994) and are also consistent with the hypothesis that low-grade in-situ cancer gives rise to low-grade invasive carcinoma and high-grade in-situ cancer to high-grade invasive carcinoma. Evidence from a study in two counties in Sweden (see Chapter 4) gave rise to an alternative hypothesis: that tumours progress from low to high grade, as the proportion of high-grade tumours increases with tumour size (Tabár *et al.*, 1992).

Recent studies, and particularly those in which comparative genomic hybridization was used to investigate DCIS, prompted the proposal of a hypothetical model for the pathogenesis of DCIS in which genetic lesions are associated with particular morphological subtypes (Buerger *et al.*, 1999). Different morphological classes of DCIS have specific genetic changes that are not shared by other types. In particular, low-grade and high-grade DCIS appear to be distinct, separate entities, on the basis of morphology, phenotype and molecular genetics. Well-differentiated DCIS is associated with loss of 16q and 17p, while tumours of intermediate and high grades often have losses of significantly more allelic chromosomal arms, frequently including 1p, 1q, 6q, 9p, 11p, 11q, 13q and 17q (Fujii *et al.*, 1996). High-grade DCIS in particular is associated with gains at 17q but also at 11q and 13q (Chuaqui *et al.*, 1997). Intermediate-grade DCIS appears to have a combination of lesions, which show 16q loss but gains at other chromosomes, particularly 1q; some cases show gain at 11q13q but lack the gain at 17q12 which is a feature of high-grade DCIS (Buerger *et al.*, 1999). Similarly, atypical lobular hyperplasia and LCIS show the same genetic mutations, with loss of material from 16p, 16q, 17p and 22q and gain at 6q (Lu *et al.*, 1998). Interestingly, although low-grade DCIS and atypical ductal hyperplasia have no molecular genetic similarity to high-grade DCIS, they have similarities to LCIS and atypical lobular hyperplasia. These observations challenge the existing assumptions that lobular and ductal lesions are distinct and that DCIS is a homogeneous disease. They also raise the possibility that future molecular markers will provide better discrimination among morphologically similar cells.

Implications for screening

Breast screening detects a wide spectrum of breast cancer, ranging from microfocal low-grade DCIS to large high-grade invasive cancer (Cowan *et al.*, 1991; Klemi *et al.,* 1992; Rajakariar & Walker, 1995). It has been proposed that detecting in-situ cancer, particularly high-grade DCIS, would prevent the development of high-grade invasive cancer (Lampejo *et al.*, 1994; Evans *et al.*, 1997, 2001a,b). It is well recognized that many low-grade, special invasive cancers are identified at screening (Cowan *et al.*, 1991; Klemi *et al.*, 1992; Porter *et al.*, 1999) (Figure 11). Such tumours have an excellent prognosis but may be so indolent that they would never have presented clinically or have threatened the

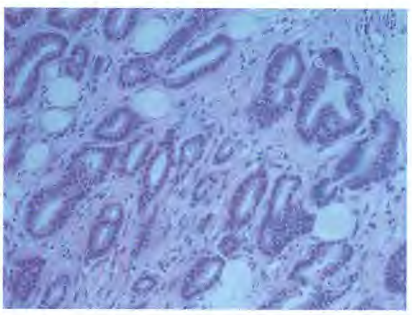


Figure 11 A well-differentiated invasive adenocarcinoma of the breast of tubular type. The tumour cells are arranged in rounded or elongated glandular or tubular structures with a central luminal space which closely mimics the normal breast terminal duct or lobule. An associated stromal fibrous reaction produces the typical stellate mammographic appearance that allows detection of these tumours by screening.

life of the patients. It has been proposed alternatively that a proportion of these low-grade invasive tumours might de-differentiate over time into more aggressive, less well-differentiated tumours (Tabár *et al.,* 1999), although this was not found in another screening programme (Hakama *et al.*, 1995). Identification and removal of such cancers when they are at a low grade would avoid such progression. Detection of high-grade invasive cancers when they are small is clearly a means by which screening could reduce breast cancer mortality. In support of this possibility, it was shown in the two-county trial in Sweden that histological grade 3 invasive cancers detected when less than 10 mm have an excellent prognosis (Tabár *et al.*, 1999), while it is widely recognized that large high-grade invasive cancers have a poor prognosis.

Ductal carcinomas of no specific type have time-dependent prognostic factors (i.e. size and lymph node stage) that are, in general, moderately good, suggesting that their detection at screening is effective. However, lobular cancers and lobular mixed cancers are larger and more frequently extended to lymph nodes at the time of mammographic detection; thus, identification of cancers with a lobular component by breast screening is not likely to be beneficial. This appears to be a consequence of the subtle mammographic features of lobular carcinoma, which are more commonly seen on only one mammographic view and less frequently contain calcification than ductal carcinomas not otherwise specified (NOS) (Cornford *et al*., 1995).

The same group examined the value of detecting DCIS at mammographic screening and showed that identification of high-risk types of calcification allows diagnosis of otherwise occult, co-existing, small grade 3 invasive carcinomas associated with calcific high-grade DCIS (Evans *et al.*, 1997). In addition, comparison in their series of the biological characteristics of DCIS detected at screening with symptomatic DCIS lesions showed a higher proportion of adverse characteristics in those detected at screening. The most likely explanation for these findings is suggested by a comparison of the radiological findings of different DCIS sub-types. High-grade DCIS more frequently showed abnormal mammographic features than low-grade DCIS. The granular and punctate calcifications seen in low-grade DCIS (Evans *et al*., 1994b) are more subtle, less specific and often not picked up at mammographic screening, as they are similar to those seen in common benign conditions (Holland *et al*., 1990; Evans *et al*., 1994a).

Invasive carcinoma

Definition

Invasive carcinoma of the breast is defined as a malignant tumour, part or all of which penetrates the basement membrane of the epithelial site of origin (i.e. the duct or lobule). The vast majority of these tumours are adenocarcinomas and are believed to be derived from the mammary parenchymal epithelial cell population, particularly cells of the terminal duct lobular unit. The morphological appearance of these tumours varies widely, and many of the recognized morphological types have particular prognostic or clinical characteristics. More recently, specific genetic lesions have been identified in some types.

Pathological classification of breast cancer

The prognosis of a patient with breast cancer is dependent on two distinct groups of variables. The first are those time-dependant variables that influence tumour stage, particularly the histological size of the tumour, the presence and extent of lymph node metastatic disease and the presence of systemic metastatic disease. The second group of variables, sometimes referred to as intrinsic characteristics, are related to the inherent biology of the individual tumour. This group includes the histological grade, tumour type, growth fraction, hormone and growth factor receptor status and an ever-lengthening list of molecular characteristics.

Of these features, tumour size, histological type, histological grade, vascular invasion status and lymph node status have been shown to be related to clinical outcome (Elston & Ellis, 1998). These features can be used:

- to decide on the most appropriate treatment for a particular patient, including the extent of surgery and the use and choice of adjuvant therapy;
- to monitor breast screening programmes, the success of which is reflected by more favourable prognostic features of the cancer detected; and
- to monitor changing patterns of disease incidence, particularly by cancer registries.

For these reasons, there is increasing international recognition that pathological classification of breast cancer should conform to a minimum dataset, which includes these key variables (Royal College of Pathologists, http). One common approach is based on a combination of invasive tumour size, nodal involvement and metastases (TNM). As it has three dimensions, it is commonly collapsed into one summary number from 0 to IV, where 0 is in-situ disease (UICC, 2002; American Joint Committee on Cancer, 2002)

Morphological features of invasive breast carcinoma relevant to prognosis and screening

The factors described below have been shown to provide clinically relevant prognostic information and are valuable in evaluating breast screening programmes.

Tumour size

Ideally, the size of tumours should be assessed on resected pathological specimens. In situations in which pathological size cannot be determined, such as in patients receiving primary systemic or neoadjuvant therapy or when several estimates of size have been made, alternative means should be used, including magnetic resonance imaging, ultrasound and clinical examination (UICC, 2002; American Joint Committee on Cancer, 2002).

As tumour size is a time-dependent factor, it has been shown consistently in many studies to influence prognosis (Cutler *et al.*, 1969; Elston *et al.*, 1982; Fisher *et al.*, 1984; Carter *et al.*, 1989; Neville *et al.*, 1992). Patients with smaller tumours have better long-term survival rates than those with larger tumours (Figure 12).

Estimation of tumour size has assumed particular importance in breast screening. The term 'minimal breast cancer' was originally introduced to identify forms of breast cancer for which there was an exceedingly good prognosis (Gallager & Martin, 1971); these included all cases of in-situ carcinoma (ductal and lobular) and invasive carcinomas measuring 5 mm or less. Subsequently, for no clearly defined reason, the invasive component was redefined by various groups. The Breast Cancer Detection Demonstration Projects (Beahrs *et al.*, 1979) and the American Cancer Society (Hartmann, 1984) used 9 mm or less as the maximum diameter, while the American College of Surgeons (Bedwani *et al.*, 1981) favoured up to and including 10 mm. This lack of uniformity in definition causes problems in the interpretation of data from different studies.

Tumour size is also an important quality assurance measure for breast screening programmes (Hartmann, 1984; Tabár *et al.*, 1987a; Royal College of Radiologists, 1997) and can be used in part to judge the ability of radiologists to detect small, impalpable invasive carcinomas on mammography. For example, the National Health Service Breast Screening Programme in the United Kingdom requires that 50% of the invasive cancers detected must measure less than 15 mm (Royal College of Radiologists, 1997). It is therefore incumbent on pathologists to measure tumour diameter as accurately as possible. As size decreases, so the risk for errors in measurement increases, and inconsistencies have been reported (Beahrs *et al.*, 1979; Sloane *et al.*, 1994).

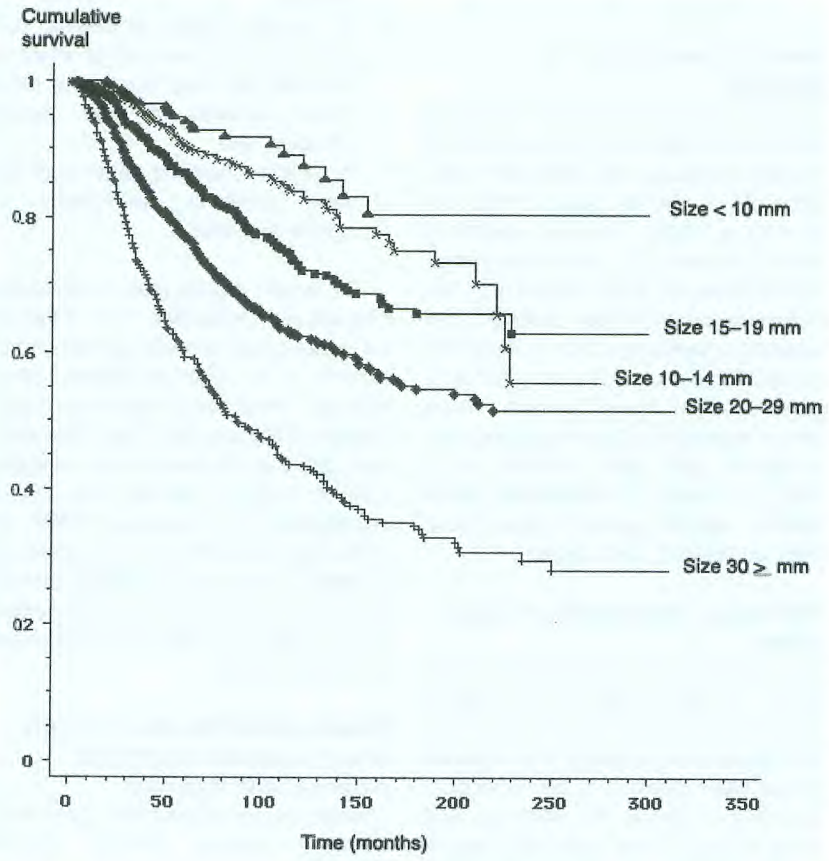


Figure 12 Relationship between tumour size and survival rate of patients with primary unoperable breast cancer. Kaplan-Meier plot of cumulative survival by size of invasive cancer.

Histological type

A wide range of morphological patterns can be seen in invasive carcinoma of the breast (Fisher, E.R. *et al.*, 1975; Azzopardi *et al.*, 1979; Page *et al.*, 1987; Ellis & Fidler, 1995), and many types have distinct prognostic characteristics (Page *et al.*, 1987; Ellis *et al.*, 1992). The diagnostic criteria are described in detail elsewhere (Page *et al.*, 1987; Ellis *et al.*, 1992; National Coordinating Group for Breast Screening Pathology, 1995; Rosen, 1997; Elston & Ellis, 1998; Tavassoli & Stratton, 2002) and will not be repeated here. It must be appreciated that a considerable subjective element remains, and there is not yet universal agreement on the criteria for all types. This is reflected in the relative

Prognostic characteristics of different tumour types

Group 1: **Excellent prognosis**
Tubular, tubulolobular, invasive cribriform, mucinous

Group 2: **Good prognosis**
Tubular variant or mixed, alveolar lobular, mixed ductal not otherwise specified and other special types

Group 3: **Average prognosis**
Medullary, atypical medullary, classicular lobular, lobular mixed

Group 4: **Poor prognosis**
Ductal not otherwise specified, solid lobular, mixed ductal not otherwise specified and lobular

From Pereira *et al*. (1995)

proportions of different types in published series (Elston & Ellis, 1998) and the observation that the consistency of diagnosis of histological type was disappointingly low in pathology quality assurance schemes (Sloane *et al.*, 1994, 1999), implying that pathologists should work to the same diagnostic protocols.

The favourable prognosis of certain histological types of invasive carcinoma of the breast is well established (see box; Pereira *et al*., 1995). Thus, tubular carcinoma (Cooper *et al*., 1978; McDivitt *et al.,* 1982; Carstens *et al*., 1985), mucinous carcinoma (Lee *et al.*, 1934; Clayton, 1986), invasive cribriform carcinoma (Page *et al.,* 1983), medullary carcinoma (Bloom *et al.*, 1970; Ridolfi *et al.,* 1977), infiltrating lobular carcinoma (Haagensen *et al*., 1978) and tubulolobular carcinoma (Fisher *et al*., 1977) have all been reported to have a more favourable prognosis than invasive ductal carcinomas NOS, but few comprehensive long-term follow-up studies of histological type in relation to survival have been carried out. Dawson and colleagues (1982) found a higher proportion of tubular, mucinous, medullary and infiltrating lobular carcinomas in patients who had survived at least 25 years after mastectomy than among those who had survived for less than 10 years. These findings were confirmed in a similar study from Edinburgh (Dixon *et al.*, 1985), with the addition of papillary and invasive cribriform carcinomas among the cancers in long-term survivors. These 'special' or 'specific' forms of invasive carcinoma have also been found at higher frequency in the prevalence round of mammographic breast screening programmes (Anderson *et al.*, 1991; Ellis *et al.,* 1993) and more frequently in carcinomas detected at screening than in cancers found between screening rounds (interval cancers) (Porter *et al.,* 1999).

A study of one series comprising over 1500 patients with primary operable invasive carcinoma who were followed up for a minimum of 10 years confirmed the excellent prognosis of pure tubular, invasive cribriform and mucinous carcinomas (Ellis *et al*., 1992). This study also showed that the categories of carcinoma with special characteristics, tubular mixed carcinoma and mixed ductal NOS and special type, are worth recording, as they carry a considerably better prognosis than ductal carcinoma NOS and form a significant proportion of all invasive cancers (15%). In previous studies, such mixed types were rarely recognized and the tumours were included in the general category of ductal carcinomas NOS.

It has become accepted dogma that medullary carcinoma (Figure 13) has an excellent or good prognosis (Moore & Foote, 1949; Richardson, 1956; Bloom *et al.,* 1970; Ridolfi *et al.,* 1977; Rapin *et al.,* 1988). It is interesting that this view has persisted, despite the fact that other studies have been unable to confirm better survival after medullary carcinoma than after ductal carcinoma NOS (Cutler *et al.,* 1966; Pedersen *et al.,* 1988; Fisher *et al.,* 1990; Ellis *et al.,* 1992). However, some of the latter studies showed that medullary carcinoma does have a better prognosis than ductal carcinoma NOS of grade 3 (Pedersen *et al.,* 1988; Fisher *et al.,* 1990; Ellis *et al.,* 1992). Some authors (Ellis *et al.,* 1992) therefore concluded that medullary carcinoma should be regarded as having a moderate rather than a good prognosis.

Overall, patients with infiltrating lobular carcinoma (Figure 14) have a slightly better prognosis than those with ductal carcinoma NOS (Haagensen *et al*., 1978; Ellis *et al.*, 1992), although the 10-year survival rate of 54% in the latter study clearly implies no more than a moderate prognostic outcome. However, Dixon and colleagues (1982) found significant differences in the survival of patients with different morphological subtypes of lobular carcinoma, and this has been confirmed (Ellis *et al.*, 1992). Thus, the classical type has a good prognosis

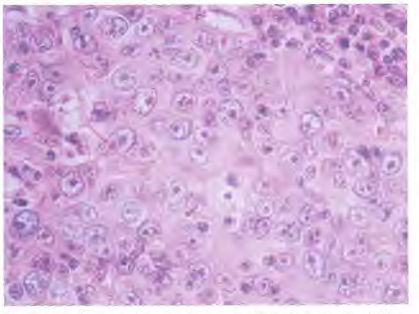


Figure 13 A medullary carcinoma of the breast composed of syncytial sheets of large pleomorphic tumour cells surrounded by stroma rich in lymphocytes and plasma cells

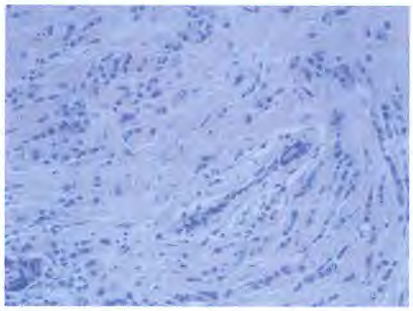


Figure 14 Classical invasive lobular carcinoma of the breast is composed of narrow files of small, regular tumour cells, which typically infiltrate the breast, surrounding existing parenchymal structures and causing little disturbance to the tissue architecture. This infiltrative patterns produce few mammographic signs, and lobular carcinoma is a cause of false-negative results in mammographic examinations, due to occult disease.

(60% 10-year survival rate), the mixed lobular type an average prognosis (55% at 10 years) and the solid lobular type a poor prognosis (40% at 10 years). Tubulolobular carcinoma, which has an excellent prognosis (over 90% 10-year survival rate), is currently considered a separate, distinct type because of lack of agreement about its assignment as a tubular or lobular variant.

The detection by breast screening of carcinomas with tubular features is facilitated by the high frequency of spiculation seen at mammography (Elson *et al.*, 1993). It is well recognized that pure tubular carcinomas detected at screening have a good prognosis. This is confirmed by the finding of Evans *et al.* (2001a) of a very low incidence of pure tubular cancers among women who subsequently developed metastatic disease (three of 173 patients (2%), three of the 16 grade 1 lesions (20%)). This suggests that these tumours may be overdiagnosed. The value of detecting pure tubular cancer at screening is therefore likely to be of benefit only if a proportion of tubular cancers de-differentiate if left in the breast. Overdiagnosis of tumours with some tubular features (tubular variant or mixed carcinoma) is less clear, as 10% of cancers that metastasize are of the tubular mixed type and these tumours do not have the exceptionally good prognosis of pure tubular carcinoma.

Grading of invasive carcinoma

Despite the diversity of methods used, many studies have demonstrated a significant association between histological grade and survival from invasive breast carcinoma. Grade is now recognized as a powerful prognostic factor that represents a simple method for classifying differentiation in all invasive breast cancers (Figure 15). Grade should be included as a component of the minimum data set for histological reporting of breast cancer (Henson *et al.*, 1991; Elston & Ellis, 1998; Royal College of Pathologists, http).

Various grading systems have been described, which are based on assessment of multiple cellular and architectural variables (Greenhough, 1925; Patey, 1928; Bloom, 1950a,b; Bloom & Richardson, 1957; Fisher *et al.*, 1984; Contesso *et al.*, 1987; Elston & Ellis, 1991) or nuclear variables (Hartveit, 1971; Black *et al.*, 1975; Le Doussal *et al.*, 1989). The absence of uniform defintion makes comparison of findings difficult.

Given the nature of the methods, assessment of histological differentiation will always carry an underlying subjective element; however, one of the fundamental problems with many of the early systems was the lack of strictly defined written criteria. Bloom and Richardson (1957) made a useful contribution by adding numerical scoring to the method described by Patey (1928) but did not provide clear criteria for their cut-off points. Elston and Ellis (1991) added further modifications to the above method and to their system and achieved greater objectivity and acceptable concordance. This method has been shown to be highly reproducible (Dalton *et al.*, 1994; Frierson *et al.*, 1995; Robbins *et al.*, 1995) and has been adopted internationally as the method of choice (National Coordinating Group for Breast Screening Pathology, 1995; Connolly *et al.*, 1996; Commission of the European Communities, 1996; American Joint Committe on Cancer, 2002; Tavassoli & Stratton, 2002; UICC, 2002). In this system, three characteristics of the tumour are evaluated: tubule formation as an expression of glandular differentiation, nuclear pleomorphism and mitotic counts (Table 3). A numerical scoring system on a scale of 1–3 is used to ensure that each factor is assessed individually, and an overall grade is assigned as follows:

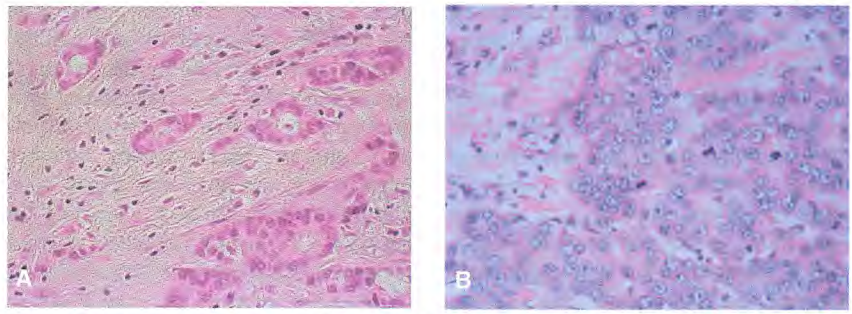


Figure 15 **A** A grade 3 or poorly differentiated invasive breast carcinoma. The tumour cells are arranged in sheets with no apparent gland formation. The cells are large and vary in size, and obvious mitotic figures are present. **B** Vascular invasion seen as a group of tumour cells present in a peritumoral lymphatic vascular channel

Table 3. Summary of semi-quantitative method for assessing histological grade of breast carcinoma

Feature	Score
Tubule and gland formation	
Majority of tumour (> 75%)	1
Moderate degree (10–75%)	2
Little or none (< 10%)	3
Nuclear pleomorphism	
Small, regular, uniform cells	1
Moderate increase in size and variation	2
Marked variation	3
Mitotic counts	
Dependent on microscope field area	1–3

Reproduced from Elston and Ellis (1991)

- Grade 1: well differentiated; 3–5 points
- Grade 2: moderately differentiated; 6–7 points
- Grade 3: poorly differentiated; 8–9 points

Lymph node stage

Involvement of loco-regional lymph nodes in breast cancer has long been recognized as one of the most important prognostic factors. Clinical assessment of lymph node status is not sufficiently accurate for therapeutic use, and evaluation of lymph node stage should be based on histological examination of excised nodes (Barr & Baum, 1992). Patients who have histologically confirmed loco-regional lymph node involvement have a significantly poorer prognosis than those without nodal involvement (Cutler *et al.*, 1969; Fisher, E.R. *et al.*, 1975; Elston *et al.*, 1982; Ferguson *et al.*, 1982; Haybittle *et al.*, 1982; Galea *et al.*, 1992; Veronesi *et al.,* 1993a). The overall 10-year survival rate is reduced from 75% for patients without nodal involvement to 25–30% for those with involved nodes. Prognosis is also related to the number and level of loco-regional lymph nodes involved: the greater the number of nodes involved, the poorer the patient survival (Nemoto *et al.*, 1980; Fisher *et al.*, 1984). Most groups stratify patients into two groups for therapeutic purposes: those with one to three positive nodes and those with four or more (American Joint Committee on Cancer, 2002; UICC, 2002). Similarly, involvement of nodes in the 'higher' levels of the axilla, and specifically the apex, carries a worse prognosis (Handley, 1972; Haagensen, 1986; Veronesi *et al.*, 1993a), as does involvement of the internal mammary nodes (Handley, 1972).

The frequency of disease with involved lymph nodes is significantly lower in women in whom disease is detected at screening. Approximately 40–50% of symptomatic patients have iinvolved nodes, in contrast to approximately 10–20% of patients with disease detected at screening (Cowan *et al.*, 1991; Klemi *et al.*, 1992; Rajakariar & Walker, 1995). This finding has raised concern that routine axillary lymph node dissection is over-treatment for many women with breast cancer detected at screening and has led to interest in use of sentinel lymph node biopsy for effective staging of the axilla (Krag *et al.*, 1998; Bundred *et al.*, 2000).

Vascular invasion (Figure 16) is defined as the presence of tumour cells in vascular spaces (lymphatic or blood) in tissues surrounding an invasive tumour (Örbo *et al.*, 1990; Pinder *et al.*, 1994). Vascular invasion correlates very closely with survival and loco-regional lymph node involvement (Rosen, 1983; Davis *et al.*, 1985; Örbo *et al.*, 1990; Pinder *et al.*, 1994). Possibly because of this association, it has been claimed that the prognostic information it provides is as powerful as lymph node stage (Bettelheim *et al.*, 1984). There is certainly a correlation between the presence of vascular invasion and early recurrence in patients with no lymph node involvement (Rosen *et al.*, 1981; Roses *et al.*, 1982; Bettelheim *et al.*, 1984), and some (Roses *et al.*, 1982; Pinder *et al.,* 1994) have shown that adverse prognostic effects are also independent of occult axillary node involvement. In addition, vascular invasion is a predictor of local recurrence after conserving therapy (Roses *et al.*, 1982; Locker *et al.*, 1989a; Rosen, 1991; Pinder *et al.*, 1994) and of flap recurrence after mastectomy (O'Rourke *et al.*, 1994).

As stated above both nodal and vascular invasion status are powerful independent prognostic factors in patients with invasive breast carcinoma (Bettelheim *et al.*, 1984; Todd *et al.*, 1987; Pinder *et al.*, 1994; Seidman *et al.*, 1995; Tabár *et al.*, 1999). In a study of the features associated with the development of metastatic disease after a previous breast cancer (Evans *et al.*, 2001a), 72% of 173 women who developed metastatic disease had nodal metastases and 59% had definite vascular invasion; 84% had either lymph node metastases or vascular invasion, or both. This finding was consistent, whatever the histological grade of the primary tumour. The absence of vascular invasion and nodal involvement in invasive breast cancer indicated a low risk for subsequent development of metastatic disease. Trends in the frequency of

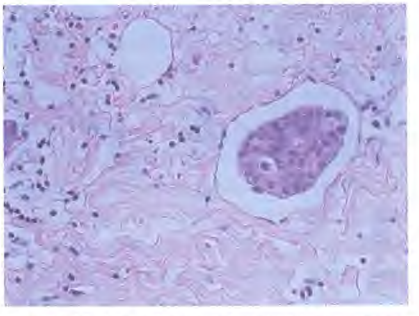


Figure 16 Vascular invasion seen as a group of tumour cells present in a peritumoral lymphatic vascular channel

nodal involvement and vascular invasion status according to histological grade, invasive size and tumour type were then examined in a group of 573 women with invasive cancers detected at screening, in order to predict the likelihood of development of systemic disease. Grade 1 invasive cancers less than 20 mm in diameter and grade 2 and 3 cancers less than 10 mm were associated with low rates of nodal involvement and vascular invasion. The criteria for selecting groups for analysis were the intrinsic morphological features of the invasive tumour (i.e. histological grade and tumour type) at different sizes. Of the well-differentiated, less intrinsically aggressive (grade 1) carcinomas, only those over 20 mm were associated with a high rate of lymph node involvement. Nevertheless, 9% of the primary breast cancers that metastasized were grade 1 lesions. Thus, large grade 1 invasive cancers can, and do, spread. Detection of these lesions when they are small might be seen as overdiagnosis but could prevent progression to a size associated with metastasis. Some types of low-grade invasive breast carcinoma have, however, an exceptionally good prognosis even when metastatic disease is present (Diab *et al.*, 1999). The detection of low-grade invasive and in-situ breast carcinoma therefore remains of questionable value.

The low rates of nodal positivity and vascular invasion of grade 2 invasive cancers less than 10 mm in diameter identified by screening indicate the value of detecting them at this size. Grade 2 cancers of 10–15 mm were associated with moderately high rates of nodal involvement but low rates of vascular invasion. The benefit of detecting grade 2 cancers 10–15 mm in size is therefore less clear. Larger grade 2 cancers (over 15 mm) were already associated with high rates of nodal involvement and vascular invasion at the time of diagnosis. Their detection by mammographic screening may therefore be of limited benefit.

Low rates of both nodal positivity and vascular invasion were seen in grade 3 invasive cancers less than 10 mm in diameter detected at screening, suggesting that detection of these small high-grade tumours is valuable, especially as larger grade 3 invasive cancers have such a poor prognosis. Women with grade 3 cancers over 20 mm in this series had high rates of affected lymph nodes and vascular invasion; therefore, detection at this stage is unlikely to influence survival. The moderate rates of nodal involvement and vascular invasion in grade 3 cancers of intermediate size (10–20 mm) suggest that their detection is less likely to be beneficial than when they are small. Similar views have been developed from reviews of other mammographic screening populations (Tabár *et al.*, 1999).

Molecular markers

Many molecular alterations have been identified which reflect the biological characteristics of invasive breast carcinomas. Some are related to survival, but, more importantly, these changes indicate which molecular pathways affect a tumour and could therefore predict benefit from specific forms of molecular therapy.

One such marker is steroid hormone receptors (the estrogen receptor (ER) and the progesterone receptor (PR). Estrogen is an important mitogen, which expresses its activation by binding to its nuclear receptor (ER) (Figure 17). ER is expressed in 60–80% of invasive breast tumours, and ER-positive tumours have a better initial prognosis than ER-negative tumours. The presence of nuclear hormone receptors is useful for predicting response to hormone therapy, such as adjuvant tamoxifen (Osborne, 1998; Bundred, 2001; Isaacs *et al.*, 2001). ER- and PR-positive tumours have a 60–70% response rate, while that of ER- and PR-negative tumours is less than 10%. ER-positive, PR-negative tumours have an intermediate response of approximately 40%.

The *ERBB2/HER2* oncogene, located on 17q21, is amplified in approximately 20% of invasive breast cancers, leading to overexpression of the coded HER2 protein, a transmembrane receptor with tyrosine kinase activity (Figure 18). The prognostic value of *HER2* overexpression, first reported in 1987 (Slamon *et al.*, 1987), has been studied extensively (Tsuda *et al.*, 2001; Yamauchi *et al.*, 2001). *HER2* overexpression is a weak to moderate independent predictor of survival, at least for patients with node involvement. Amplification or overexpression can be measured by Southern blot analysis, fluorescent in-situ hybridization (FISH) or differential polymerase chain reaction to detect

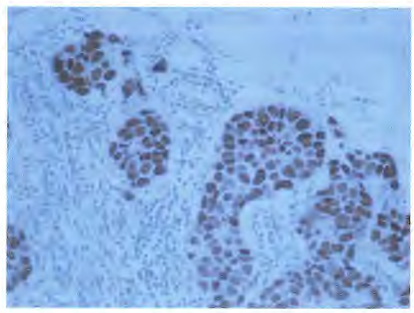


Figure 17 An invasive breast cancer stained immunocytochemically for estrogen receptors. The estrogen receptor protein is seen as a brown pigment in the tumour cell nuclei.

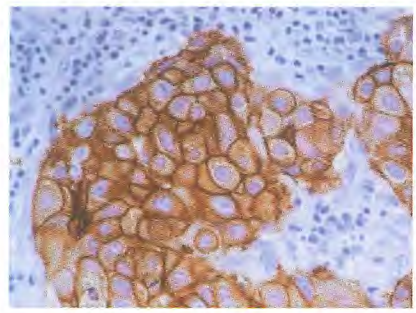


Figure 18 An invasive carcinoma stained immunocytochemically for HER 2 protein. The protein, seen as a brown pigment, is overexpressed on the tumour cell membranes.

gene amplification and immunohistochemistry or enzyme-linked immunosorbent assay to detect protein expression (Tsongalis & Ried, 2001). The results of studies of the predictive value of HER2 status have not been consistent. A review by Yamauchi *et al.* (2001) concluded that HER2 is a weak-to-moderate negative predictor of response to alkylating agents and a moderate positive predictor of response to anthracyclines and that the data are insufficient to draw conclusions on the response to taxanes or radiotherapy. A humanized anti-HER2 monoclonal antibody, trastuzumab (Herceptin), has been developed as an anti-cancer drug targeting amplified and overexpressed *HER2* (Cobleigh *et al.*, 1999).

Markers of proliferation have been investigated extensively in relation to prognosis (Fitzgibbons *et al.*, 2000; Isaacs *et al.*, 2001). Mitotic count is part of histological grading (see above). Other methods include DNA flow cytometry measurement of the S-phase fraction and immunohistochemistry with antibodies directed against antigens present in proliferating cells like Ki-67. Several hundred studies on the S-phase fraction, with various techniques, indicated that a high S-phase fraction is associated with inferior outcome. Ki-67 is a labile, non-histone nuclear protein that is not expressed in resting cells but is detected in the G1 through M phases of the cell cycle. The percentage of Ki-67-positive cells can be used to stratify patients into good and poor survivors.

Genetics and invasive cancers

In the past decade, the ability to measure both molecular markers of cancer activity and the genes that control cell growth has increased tremendously. In future, this information may complement (and even supplant) the histological categorization described above. The basic approach is to relate the pattern of expression of multiple genes to the rate of growth of the tumour. This process may help clinicians to predict which cancers will grow fast and which will not.

Genetic changes in specific types of invasive breast cancer. Specific genetic lesions or regions of alteration are associated with specific histological types of cancer and are related to grade in large ductal carcinoma NOS. The latter group appear morphologically similar but include a number of tumours with unrelated genetic evolutionary pathways (Buerger *et al.*, 2001). They also show fundamental differences from some special type tumours, including lobular (Gunther *et al.*, 2001) and tubular carcinoma (Roylance *et al.,* 1999). Furthermore, recent cDNA microarray analyses have shown that ductal tumours NOS can be classified into subtypes on the basis of expression patterns (Perou *et al.*, 2000; Sorlie *et al.*, 2001) .

Genetic changes have also been found in invasive lobular carcinoma (Frixen *et al.*, 1991; Vleminckx *et al.*, 1991; Gamallo *et al.*, 1993; Rasbridge *et al.,* 1993; Berx *et al.*, 1995; Nishizaki *et al.*, 1997; Flagiello *et al.*, 1998), but they are identified less frequently than in ductal cancers (Nishizaki *et al.*, 1997; Flagiello *et al.*, 1998).

Tubular carcinomas of the breast have a lower frequency of genetic alterations than other types of breast carcinoma (Man *et al.*, 1996; Roylance *et al.*, 1999; Waldman *et al.*, 2001). Of particular interest is the observation that sites of chromosomal alteration frequently affected in other types of breast cancer are not seen, implying that tubular carcinoma of the breast is a genetically distinct group of breast cancers.

Up to 13% of carcinomas arising in carriers of the *BRCA1* gene are of the medullary type (Marcus *et al.*, 1996; Breast Cancer Linkage Consortium, 1997), and 35–60% exhibit medullary-like features (Marcus *et al.*, 1996; Armes *et al.*, 1998). Reciprocally, in a group of medullary cancers, germ-line mutations of *BRCA1* were observed in 11% of cases (Eisinger *et al.*, 1998). There is thus a large overlap between medullary features and the phenotype of *BRCA1* germ-line-associated tumours, but not all *BRCA1* mutations lead to the medullary phenotype. Medullary carcinomas are also characterized by a high rate of *P53* alterations (de Cremoux *et al.*, 1999)

Gene expression. Gene expression profiles may offer more information than morphology and provide an alternative to morphology-based tumour classification systems. The recent development of laser capture microdissection and high-density cDNA arrays provides a unique opportunity to generate such profiles of cells from tumours in various stages of progression (Kitahara *et al.*, 2001). Although this field is still in its infancy, it has already been shown that variations in gene expression can be used to classify breast cancers into a basal epithelial-like group, a luminal epithelial/ER-positive group, an HER2-overexpressing group and a normal breast-like group (Perou *et al.*, 2000; Sorlie *et al.*, 2001). The luminal group has since been divided into at least two subgroups, each with a distinctive expression profile. It is expected that gene sets will be identified that correlate with patient outcome or predict patient response to treatment.

Expression profiles based on microarrays will make it possible to analyse the expression of thousands of genes simultaneously and will allow the classification of tumours into new groups according to gene expression patterns (Alizadeh *et al.*, 2001; Gruvberger *et al.*, 2001; Hedenfalk *et al.*, 2001; Perou *et al.*, 2000; Sorlie *et al.*, 2001; West *et al.,* 2001). Expression patterns have shown biological differences between tumours: hereditary breast cancers with *BRCA1* mutations could be distinguished from those in *BRCA2* carriers (Hedenfalk *et al.*, 2001), and ER-positive and ER-negative cancers had different expression

profiles (Gruvberger *et al.*, 2001; West *et al.*, 2001). Analysis of a number of breast cancer series has resulted in identification of at least five different subtypes with different prognostic outcomes (Perou *et al.*, 2000; Sorlie *et al.*, 2001; van't Veer *et al.*, 2002).

Familial risk. Breast cancer has been recognized for over 100 years as having a familial component (Brocca, 1866). Its genetic basis is discussed in Chapter 4.

Can a patient be 'cured' of breast cancer?

The concept of cure in breast cancer has been problematic, as deaths occurred at all intervals in short-, medium- and long-term follow-up studies. Three concepts of cure have been defined — statistical, clinical and personal — and the evidence for the curability of female breast cancer according to each of these concepts has been examined (Haybittle, 1991). The author concluded that there was no convincing evidence of statistical or clinical cure in series of treated patients, but that one-fourth of such patients had experienced individual cure, in that they died from some other cause without overt signs of breast cancer. This view was based on the few large studies with long-term follow-up, which showed persistently worse survival up to 30 years after diagnosis when compared with aged-matched controls, some of the later deaths being attributed to treatment rather than metastatic disease (Haybittle *et al.*, 1989). The level of individual cure in series of patients treated more recently should be higher, due mainly to better stage distribution. It has been shown that deaths rarely occur 20 years after diagnosis (Joensuu & Toikkanen, 1995), indicating that cure may be achieved. Analysis of one large, long-term follow-up study showed that the time to death of patients dying of breast cancer is influenced not by the time-dependent factors of tumour size and lymph node status (which appear to predict the risk for death) but by the intrinsic factor, histological grade (Blamey *et al.*, 2000). Of women who died, 90% of the deaths occurred within 8 years of diagnosis in patients with grade 3 tumours and within 13 years in patients with grade 2 tumours and were projected to occur within 30 years in patients with grade 1 tumours. The survival curves of patients with grades 2 and 3 tumours mirrored that of the general population after 90% of deaths had occurred. Patients with grade 1 tumours had a low overall risk of dying. These results suggest that not all patients with invasive breast carcinoma have systemic disease at diagnosis, and that patients could be offered advice on their risk of death depending on the grade of their tumour. Those patients who live for defined times after diagnosis could be reassured that their risk for death from breast cancer is the same as if they had not had breast cancer and is equivalent to cure.

Diagnosis and treatment

Diagnostic and treatment approaches have changed throughout the history of breast screening. In the early 1960s, when the first evaluation of mammography began, radical mastectomy was the predominant form of therapy, and this did not vary with tumour or patient characteristics. At present, breast conserving techniques with radiation therapy, adjuvant chemotherapy and hormonal treatments are used in a variety of ways, depending on age and tumour size and stage. Diagnostic approaches have also evolved over the past 40 years to accommodate the need to find smaller and smaller tumours. European guidelines for quality assurance in diagnosis and treatment provide a reference for implementing present practice (Commission of the European Communities, 2001). This section gives a summary of diagnosis and treatment strategies, reflecting current evidence-based practices in high-income regions.

Current diagnostic strategy

Diagnosis of breast cancer depends on whether or not a lesion can be felt (whether it is palpable). When there is a palpable lesion, a diagnosis is made on the basis of the results of three techniques: inspection and palpation, mammography and core-cutting needle biopsy or fine-needle aspiration biopsy. If there is a suspicion of malignancy, operative excision is recommended.

Non-palpable lesions pose a greater diagnostic challenge. In this circumstance, a suspect area on a mammogram is localized by further magnification, stereotactic mammography and/or ultrasound. Biopsy is performed by core cutting or fine-needle aspiration, usually with guidance by imaging techniques, such as ultrasound or mammography. Operative excision of the area is again undertaken for any suggestion of malignant change.

Evolution of treatment guidelines

Operative treatment

Breast-conserving surgery. Breast cancer treatment was based for a long time on the Halsted hypothesis, according to which breast cancer spreads only by direct infiltration or via the lymphatic vessels into the lymph nodes. Halsted radical mastectomy was the predominant method of operation until the 1970s, when two prospective randomized trials confirmed that the prognosis was similar with less extensive operation (Turner *et al.*, 1981; Maddox *et al.*, 1983). It was thus concluded that breast cancer can send distant metastases at an early stage, and lymph node metastases are not necessarily a result of cancer spread along the lymphatic vessels but rather an indicator of systemic disease. Consequently, the extent of local treatment will not affect survival. This hypothesis has since been replaced by the view that breast cancer is a heterogeneous disease, some forms remaining local for a long time and others becoming systemic relatively early. According to this third hypothesis, the role of local treatment is

often crucial (Hellman, 1994). Prospective randomized studies conducted since the 1970s showed that survival after breast-conserving surgery combined with radiotherapy was similar to that after mastectomy (Sarrazin *et al.,* 1989; Fisher *et al*., 1989; Veronesi *et al.,* 1990; Blichert-Toft *et al*., 1992; van Dongen *et al.*, 1992b; Fisher *et al.*, 1995; Jacobson *et al*., 1995; van Dongen *et al.*, 2000).

Although breast-conserving surgery has become more popular since the 1980s, there is wide variation in the proportion of breast-conserving operations, due to differences in patient populations, hospital resources and surgeons' abilities and attitudes (Kotwall *et al.*, 1996; Margolese, 1999). Breast-conserving surgery was first shown to be safe for patients with tumours less than 2 cm in diameter. Initially, the local relapse rate was higher after breast-conserving surgery for patients with tumours 2–5 cm in diameter and for those with axillary node involvement (van Dongen *et al.*, 1992b). Later, breast-conserving surgery combined with various adjuvant treatments was shown to give results similar to those of mastectomy (van Dongen *et al*., 2000). Currently, breast-conserving surgery is preferred whenever possible, i.e. for invasive tumours less than 3 cm in diameter and for DCIS with tumour-free margins.

Axillary lymph nodes. Axillary lymph nodes are removed primarily for staging, but the operation also has therapeutic significance, preventing axillary recurrence. The number of metastatic lymph nodes among all the lymph nodes removed is reported. At least 10 nodes should be removed (Grabau *et al.*, 1998; Orr, 1999). The number of metastatic lymph nodes is an important prognostic factor. Thus, if more than 10 lymph nodes are removed from I and II axillary levels and they are all free of metastasis, there will be no local recurrence in the subsequent 5 years (Axelsson *et al.*, 1992). If no involvement of axillary lymph nodes is detectable by palpation and ultrasound examination and the nodes are not excised, survival will be reduced by 5% (Orr, 1999). All I and II level axillary lymph nodes are removed from patients with invasive breast carcinoma. If, during the operation, lymph nodes suspected of containing metastasis are detected, III level axillary lymph nodes are also removed.

Post-operative radiotherapy
Post-operative radiotherapy with approximately 50 Gy for 5 weeks reduces local recurrence after breast-conserving surgery (Fisher *et al.*, 1989; Clark *et al.,* 1992; Veronesi *et al.*, 1993b) and after mastectomy (Overgaard *et al*., 1997). Post-operative radiotherapy is given after breast-conserving surgery. For patients with lymph node metastases or tumours in stage 3 or 4, post-operative radiotherapy is also given after mastectomy.

Adjuvant cytostatic chemotherapy
For most of the past century, breast cancer was considered to require mainly local treatment. In the 1970s, it was shown in controlled trials that adjuvant cytostatic chemotherapy reduced local recurrence in patients with lymph node involvement (Fisher, B. *et al*., 1975; Bonadonna *et al.,* 1977) and improved the disease-free and overall survival rates by 15–20% (Bonadonna *et al.*, 1995). The standard regimen until the late 1990s was 4–6 months of cytoxan, methotrexate and 5-fluorouracil. This has been replaced gradually by anthracycline-based combinations, especially for younger patients.

Adjuvant therapy is recommended when the risk for recurrence is intermediate or high, i.e. more than 10% over 10 years (Fisher *et al.*, 1992). Adjuvant treatment is given to all women under 35 years of age. Currently, a growing number of patients with no node involvement receive adjuvant cytostatics, according to their tumour characteristics (Fisher *et al*., 1997; Kroman *et al*., 2000).

Adjuvant hormonal therapy
Adjuvant hormonal treatment with the anti-estrogen tamoxifen improves the disease-free and overall survival rates of patients who have undergone radical surgery (Nolvadex Adjuvant Trial Organisation, 1985, 1988). The current standard treatment for post-menopausal, ER- and/or PR-positive patients is 20 mg/day for 5 years. This treatment increases the 5-year survival rate by 15% (Veronesi *et al*., 1998). Newer selective ER modifiers and/or aromatase inhibitors may improve the survival of patients who would otherwise have received tamoxifen (Bonneterre *et al*., 2001).

Screening for breast cancer: Conceptual considerations

The core concept of screening is that detection of early disease offers the opportunity to change its prognosis. Earlier diagnosis may improve prospects for survival because early intervention permits treatment at a more tractable stage (Morrison, 1992). However, as experience with screening has accumulated and understanding of cancer biology has evolved, it is apparent that there is substantial heterogeneity among cancers at particular sites, and this heterogeneity may well influence the impact of screening. Models of screening should incorporate this heterogeneity.

The epidemiological model discussed in this section is an operational one for screening and incorporates heterogeneitiy among cancers. It makes no assumption about the biological nature of the process of cancer progression. The model applies principally to mammographic screening for breast cancer, in which the great majority of detected lesions are invasive cancer, and it is assumed that these will not progress.

General definitions

Several definitions are needed to understand this simplified screening model, and these are shown in Figure 19. First, the model assumes that there is a period in which there is no detectable disease, but early malignant changes may have taken place and a clone of cells is dividing and de-differentiating until it attains a size that could be detected by screening. The point at which a tumour could be found by screening begins the sojourn time (Zelen & Feinleib, 1969) or 'detectable preclinical phase'. 'Lead time' refers to the period between when a cancer is found by screening and when it would appear through clinical signs and symptoms (Morrison, 1992). Sojourn time is a characteristic jointly of the lesions and the screening test. Lead time will in addition be affected by the frequency of screening. Neither the sojourn time nor the lead time is directly observable for an individual, unless a screening test is repeated at frequent intervals, the results of a positive screening test are ignored and the person is observed until she becomes symptomatic. Such a situation is clearly not tenable. However, in a population that has undergone screening, the distribution of lead time and sojourn time can be estimated (see below).

Sojourn time is a characteristic of a particular lesion. It is expected to vary widely for different lesions, reflecting the wide biological heterogeneity of breast cancer. Sojourn time notably depends for example on histological grade.

In addition to sojourn time and lead time, two parameters of traditional importance in screening are sensitivity and specificity. For a condition which either exists or does not, such as Tay Sachs disease, these two parameters are defined in terms of a 2 x 2 table:

		Result of diagnostic test	
		Positive	Negative
Result of screening	Positive	a	b
	Negative	c	d

Sensitivity = a/(a + c)
Specificity = d/(b + d)

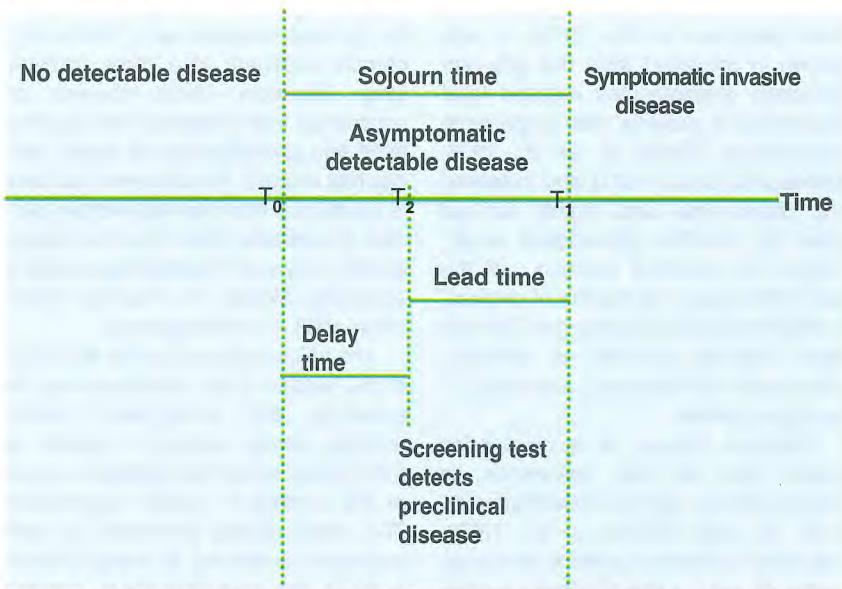


Figure 19 Scheme of progression of a chronic disease, with the intervention of an early-detection screening test

The situation is more complex for screening for breast cancer, because it is a progressive condition. At the time at which screening is performed, there is no 'gold standard' diagnostic test for the disease: the condition being screened for is a future clinical disease. The 'true' disease state at the time of screening is a lesion that will progress into a clinically invasive cancer. This state can be determined for an individual only by following her forward in time. Since, however, a positive result at screening should lead to an intervention to prevent the development of a clinical cancer, much of the information required for direct estimates of sensitivity and specificity will be missing. There is no direct measure of 'a' in the table. The quantity 'c', however, can be estimated directly, since, if one follows forward in time a group of individuals who showed no lesion on the screening test, some will develop clinical invasive disease. The length of time after screening that is used to define this group of 'screen-negative' and 'disease-positive' individuals is commonly 1 year, but that is a somewhat arbitrary interval. The women presenting with clinical disease in the year after a negative result thus constitute the cell entry 'c' in the above table.

In a number of programmes, a 2-year interval is used to define sensitivity. This has the advantage that it is less subject to statistical variation due to small numbers and less dependent on the exact date of diagnosis, although more affected by new cancers. Clearly, the longer the interval used to define sensitivity, the lower will be the resulting estimate (as follows from the discussions below and Figure 20).

Attention should also be paid to the definition of the screening test. Mammographic screening is essentially a multiple-step process, with the initial screening mammogram leading, if positive, to a series of more detailed

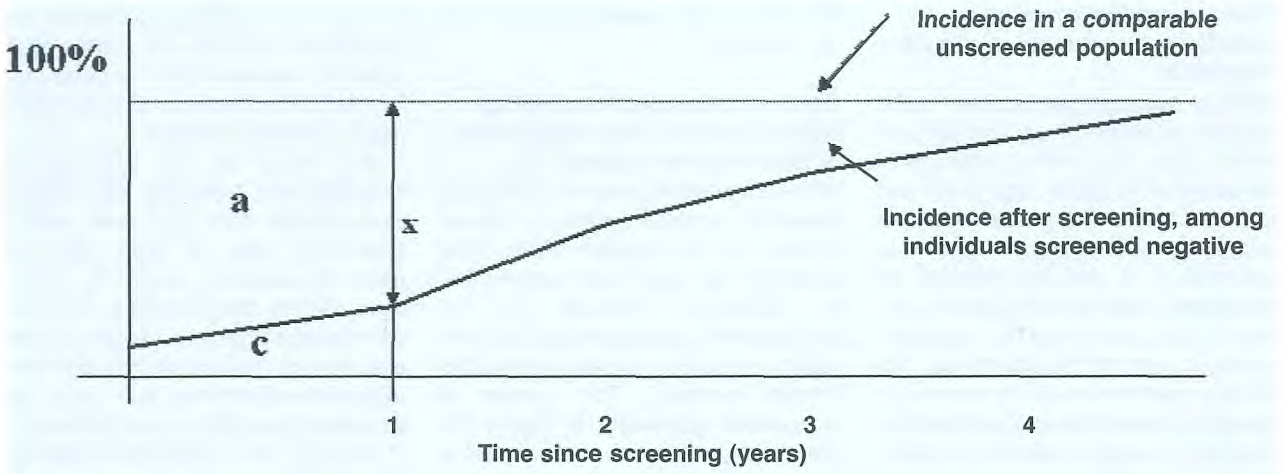


Figure 20 Sensitivity defined in terms of 1-year proportionate incidence: incidence of interval cancers as a proportion of the incidence in comparable unscreened population
c = interval cancers in first year; a = deficit of cancers in first year by comparison with an unscreened population; x = deficit of incidence at 1 year

investigations, culminating in a biopsy for a definitive diagnosis of malignancy. The definitions of sensitivity and specificity discussed in this section refer to the complete screening episode, the final assessment of positivity or negativity being based on the results of the mammogram and all further assessment. It is a common experience that women with a positive mammogram but classified as negative, i.e. disease free, on further assessment are at higher risk of subsequent disease than the general population. The implication is that if only the screening mammogram is considered as the screening test, it will be more sensitive than the overall screening episode, although of course with less specificity. The sensitivity of screening mammograms could be estimated in analogous fashion to the sensitivity of the complete screening episode, but in practice such estimation is rarely attempted. It would refer strictly to the sensitivity of the test itself.

To estimate sensitivity, one then must identify the individuals, or indirectly estimate their number, who constitute the cell entry 'a'. As the 'true' disease state is agreed to be clinical cancer appearing within 1 year of a screening test, to estimate 'a' one needs to estimate the number of true cancers that were detected at screening and treated, and thus prevented from presenting clinically in this period of 1 year after screening. This group forms the screen-positive, disease-positive group. The quantity 'a + c' is the number of cancers that would have presented clinically in the group that was screened if no screening had taken place. Thus, if one has a directly comparable unscreened population, e.g. as in a randomized trial, the quantity 'a + c' is observable. In the absence of a comparison group strictly defined by randomization, other approaches would be needed, but for any general population sample, estimates based on age-adjusted cancer incidence data from a comparable population or time when screening was not practised should provide a good approximation if used judiciously. The quantity 'a' is then obtained by subtraction, and the sensitivity estimate is given as before (Day, 1985):

$$a/(a + c).$$

For the test sensitivity, the same expression, $a/(a + c)$, applies, except one moves from 'c' to 'a' the interval cancers that were positive on the mammography test but negative on follow-up. This approach to the estimation of sensitivity, called the 'incidence method', can be expressed graphically as in Figure 20 and can be used to estimate sensitivity as the 1-year proportionate incidence of interval cancers (see below). In a definition of sensitivity that was sometimes used in the past, the 'gold standard positive' tumours were considered to be all those diagnosed at screening or within 1 year after screening, so that, in the above terminology, sensitivity was given by $(a + b)/(a + b + c)$. This quantity has no clear interpretation, since the cancers diagnosed at screening could have surfaced at any (including infinite) time in the future or never (overdiagnosis), whereas the false-negative results at screening surfaced in the first year after screening. The two groups are clearly not comparable.

Positive predictive value, specificity and the issue of overdiagnosis

A similar approach can be taken to the definition of specificity and positive predictive value, as, if one has estimates of the values of 'a' and 'c', and 'a + b' and 'c + d' are known from the results of screening, then clearly one has estimates of 'b' and 'd'. However, for determining specificity and positive predictive value, an interval of 1 year after screening may not be appropriate. For positive predictive value, for example, it would be of more interest to estimate the proportion of lesions detected at screening that would have progressed to clinical cancers (*i.e.* a/(a + b)) by the next round in a periodic screening programme. This parameter is of direct relevance to the question of overdiagnosis, which relates to invasive cancers detected at screening that will not progress to a clinical cancer within some defined time interval.

For specificity, or rather its complement, one might be interested in the proportion of individuals who had a positive screen among those who would not have developed a clinical cancer in the interval between screening tests (i.e. b/(b + d)). For the test specificity, the false-positive results should be included. The test validity indicators correspond to each other like those of episode validity. In particular, it is deficient to report (only) episode specificity and only test sensitivity.

In screening for a progressive disease, such as breast cancer, it is important to define the interval over which sensitivity and the other parameters are being defined and to ensure that they are comparable across populations. As demonstrated by Rosenberg *et al*. (2000), lengthening the observation period after a screening mammogram will decrease the sensitivity. Conversely, the shorter the interval the more important it is to remove the time assumed by episode from the woman–years of interval cancer incidence.

Cancers detected at screening, interval cancers and distribution of lead time and sojourn time

When a population of women undergoes screening, a certain number of breast cancers will be detected at the initial screening test, and further cancers will be diagnosed clinically in the post-screening period among those with negative results at screening (so-called 'interval cancers'). This process is represented graphically in Figure 20. The probability of a cancer being detected at screening clearly depends on the length of time the lesion is detectable preclinically, i.e. on the sojourn time: the longer the sojourn time, the greater the chance that the lesion will be detected. Cancers detected at screening thus represent a biased sample of preclinical lesions, with an undue proportion of cancers with a long sojourn time and probably a good prognosis. This bias is known as length bias.

In Figure 21, the incidence rate of breast cancer after screening is expressed as a proportion of the incidence rate in an equivalent unscreened population. The deficit in incidence in comparison with the unscreened represents those cancers detected at screening, as described in the definition of sensitivity. The curve of incidence after screening not only gives the proportion of cancers that are detected at screening (sensitivity), but also the time at which after screening the cancers detected would have presented clinically. Thus, in Figure 21, there is an incidence deficit of 'x' 1 year after screening. According to the definition of lead time, this deficit of 'x' corresponds to cancers detected at screening with a lead time of 1 year. The complete distribution of lead time among the cancers diagnosed at screening that would have presented clinically is therefore given by the difference between the unscreened incidence rate and the post-screening incidence rate, from time zero through to the maximum time for which observations are available.

The curve of the proportionate incidence after screening will increase monotonically from time zero until it approaches unity, at which time the effect of screening essentially disappears. Shortly after screening, the curve will represent largely the cancers missed at screening. The increase with time then represents cancers that were not in the preclinical detectable phase at the time of screening. So, 1 year after screening, the interval cancers will consist of all those with a sojourn time less than 1 year, and hence not detectable 1 year previously, plus those with a sojourn time greater than 1 year but missed at screening. The curve of proportionate incidence after screening thus represents a combination of sensitivity and the cumulative distribution of sojourn time among cancers diagnosed at screening that would have presented clinically. Separation of the two is difficult (Walter & Day, 1983; Day & Walter, 1984), but broad areas of acceptable (and correlated) values can be identified.

Periodic screening: Length bias and the unbiased set

Population screening programmes usually aim at screening each woman at regular intervals, normally between 1 and 3 years. In this situation, cancers would be detected at each screening test, and clinically detected cancers would present in each of the intervals between screening rounds. Figure 21 shows the process graphically. One can define a screening cycle as the period between the ends of two screening rounds.

A useful extension of the concept of sensitivity can be derived from programme sensitivity. This refers to the cumulative incidence of interval cancers during the screening cycle, as a proportion of the cumulative incidence

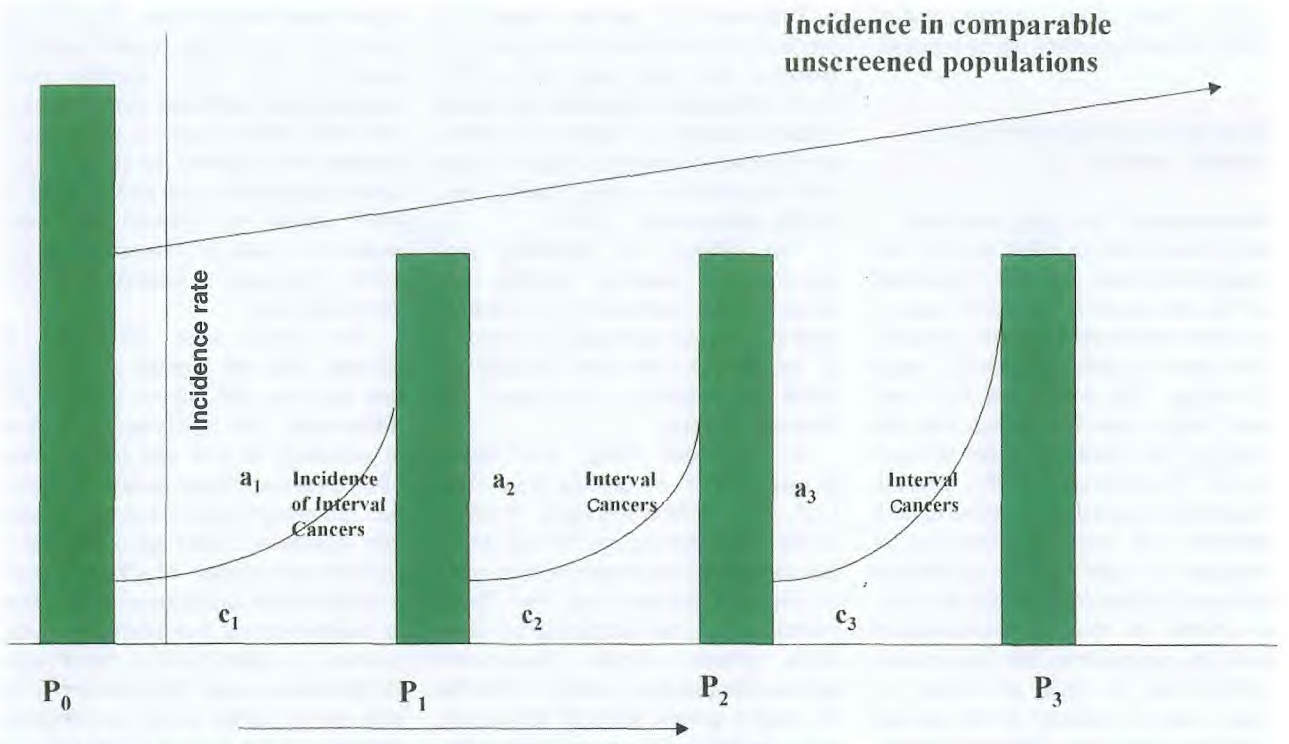


Figure 21 Cancers occurring in a population screened periodically, at a screening interval, T
P_0, cancers detected at prevalence screen; P_i, cancers detected at the 'i'th incident screen; c_i, interval cancers diagnosed in the 'i'th inter-screening interval; a_i, incidence gap in the 'i'th inter-screening interval

during this interval that would have occurred in the absence of screening, programme sensitivity being 1 minus this proportion. Programme sensitivity thus gives a measure of the proportion of incident cases that would be diagnosed by screening among women who are screened according to the programme schedule. The denominators can be woman–years among the screened, invited or the target population. The last results in a measure that is relevant to organized programmes and is comparable with overall mortality reduction in the target population.

As described in the previous section, cancers detected at screening represent a biased sample of cancers in the population. Those detected at the first screening test, the prevalence screen, will be the most biased, as lesions with a sojourn time that is long in comparison with the inter-screening interval will be more strongly overrepresented. After the prevalence screen, the successive screening cycles soon approach a steady state. The cancers diagnosed in one screening cycle, i.e. from immediately after one screening test to immediately after the next, thus approximate the incident cancers during that period, although with a threshold of detection lower than for incident clinical cancers. They thus form a set of cancers from which length bias has been removed and have been termed the 'unbiased set' (Tabár *et al.,* 1992). Their prognostic characteristics can be compared with those of clinically incident cancers, a comparison from which length bias has been approximately removed. Unbiased sets should include nonresponders as well.

A more complex view of cancer

The model shown in Figure 19 describes the operational process of screening, incorporating no information on the biology of the carcinogenic process. Current knowledge of the neoplastic process allows us to distinguish a number of steps, which may begin with mutation at specific genetic loci and other cellular events and continue until cells divide and disseminate throughout the organism. Cancer development is a long process, and not all the steps are necessarily irreversible. In the future, screening modalities may be developed to target these early molecular changes.

In that case, more complex models of the screening process will be required.

Evaluation of screening for breast cancer

Mammography has been evaluated in randomized trials in which women with breast cancer were excluded. These trials are fully discussed in Chapter 4, where it is shown that the effect of early detection of invasive disease can take 5–7 years to emerge. The emergence may take even longer when the women who are screened are under 50 years at entry into the trial (Tabár *et al*., 1997). With the introduction of population screening programmes, the methods developed for evaluation of trials must be adapted to the more complex public health situation. In contrast to trials, population-based screening programmes will take considerably longer to have an impact on breast cancer mortality in the general population. Unlike the participants in trials, the general population have staggered entry into a programme, and women with a pre-exisiting diagnosis of breast cancer are not easily excluded from the overall mortality estimate (Blanks *et al*., 2000b; Jonsson *et al.,* 2001). The conditions for estimating refined mortality rates imply the existence of a cancer register and linkage to it of screening data which is not prevented by data protection legislation. Therefore, predictive measures of the process of cancer screening based on short-term outcomes are useful for evaluating the potential of a programme to make long-term reductions in mortality that are quantitatively comparable to those seen in randomized trials. Short-term parameters for this purpose that have been partially validated as accurate predictors of long-term reductions in breast cancer mortality include sensitivity and sojourn time distribution, both expressible in terms of the post-screening incidence of interval cancer (Day & Duffy, 1996).

Estimates of benefit based on predicted breast cancer mortality may be useful in the initial stage of a public health screening programme but cannot replace analysis of observed mortality, as discussed at length in Chapter 5 (see also Hakama *et al*., 1999; Blanks *et al*., 2000b; Jonsson *et al.*, 2001).

The efficacy of screening programmes for reducing mortality from breast cancer, particularly by mammography, has been analysed in a number of randomized trials (see Chapter 4), which are referred to throughout the following sections.

In December 1963, the Health Insurance Plan of Greater New York, USA, had 490 000 members, of whom 80 000 were women aged 40–64. About two-thirds were employees of local, state or Federal agencies and their family members. The next largest group were union groups outside Government service (Shapiro *et al.*, 1966). In 23 of the 31 medical groups, about 62 000 women were randomized to annual mammography screening and clinical breast examination for 4 consecutive years. Randomization was done by pair-matching by age, size of insured family and employment group through which the family had joined the Plan. Sixty-seven per cent attended the first round. There were no differences between attenders, a 10% sample of non-attenders and a 10% sample of controls with respect to age, socioeconomic status and histories of pregnancy and breastfeeding (Shapiro *et al.*, 1988a). This study is referred to as the 'Health Insurance Plan study'.

In Edinburgh, Scotland, between 1978 and 1981, 87 general practitioners' practices covering 44 268 women aged 45–64 years, were randomized for a breast cancer screening trial (Alexander *et al.*, 1999). The 22 926 women in the intervention group practices were invited to participate in a screening programme, which included clinical breast examination every year and two-view mammography every second year. The 21 342 women in the control group practices received only usual medical care. Subsequently, additional eligible women who joined these practices and existing patients who reached 45 years of age were recruited into two further cohorts: 4867 women in 1982–83 and 5499 women in 1984–85 (Alexander *et al.*, 1999). This study is referred to as 'the Edinburgh trial'.

Two trials were conducted in Canada, one with women aged 40–49 and the other with women aged 50–59 (Miller *et al*., 1981). Women randomized to screening in both age groups were offered annual clinical breast examination and mammography and were taught how to practise breast self-examination. Control women aged 40–49 were given a single clinical examination, taught how to practise breast self-examination and received a questionnaire every year. Control women aged 50–59 were offered only annual clinical breast examinations and were taught how to practise breast self-examination, as the objective was to evaluate the contribution of mammography over and above that of clinical breast examination and breast self-examination. Women were eligible for the trials if they had not had breast cancer, had had no mammogram in the previous 12 months, were currently not pregnant and completed a questionnaire giving full identification and data on risk factors for breast cancer (Miller *et al.*, 1981). Before randomization, all participants gave written informed consent and were told that they had a 50% chance of having a mammogram. They then received a screening clinical breast examination (and instruction in breast self-examination), and the findings were recorded. While the participant remained in the examining room, the examiner went to receive the results of randomization from the centre coordinator and then told the participant whether she would receive mammography. A total of 50 430 women aged 40–49 and 39 405 women

aged 50–59 were enrolled (Miller *et al.*, 1992a,b).

Several trials have been conducted in Sweden, and these are summarized below. The trials have been the subject of two overview analyses (Nyström *et al.*, 1993, 2002).

In the first of two trials in Malmö, Sweden, all women born between 1908 and 1932 were identified from the population register and randomized by a computer program within each birth year cohort. The lists were divided, the 21 088 women on the first half being invited and the 21 195 on the second serving as controls (Andersson *et al.*, 1988). Women were invited to screen–film mammography alone in the first two rounds, with two views (cranio-caudal and medio-lateral oblique), and either both views or only the oblique view, depending on the parenchymal pattern, in the subsequent rounds, every 18–24 months. A single medio-lateral oblique view was taken for women whose breasts were mainly fatty on mammography, and both views were taken for women with dense breasts. After August 1978, the investigators aimed to continue to recruit women who attained the age of 45 years and to randomize them to either receive or not receive an invitation to mammography. In this second trial, 17 786 women born in 1933–45 were ultimately recruited, with 9574 in the invited group and 8212 in the control group. Owing to financial restraints, it was not possible to include one birth-year cohort every year. The randomization and screening procedures were the same as in the first trial, and recruitment continued up to 1990 (Andersson & Janzon, 1997). These trials are referred to in this handbook as the first and second Malmö trials.

In 1975, the Swedish National Board of Health and Welfare invited five county councils to start mammography screening. Two counties, Kopparberg and Östergötland, accepted the invitation. Women in this trial were randomized by cluster within geographical areas (municipalities, parishes, tax districts). The sparsely populated municipalities in the county of Östergötland were grouped pairwise with respect to the size of the population and geographical characteristics, adjacent municipalities being constituted into pairs, as they were considered to be similar in most respects. The more populated municipalities of Linköping, Norrköping and Motala were split into six, eight and two clusters, respectively, of similar size, creating three, four and one pairs, respectively, in order to increase the number of clusters. The clusters were allocated to invitation or a control group at a meeting of the county council by tossing a coin. A total of 76 617 women aged 40–74 were randomized to mammography or usual care (Nyström *et al.*, 2002). In the County of Kopparberg, the invited group was planned to be twice as large as the control group. Thus, triplets of geographical areas were identified by dividing each block into three units of roughly equal size, two of which were randomized to receive screening and one to the control group. A total of 56 782 women aged 40–74 were randomized (Tabár *et al.*, 1985). In this handbook, this trial is referred to as 'the Two-county study'.

A trial was performed in the south-eastern part of Greater Stockholm, Sweden, in which about 60 000 women aged 40–64 years in March 1981 were randomized by day of birth to be invited for mammography or to a control group (Frisell *et al.*, 1986). Women born on days 1–10 and 21–31 of the month were invited for screening (total, 40 318), and women born on days 11–20 to the control group (about 20 000). In the overview of Nyström *et al.* (2002), women born on day 31 were not included, and the totals analysed were 39 139 in the intervention group and 20 978 in the control group. In this handbook, this study is referred to as 'the Stockholm trial'.

Between December 1982 and April 1984, all 51 611 women born between 1923 and 1944 and living in the city of Göteborg, Sweden, were randomized to mammographic screening or a control group, of whom 25 941 were aged 39–49. Randomization was by cluster on the basis of date of birth for the cohorts born in 1929–35 and by individual birth date for those born in 1936–44 (Bjurstam *et al.*, 1997). In order to be able to re-screen women every eighteenth month, despite a fixed capacity for mammography, the ratio of women randomized to the invited and the control group was 1:1.2 in the age group 39–49 years and 1:1.6 in the age group 50–59 years. In this handbook, this study is referred to as 'the Göteborg trial'.

In addition to the randomized trials described above, the Finnish national programme was evaluated after randomization. The programme was begun in 1987, when the Finnish National Board of Health recommended identification of women aged 50–59 years and invitation to screening every second year. The Finnish Cancer Society established 11 mammography centres, and local municipalities, responsible in Finland for public health services, were entitled to establish an arrangement with one of these screening centres. In 1987, 84% of the municipalities made arrangements with the Cancer Society and followed the guidelines of the National Board of Health. The programme was introduced gradually, and decisions about adding cohorts were taken at random. Thus, in 1987, women born in 1928, 1932 and 1936 were invited to be screened by mammography; in 1988, women born in 1930, 1934 and 1938 were invited, and in 1989 women born in 1931 and 1937 were invited. This facilitated comparison of these birth cohorts, considered to be the study cohorts, with the birth cohorts invited after 1990, considered to be the control cohorts (Hakama *et al.*, 1997).