

## Chapter 7

# Summary

### Breast cancer and screening

#### World-wide burden

Breast cancer is the commonest cancer among women in both high-income and low-income countries, accounting for 22% of the 4.7 million new cases of cancer occurring annually among females worldwide. Improvements in treatment and possibly breast screening by mammography have reduced mortality from breast cancer in high-income countries, but the risk continues to increase in eastern Europe and Latin America. Substantial improvements in survival have been reported in high-income countries such as the USA, where the prevalence of breast cancer is estimated to be 1.5% of the female population, whereas survival from this cancer in middle- and low-income countries remains poor, mainly because of late presentation of cases.

#### Biology, pathology and natural history

Breast cancer appears to be a heterogeneous disease. The introduction of mammographic screening has altered the range of benign lesions and the patterns of neoplastic entities that are removed surgically. In general, there are three categories of breast abnormalities: benign conditions, in-situ conditions and invasive cancer. Benign conditions are associated with a risk for breast cancer ranging from one- to fivefold, depending on the degree of epithelial proliferation and atypia. In-situ lesions are lobular or

ductal. Lobular carcinoma *in situ* is associated with an increased risk for invasive breast cancer but is usually an incidental finding and is not generally detected by mammography. Although data on the natural history of ductal carcinoma *in situ* are limited, it is likely that poorly differentiated cytonuclear lesions (high grade) are associated with a significantly higher risk for development of invasive carcinoma than well-differentiated cytonuclear lesions (low grade). High-grade lesions appear to be more biologically aggressive, with a higher rate of recurrence after breast-conserving surgery. Low-grade ductal carcinoma *in situ* is associated with low-grade invasive cancer, which is generally characterized by indolent behaviour and a good prognosis. Molecular markers may become available that will improve evaluation of prognosis.

Reliable classification of in-situ and invasive breast cancers provides important clinical information and can contribute to the evaluation and quality assurance of breast cancer screening programmes. The grading of in-situ cancers is evolving, and standardized staging of invasive cancers has become possible.

#### Conceptual considerations

The main concept in cancer screening is that detection of early disease will make it possible to reduce mortality, because treatment at early stages is more effective than treatment at later stages. The purpose of modelling the screening process is to identify the characteristics of both a screening test and a screening

programme that will determine the extent to which cancers are detected earlier, and thus the potential for reducing mortality. A model is presented, which is based on the assumption that the aim of screening for cancer is to detect lesions that, if left untreated, would progress to clinical cancers. The definitions of sensitivity, specificity and positive predictive value are therefore based on the proportion of cancers that would otherwise be diagnosed clinically during some specified period after screening but which are diagnosed at screening. The model allows for the considerable heterogeneity among cancers indicated by increasing knowledge of tumour biology. This heterogeneity is expressed partly as variation in the preclinical detectable phase, which results in variation in the potential lead time of lesions. The models can be used to identify intermediate outcomes of a screening programme that predict future reductions in mortality from the cancer in question and, as such, are valuable monitors.

### Screening techniques

#### Screening mammography

Modern mammography machines are equipped with devices to reduce scatter, automatically control exposure and optimize the quality of the image in relation to the dose of radiation. The mean absorbed dose to the average-sized breast is in the order of 1.0–2.0 mGy. The sensitivity and specificity of mammography depend on several factors,

including the density of the breast parenchyma, which in turn is related to age, parity, menopausal status and use of hormone replacement therapy, and technical variability. Sensitivity is also determined by the number of image projections used. Variability in interpretation by readers can be partly offset by training and by double reading of films. Continuing training and monitoring of the imaging process is a crucial part of a quality assessment programme for mammographic screening.

### Other and emerging imaging techniques

Many techniques have been suggested for breast cancer screening. Although several of them hold promise, a systematic review showed that few have been used to screen populations, and the studies were generally small and of poor quality, so that the evidence for the following statements is weak. A combination of ultrasound and mammography may increase sensitivity, especially in women with radiographically dense breasts, but with a concomitant reduction in specificity. Magnetic resonance imaging is more sensitive than mammography in women at high risk for breast cancer but has less specificity. Computer-aided diagnosis may improve sensitivity when used in combination with conventional mammography, although it is unclear whether the improvement is greater than with other techniques, such as double reading and special training of film readers. The role of computer-aided diagnosis in specificity is unclear. The sensitivity of full-field digital mammography may be similar to that of film mammography. Only one small study has been reported of use of positron emission tomography in screening, and none has been reported for computed tomography scanning, magnetic resonance spectroscopy, scintimammography, electrical impedance, infrared spectroscopy, light scanning or recent thermography.

### Clinical breast examination

No one technique for clinical breast examination has been shown to be better than any other for breast cancer screening. The technique generally recommended involves visual examination and systematic palpation of the entire breast and regional axillary nodes. The sensitivity of clinical breast examination alone in large studies ranged from 55% to 70% and the specificity from 85% to 95%.

### Breast self-examination

Women often find their own breast cancers. Detailed protocols for breast self-examination have been designed, and competence in the practice has been evaluated by use of silicone models of the breast. Training and reinforcement improve the quality and increase the frequency of use of breast self-examination. While many programmes have been designed to promote breast self-examination, a minority of women practise it and fewer do it well.

## Use of breast cancer screening

### Delivery and uptake of screening

Breast cancer screening is delivered in a variety of ways, including organized programmes and 'opportunistic' activities, which involve referral to mammography facilities by clinicians and self-referral by women themselves. Organized programmes include an administrative structure responsible for implementation, quality assurance and evaluation. Most programmes emphasize mammography. The characteristics of screening in various regions are summarized below.

#### Europe

Organized breast cancer screening programmes were first established in northern Europe and the United Kingdom. Currently, screening is done through organized screening programmes in 19

countries, although opportunistic screening co-exists. Seven of the programmes are organized nationally, nine are organized regionally, and three are pilot programmes. The programmes target at least women aged 50–69, but some extend invitations to women aged up to 74 or under 50. In most of the programmes, women are invited to mammography about every 2 years; in the United Kingdom, women are invited every 3 years. Seven countries offer clinical breast examination in addition to mammography in their screening policies

The proportion of women with access to organized screening programmes varies markedly, from 2% in the German pilot programme to nearly 100% in six countries. Quality review is extensive, following European or national guidelines for addressing the technical quality of mammography, external and internal control, recall rates and cancer detection rates and a wide range of other relevant indicators.

#### The Americas

In Canada, screening is done primarily through a nationally organized programme that is funded and administered at provincial level, targeting women aged 50–69. Although all women have access to screening, 79% of those aged 50–69 reported ever having had mammography, and 54% reported having had one within the previous 2 years. The technical quality of mammography is reviewed within programmes according to national standards.

In contrast, in the USA, screening is primarily opportunistic, and few organized programmes exist. The recommendations of the Preventive Services Task Force now include mammography for women aged 40–69. Assessment of mammography use in a state-based telephone survey showed that 85% of women over 40 had ever had a mammogram, and 71% had had one in the previous 2 years. Quality assurance is

nationally based and focuses on certification of mammography facilities.

Mammography is available on demand in Latin America and the Caribbean; however, most countries report either no policy regarding breast cancer screening or policies that may not completely reflect the available scientific evidence. No population-based estimates of mammography use are available.

### **Oceania and Asia**

Australia and New Zealand have organized national mammographic screening programmes; a few other countries have initiated local screening, usually not based on mammography. Screening in the organized programmes is targeted at women aged 50–69 in Australia and 50–64 in New Zealand and involves invitation for a mammogram every 2 years. In Australia, women aged 40–49 and  $\geq 70$  years may also attend, but they are not systematically invited. In Australia in 1997–98 and in New Zealand in 1999–2000, 54% of all eligible women had had mammograms in the previous 2 years. The Australian and New Zealand programmes have management structures for quality assurance and national standards for participation, recall rates, technical radiological performance, cancer detection rates and data monitoring.

### **Behavioural considerations in screening participation**

Women should be fully informed about the potential benefits and harms of periodic screening so that they can decide whether to take part. Most women tend to overestimate both the likelihood of developing breast cancer and the sensitivity and specificity of screening; they vary in their preference for numerical and verbal information about risk, and this information is often not well understood.

Some women are anxious about mammographic screening, primarily

because of fear of an abnormal result. Women experience a moderate increase in anxiety after a false-positive mammogram, although this is usually short.

Factors associated with participation in mammographic screening include: an invitation or reminder to attend within an organized programme, a recommendation from a doctor to attend, good understanding of the benefits of mammographic screening, a belief that breast cancer can be treated, a perception of personal risk, moderate anxiety about breast cancer and having had other preventive health interventions.

The effects of a number of intervention strategies on participation have been studied, including programmes targeting individual women, community strategies, health care provider programmes and strategies for special groups. Most of these strategies were found under trial conditions to be effective to some extent in increasing participation; however, the feasibility and cost-effectiveness of these strategies as part of routine programme implementation is unknown.

## **Efficacy of screening**

### **Methodological and analytical issues in assessing efficacy**

The efficacy of screening is best evaluated by means of randomized screening trials. Such trials, with mortality from the cancer of interest as the end-point, avoid selection bias and the biases associated with studying survival after diagnosis, including lead-time bias, length bias and overdiagnosis bias. Trials must be planned and conducted with attention to the necessary quality standards, particularly in the areas of randomization, confirmation that balance is achieved by randomization (especially if cluster randomization is used), delivery of the screening intervention, participation by the intervention group and little contamination from screening in the control

group, comparison of cases in the two arms of the trial with regard to early indicators of an effect of the intervention such as tumour size and nodal status, treatment according to stage of detection applied equally in both groups and adequate documentation of the study end-point, preferably after an independent review of cause of death by persons unaware of the allocation of the woman to intervention or control. Observational studies of screening, such as cohort and case-control studies, may give biased measures of effect because of self-selection of women for screening. There are no certain ways of eliminating this bias.

### **Conventional screening mammography**

The screening modality used mainly as a public health intervention at present is mammography alone. The Working Group therefore focused its attention on trials in which the efficacy of mammography alone was compared with no screening.

Of the 10 randomized trials of breast cancer screening, the effect of invitations to mammography alone was compared with that of usual care in six studies, all conducted in Sweden. In two of these, women were randomized by cluster, while various forms of individual randomization were used in the others: two according to randomly ordered birth cohort and the other two by date of birth, either exclusively or in part. Various analytical approaches confirmed that these processes achieved balance. In addition, for a short period at the beginning of the Finnish national programme, women born in even-numbered years were invited to be screened. This is equivalent to randomization, and the results of this experience were incorporated into the evaluation of screening for women aged 50–69.

The findings from the latest follow-ups for women aged 50–69 in the five trials of mammography alone that

included this age group and the Finnish programme gave a combined rate ratio of 0.75 (95% confidence interval, 0.67–0.85) and displayed no heterogeneity.

The findings for women aged 40–49 (43–49 in one trial and 45–49 in another) in the six trials of mammography alone that included this age group gave an overall rate ratio of 0.81 (95% confidence interval, 0.65–1.01). It is uncertain how much of this effect could have been due to screening after the age of 50.

The possibility of the introduction of bias into the results of the studies of screening with mammography alone by a range of methodological factors was considered. The available evidence suggested that none, if any, bias was present that could have had a sufficiently large effect to affect the overall rate ratios appreciably.

The other trials involved combined screening with mammography and clinical breast examination. In one, conducted in the 1960s in New York, USA, the results for both age groups were similar to those in the trials with mammography alone. One, conducted in Edinburgh, Scotland, as a randomized component of the Trial of Early Detection of Breast Cancer and involving combined mammography and clinical breast examination, was based on cluster randomization. There was evidence that the randomization had failed, as there were appreciable differences between the two groups in distribution by social class and mortality from causes other than breast cancer. The Working Group was not convinced that the adjustments undertaken in the analysis of the trial would satisfactorily have removed any bias due to differences between the two groups.

Of the two trials in Canada, both of which involved individual randomization of volunteers, one addressed combined screening with mammography and clinical breast examination of women aged 40–49 in comparison with usual

care. The women were also taught breast self-examination. After an average of 13 years of follow-up, there was no evidence of a reduction in mortality from breast cancer, although the confidence interval was compatible with the overall estimate of effect in this age group in the trials of mammography alone.

The other trial in Canada, of women aged 50–59, was a comparison of screening with mammography plus clinical breast examination with clinical breast examination alone. Thus, its results do not allow a direct evaluation of the efficacy of screening for breast cancer with mammography alone.

In addition to the trials, there have been one quasi-experimental study, one cohort study and four case–control studies, conducted independently of the trials. In general, the observational studies showed greater reductions in the relative risk for death from breast cancer than the trials. This difference has often been attributed to the fact that observational studies address the effect of attendance for screening rather than that of invitation to screening, as is measured in trials. However, observational studies have an inherent potential for bias, due, for example, to self-selection of women for screening, which would make such interpretation inappropriate. Estimates of efficacy should rather be based on the results of trials, after adjustment for non-participation and contamination. By making such adjustments, the Working Group estimated that attendance for screening would reduce mortality from breast cancer by about 35%.

Various frequencies of screening were used in the trials, ranging from 12 to 33 months. In view of the small number of trials, which also had many other differences, it is impossible to assess the effect of screening frequency on the reduction in mortality. One subsequent randomized trial was designed to compare the effect of annual versus three-yearly screening on the size, stage and

grade of tumours. Predictive models based on these data suggest that the effect of shortening the screening interval is modest.

The conclusions of the Working Group differ from those of the review published by the Cochrane Collaboration. In particular, the Group disagreed with the exclusion in that review of several of the randomized trials carried out in Sweden.

### Clinical breast examination

The efficacy of screening by clinical breast examination alone in reducing mortality from breast cancer has not been demonstrated in randomized controlled studies. A case–control study and an ecological study in Japan provided very weak evidence for a reduction in mortality in women screened by clinical breast examination as compared with no screening. One randomized controlled trial showed similar rates of mortality from breast cancer in women screened by clinical examination alone and by a combination of clinical examination and mammography.

### Breast self-examination

Among women who present clinically with breast cancers, the tumours detected in those who practise self-examination tend to be smaller and to be associated with longer survival than those in women who do not examine themselves. Cohort and case–control studies provide some evidence for a reduction in the risk for death from breast cancer among women who practise breast self-examination frequently and competently. Randomized trials in the Russian Federation (of which only one of two components has been reported) and in China showed that women who were taught breast self-examination were more likely than women in the control groups to detect benign breast lesions but not more likely to detect breast cancers at a less advanced stage of progression. Neither the trial in the

Russian Federation, where participation was relatively limited, nor the trial in China, where participation was high, showed a reduction in mortality from breast cancer among women taught this technique.

### Women at high risk

Women who carry mutations in either the *BRCA1* or the *BRCA2* gene have a very high lifetime risk for breast cancer, and many clinicians recommend annual mammographic screening of carriers of such mutations, beginning some time between the ages of 25 and 35. It has not yet been proven that screening of this predisposed group by mammography reduces their mortality from breast cancer, and no randomized trials with mortality as the end-point have been conducted in this group of women. Because of their high risk for cancer, both the prevalence of cancer at screening and the positive predictive value of the screening test are higher than in other women. Because the *BRCA1* and *BRCA2* genes participate in the repair of radiation-induced DNA breaks, it has been suggested that women who carry these mutations are at greater risk for radiation-induced breast cancer than are women in the general population; however, no relevant data are available. The sensitivity of magnetic resonance imaging has been reported to be greater than that of screening mammography for women at high risk because of a *BRCA* mutation or a family history. These studies, however, were based on small numbers of women.

### Effectiveness of population-based screening

#### Implementation of population-based screening in accordance with results of screening trials

The results of randomized trials of mammographic screening compared with no intervention have been used as the basis

for centrally organized screening programmes, to decide the age range of women to be screened and the screening interval. All national screening programmes cover at least women aged 50–64, and all programmes involve an interval of 3 years or fewer. Older and younger women are invited in some countries. Other components of a screening programme, such as the number of film readers and the number of mammographic views, are based largely on considerations other than the results of trials.

#### Indicators of the effectiveness of population-based screening programmes

The basic indicator used for effectiveness is the standardized mortality ratio (SMR). From the point of view of public health, a relative measure such as the SMR may be an incomplete indicator; absolute measures will provide additional information on effectiveness. In none of the national mammographic screening programmes has a reduction in mortality from breast cancer of the order demonstrated in the randomized trials yet been observed. If such a reduction is achievable in practice, it will take many years to occur.

Indicators of performance can be used as a basis for corrective action to a screening programme in the early stages and can be used to predict whether a reduction in mortality is likely to be found in the long term. These indicators include measures of coverage, participation, age-specific or age-standardized rates of detection of cancer and rates of detection of advanced disease and interval cancers (by stage). Predictions of mortality reduction can be based on modelling, and several techniques with various assumptions can be used and validated by comparison with the results of randomized experiments. The microsimulation screening analysis (MISCAN) model has been used for several populations.

Intermediate indicators and surrogate measures are important in order to obtain an early estimate of effect. They are necessary but not sufficient for an effective screening programme. The interval cancer rate is a useful determinant of programme sensitivity and the rate of advanced disease of programme effectiveness. Both are predicated on the availability of cancer registration in the target population; the identification of interval cancers also requires linkage of data sources into a coherent information system, and measurement of the rate of advanced cancer also requires that the cancer registry records clinical stage.

In the few instances in which assessment of advanced cancer rates has been possible, screening appears to have been followed by a decline in the rates of advanced disease (albeit more than offset by the large numbers of early and in situ cancers detected). In all the programmes examined, the decreases in advanced disease rates have been smaller than predicted from the data of the Two-county study in Sweden.

Given the natural history of the disease and the long implementation period of national programmes, it is too early to expect a substantial reduction in breast cancer mortality. Evidence from the United Kingdom has shown that the recent substantial declines are probably due to multifactorial causes, and the precise roles of screening and other factors, including improved therapy, are hard to determine. This is even truer in areas that depend only on overall rates of breast cancer mortality for evaluating effectiveness, as the quality of screening and the extent of information are likely to be correlated.

Cases of breast cancer diagnosed before the start of screening contribute to the mortality rates, and removal of these cases results in a better estimate of effect. Such estimates of 'refined' mortality require the existence of a cancer registry and the possibility of linkage to data on screening. Refined mortality should be

estimated for screened and unscreened populations to ensure comparability. Furthermore, cancer registration with data on treatment is likely to be the only means for differentiating the confounding effect of changes in treatment from the effect of screening.

Studies on effectiveness, including those based on modelling, have so far resulted in estimates of 5–10% reductions in mortality in the target population due to screening. The estimates of refined mortality have been higher, around 20%, closer to the effect indicated by the screening trials. In terms of prolongation of life, the effect per screen remains small.

### Hazards of screening

#### **False-positive mammograms:**

False-positive results are inevitable in screening. However, the rate varies dramatically from one area to another; it is particularly high in the USA. Depending on the setting and the frequency of examinations, the cumulative risk of a woman who receives a false-positive result after completing a screening programme can be extrapolated to be as low as 2% or as high as 50%. False-positive results increase health care use associated with screening. Women experience considerable anxiety after being told they have a positive result; this effect is largely transient and is an accepted part of screening for most women. The greatest opportunity for reducing false-positive results is in improving radiological interpretation.

#### **Overdiagnosis**

'Overdiagnosis' is the term used to describe the detection of cancers that would never have been found without screening. Patients who have such indolent cancers experience only harm: the anxiety associated with a cancer diagnosis and the complications of therapy. Overdiagnosis increases the cost of screening and complicates evaluation of the programme.

There is evidence of some over-diagnosis of breast cancer in the randomized trials of mammography and from population-based incidence rates. From 5 to 25% of cancers detected by mammography may represent over-diagnosis. The finding of a substantial breast cancer reservoir suggests that perhaps the most pressing challenge for breast cancer screening is to determine which lesions should be treated.

#### **Ductal carcinoma in situ**

Evidence from clinical studies suggests that a proportion of ductal carcinomas *in situ* will progress to invasive cancer. How small this proportion is for non-palpable lesions detected by mammography is, however, less clear. The results of the trials in Canada suggested that detection of ductal carcinoma *in situ* by mammography and its subsequent treatment did not lead to a reduction in the incidence of invasive cancer within 11 years. As current work suggests that the prognosis of ductal carcinoma *in situ* differs according to its nuclear grade, screening might offer greater benefit to women with some types of lesion than to others. It is an open question, therefore, whether the potential benefits of detecting and treating ductal carcinoma *in situ* outweigh the harmful effects of treatment (anxiety, operations, radiotherapy).

#### **Radiation**

Exposure to radiation is a known risk factor for breast cancer. The mean absorbed dose of radiation to the breast during mammography is now generally below 3 mGy per screen, and the dose of radiation to the thyroid and other organs is assumed to be negligible. The risk for radiation-induced breast cancer decreases with age and is particularly low for women after the menopause. In a model based on the assumption of a linear relationship between risk for breast cancer and dose of radiation, the number of deaths from radiation-induced breast cancer during the remaining life

span when screening is begun at the age of 50 is estimated to be 10–50 per million in regularly screened women (10–20 screens, 2–5 mGy per screen). These numbers can be compared with the 30 000–40 000 deaths from breast cancer over a lifespan after 50 years of age per million women in the whole population, of which some 10 000–15 000 may be preventable by screening. If screening is begun at the age of 40, the number of radiation-induced breast cancers is estimated to be 100–200 per million regularly screened women.

In relation to the expected benefit, the risk is negligible when screening is started at the age of 50 but is higher when screening is begun between the ages of 40 and 50. The risk for radiation-induced breast cancer should be taken into account if screening is started at a younger age.

### Cost-effectiveness of population-based screening

In many countries, it has become routine policy to assess the costs of new, promising health care interventions in relation to their expected benefits, before implementation. The screening policy for breast cancer that is most cost-effective in a particular country depends on various factors, including the incidence of breast cancer, its stage distribution and mortality rate, the expected quality of the screening programme, the national health care setting and economics. Although the final ratio, cost per life-year gained, is considered most important by some, the hierarchy in cost-effectiveness analyses is, first, to assess the benefits (breast cancer mortality reduction and life-years gained), second, to assess the possible harm and benefits other than reduction in mortality (quality of life) and, finally, to weigh these against induced costs and possible savings.

Given the evidence about reduction of mortality from breast cancer in randomized trials of breast cancer screening, screening programmes for women aged 50–69 at a 2- or 3-year interval are expected to be cost-effective in high-incidence countries with well-organized programmes. National reductions in breast cancer mortality may be of the order of 10–20%; women who do not die from breast cancer may gain approximately 15 years of life, and the cost-effectiveness ratio is 3000–8000 euros per life-year gained. Very high referral rates of about 10% unfavourably influence cost-effectiveness ratios, as does the delicate balance between favourable and unfavourable effects. In

general, the harm inflicted on a group of screened women is less than the benefits achieved by some part of the same screened group. Correction for all anticipated unfavourable effects in terms of quality-adjusted life-years gained may diminish the total number of life-years gained by 5–15%.

The most important cost elements to consider are the cost of screening and the cost of treating advanced disease. Savings in the cost of screening of up to 30% may be achieved by the reduction in the cost of treating advanced disease, but breast cancer screening will always lead to substantial additional cost for a country. It should be compared with other health care priorities, preferably by

cost-effectiveness ratios too. Low-risk and low-income countries are likely to give higher priority to other activities.

The marginal cost-effectiveness of expanding a programme to younger women (40–49) greatly depends on its effect on reducing breast cancer mortality as estimated from randomized controlled trials. Under the assumption of less or relatively low benefits of screening younger women, it would be more cost-effective to increase the upper age limit to 74 or to narrow the screening interval from 3 to 2 years for the age group 50–69, rather than expand the programme to women aged 40–49.