5. EFFECTIVENESS OF BREAST CANCER SCREENING

This section considers measures of screening quality and major beneficial and harmful outcomes. Beneficial outcomes include reductions in deaths from breast cancer and in advanced-stage disease, and the main example of a harmful outcome is overdiagnosis of breast cancer. The absolute reduction in breast cancer mortality achieved by a particular screening programme is the most crucial indicator of a programme’s effectiveness. This may vary according to the risk of breast cancer death in the target population, the rate of participation in screening programmes, and the time scale observed (Duffy et al., 2013). The technical quality of the screening, in both radiographic and radiological terms, also has an impact on breast cancer mortality. The observational analysis of breast cancer mortality and of a screening programme’s performance may be assessed against several process indicators. The major indicators of both the screening process and the clinical outcome, and the associated analytical methodologies, are described below.

5.1 Indicators for monitoring and evaluating effectiveness

5.1.1 Performance indicators

As a general principle, the most important indicator of the effectiveness of a screening programme is its effect on breast cancer mortality. Nevertheless, the performance of a screening programme should be monitored to identify and remedy shortcomings before enough time has elapsed to enable observation of mortality effects.

(a) Screening standards

The randomized trials performed during the past 30 years have enabled the suggestion of several indicators of quality assurance for screening services (Day et al., 1989; Tabár et al., 1992; Feig, 2007; Perry et al., 2008; Wilson & Liston, 2011), including screening participation rates, rates of recall for assessment, rates of percutaneous and surgical biopsy, and breast cancer detection rates. Detection rates are often classified by invasive/in situ status, tumour size, lymph-node status, and histological grade.

Table 5.1 and Table 5.2 show selected quality standards developed in England by the National Health Service (NHS) (Wilson & Liston, 2011; Department of Health, 2013) and in the USA by the Agency for Health Care Policy and Research and endorsed by the American College of Radiology, respectively (Bassett et al., 1994; D’Orsi et al., 2013). Similar sets of standards exist for screening in Australia, Canada, and Europe (National Quality Management Committee of BreastScreen Australia, 2008; Perry et al., 2008; CPAC, 2013) (see Section 3.2). The programmes specify standards – related mainly to the screening process and not directly to technical
Table 5.1 Minimum quality standards and targets considered in the National Health Service breast screening programme in England

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Standard</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendance at screening</td>
<td>≥ 70%</td>
<td>80%</td>
</tr>
<tr>
<td>Invasive cancers detected, prevalent screen</td>
<td>≥ 3.6/1000</td>
<td>≥ 5.1/1000</td>
</tr>
<tr>
<td>Invasive cancers detected, incident screen</td>
<td>≥ 4.1/1000</td>
<td>≥ 5.7/1000</td>
</tr>
<tr>
<td>In situ cancers detected, prevalent screen</td>
<td>≥ 0.5/1000</td>
<td>None specified</td>
</tr>
<tr>
<td>In situ cancers detected, incident screen</td>
<td>≥ 0.6/1000</td>
<td>None specified</td>
</tr>
<tr>
<td>Standardized detection ratio</td>
<td>≥ 1.0</td>
<td>≥ 1.4</td>
</tr>
<tr>
<td>Invasive cancers &lt; 15 mm, prevalent screen</td>
<td>≥ 2.0/1000</td>
<td>≥ 2.8/1000</td>
</tr>
<tr>
<td>Invasive cancers &lt; 15 mm, incident screen</td>
<td>≥ 2.3/1000</td>
<td>≥ 3.1/1000</td>
</tr>
<tr>
<td>Mean glandular radiation dose for standard breast</td>
<td>≤ 2.5 mGy</td>
<td>None specified</td>
</tr>
<tr>
<td>Number of repeat examinations (% of total examinations)</td>
<td>&lt; 3%</td>
<td>&lt; 2%</td>
</tr>
<tr>
<td>Recall for assessment (% of prevalent screens)</td>
<td>&lt; 10%</td>
<td>&lt; 7%</td>
</tr>
<tr>
<td>Recall for assessment (% of incident screens)</td>
<td>&lt; 7%</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>Short-term recall (% of screened women)</td>
<td>&lt; 0.25%</td>
<td>≤ 0.12%</td>
</tr>
<tr>
<td>Non-operative diagnosis (% of cancers)</td>
<td>≥ 90%</td>
<td>≥ 95%</td>
</tr>
<tr>
<td>Non-operative diagnosis (% of DCIS)</td>
<td>≥ 85%</td>
<td>≥ 90%</td>
</tr>
<tr>
<td>Benign biopsies (prevalent screens)</td>
<td>&lt; 1.5/1000</td>
<td>&lt; 1.0/1000</td>
</tr>
<tr>
<td>Benign biopsies (incident screens)</td>
<td>&lt; 1.0/1000</td>
<td>&lt; 0.75/1000</td>
</tr>
<tr>
<td>Interval cancers within 24 months (screened women)</td>
<td>≤ 1.2/1000</td>
<td>None specified</td>
</tr>
<tr>
<td>Interval cancers within 25–36 months</td>
<td>≤ 1.4/1000</td>
<td>None specified</td>
</tr>
<tr>
<td>Percentage rescreened within 36 months</td>
<td>≥ 90%</td>
<td>100%</td>
</tr>
<tr>
<td>Percentage receiving screening result within 2 weeks</td>
<td>≥ 90%</td>
<td>100%</td>
</tr>
<tr>
<td>Assessed within 3 weeks (% of total assessed)</td>
<td>≥ 90%</td>
<td>100%</td>
</tr>
<tr>
<td>Percentage non-operative biopsies with result within 1 week</td>
<td>≥ 90%</td>
<td>100%</td>
</tr>
<tr>
<td>Percentage referred to surgeon receiving surgical assessment within 1 week</td>
<td>≥ 90%</td>
<td>100%</td>
</tr>
<tr>
<td>Percentage admitted for treatment within 2 months of referral</td>
<td>≥ 90%</td>
<td>100%</td>
</tr>
</tbody>
</table>

DCIS, ductal carcinoma in situ.
Adapted from Wilson & Liston (2011) and Department of Health (2013).

Aspects of image quality – that all units should attain, as well as achievable targets at which units should aim.

Table 5.1 pertains to a programme that targets women aged 50–70 years with a maximum screening interval of 36 months in high-incidence countries. In the example in England, two-view mammography is used, and the programme changed from film to digital mammography during 2010–2014.

Minimum standards are specified for screening attendance and detection rates, in particular detection rates of small cancers, which are expected to be high in an effective screening programme. Maximum standards are specified for adverse effects of screening, such as radiation dose, and for rates of interval cancers, repeat examinations, and recalls for assessment. In addition, maximum times to events in the screening, diagnostic, and treatment processes are specified; these are important for the patient’s experience and quality of life, although they do not necessarily reflect clinical or radiological quality.

Some of the criteria and standards are very specific to the programme. For example, the randomized trials of breast screening observe a higher rate of breast cancer detection at the prevalent (first) screen than at incident (subsequent) screens (see, for example, Tabár et al.).
Table 5.2 Minimum quality standards for mammography in the USA

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recall rate for assessment (% of screened women)</td>
<td>&lt; 10%</td>
</tr>
<tr>
<td>Cancer detection rate, prevalent screen (per 1000 screened)</td>
<td>6–10</td>
</tr>
<tr>
<td>Cancer detection rate, incident screen (per 1000 screened)</td>
<td>2–4</td>
</tr>
<tr>
<td>Positive predictive value of recall for assessment</td>
<td>5–10%</td>
</tr>
<tr>
<td>Positive predictive value of biopsy</td>
<td>25–40%</td>
</tr>
<tr>
<td>Proportion of screen-detected cancers in situ or TNM stage 0–I</td>
<td>&gt; 50%</td>
</tr>
<tr>
<td>Proportion of screen-detected node-positive cancers</td>
<td>&lt; 25%</td>
</tr>
</tbody>
</table>

* Values are specified by the United States Agency for Health Care Policy and Research and endorsed by the American College of Radiology. TNM, tumour–node–metastasis staging system of malignant tumours (see Section 1, Table 1.9).
Adapted from Bassett et al. (1994) and D’Orsi et al. (2013).

1992). However, the detection rate standards are expected to be higher for incident screens because these values are based not on observations of a cohort recruited at the prevalent screen and followed up thereafter but on a programme in which prevalent screens usually take place at about age 50 years and incident screens on average at about age 60 years (when the underlying risk of cancer is higher).

Another measure that is used in the United Kingdom is the standardized detection ratio, obtained by comparing the observed detection rates of invasive cancers by age with those of the Swedish Two-County trial (Tabár et al., 1992), on which the United Kingdom breast screening programme was modelled. At present, the standard is almost invariably exceeded (NHSBSP, 2009), probably at least partly due to the fact that breast cancer incidence in the United Kingdom in the 21st century is higher than that in Sweden in the 1970s and 1980s. This example implies that standards should be revised over time, although it has also been observed that lower standards followed by remedial action have conferred substantial improvements in programme performance (Blanks et al., 2002). Wallis et al. (2008) gave a demonstration of how careful surveillance of audit standards can lead to changes in practice and improved performance at the local and national levels.

Indicators such as detection rates are typically part of the monitoring system of most screening programmes, but the actual target values will vary according to the screening regimen, the target population, the underlying incidence in the programme’s location, and possibly aspects of the health-care delivery systems and the medicolegal environment (Klabunde et al., 2001).

Table 5.2 shows selected standards developed in the USA. These standards include acceptable ranges for positive predictive values (PPVs) of recall for assessment and for recommendation for biopsy. They specify that the proportion of cases recalled for assessment that result in diagnosis of cancer should be 5–10%, and that the proportion of biopsies that result in diagnosis of cancer should be 25–40%. These are powerful measures of the process since they reflect detection rates, recall rates, and biopsy rates.

(b) Screening sensitivity and interval cancers

In a screening setting, the prevalence of the disease in screened subjects, expressed as a proportion, is usually very low; a very small number of those screened at each screening round are diagnosed with cancer, whereas thousands of women are screened negative. Typically, in European screening programmes, per 10 000 women screened, about 9500 will have a normal initial result and about 500 will be recalled for further assessment, of whom about 70 will have
breast cancer. After the screen, about 10–30 will present with symptomatic interval cancer.

Components of the quality monitoring data listed above can be useful to estimate some important attributes of the screening programme, notably the specificity and sensitivity (the correct classification of negative and positive subjects) and the PPV. Specificity estimates the false-positives, or the complement of the proportion of screened-negative cases that are recalled for further assessment. The classic definition of test sensitivity is the probability that if the screening test is applied to someone with the disease, a positive diagnosis will result. PPV is the proportion of test-positive subjects who are diagnosed as cases at the end of the screening episode and is a function of the prevalence of the lesion. There are costs, both human and economic, to achieving a good balance of these performance parameters.

Other parameters of cancer detection have been defined by Hakama et al. (2007): test sensitivity, programme sensitivity, and episode sensitivity.

(i) Test sensitivity

In a clinical setting, test sensitivity is usually measured by comparison with a “gold standard”. This is rarely possible in a screening setting, where the objective of the test is the detection of a lesion in the preclinical detectable phase, and where only those with suspicious initial screening findings receive further investigation. Test sensitivity is the number of cancers detected at a screen divided by the sum of those detected at the screen plus the false-negatives. In principle, the false-negatives can be identified by a radiological audit of the original screening mammograms in those screened negative and subsequently diagnosed with interval breast cancer (Houssami et al., 2006; Perry et al., 2006). This method of estimation involves assumptions about the audit quality, and the audit itself consumes resources, but it is a crucial learning tool and has the potential to improve the programme’s ability to detect early-stage cancers.

In the past, a common convention has been to estimate sensitivity as the number of cancers detected at a screen divided by the sum of those detected at the screen plus the interval cancers arising within 1 year. Two main sources of error have been identified: first, the interval cancers arising within 1 year will include true negatives that have entered the preclinical detectable phase during that year, and, second, they will not exclude those cancers missed at the screen but taking longer than 1 year to arise symptomatically (Day, 1985). The reasoning implies that interval cancers are a mixture of missed and newly arising cancers, which tend to be more rapidly developing tumours. This, in turn, suggests that interval cancers will also be a mixture with respect to the aggressive potential of the cancers. In the epoch of film mammography, test sensitivity was reported to range from 83% to 95%, with the higher values observed for screening women older than 50 years (Mushlin et al., 1998). In the epoch of digital mammography, the difference in sensitivity between age groups may be smaller (Vinnicombe et al., 2009).

(ii) Programme sensitivity

Programme sensitivity may be defined as the proportion of cancers diagnosed among women attending a screening programme or as the proportion of cancers diagnosed in the screening-eligible population. The first definition is the number of screen-detected cases divided by the sum of the screen-detected cancers plus the interval cancers. The second definition includes in the denominator cancers diagnosed among those who were invited but did not attend screening. Programme sensitivity is often described as the ability of the programme to detect cancers. It is generally estimated from steady-state screening, from the numbers of cancers diagnosed at several incident screens (not from prevalent screening).
and the symptomatic cancers occurring in the same number of intervals between screens.

Programme sensitivity depends on the test sensitivity, the screening interval, and (depending on which measure is used) the attendance rate. It is typically estimated to be 50–60% (Anttila et al., 2002; Zorzi et al., 2010). This means that in organized programmes, about half of the cancers in the target population are detected by screening. Of course, this will depend strongly on the rate of participation in screening.

(iii) Episode sensitivity

Hakama et al. (2007) defined episode sensitivity as the incidence reduction in a specified period after screening compared with the expected incidence in the absence of screening, that is \(1 - (P_1/P_0)\), where \(P_1\) is the incidence among the screened subjects in the specified period after screening and \(P_0\) is the expected incidence in the absence of screening (which, in practice, is difficult to estimate).

Taylor et al. (2002, 2004) reviewed estimates of the proportional incidence in the first year of the screening interval, comparing international data published since 1975 and including results from randomized trials and service screening programmes in Australia, Canada, Italy, the Netherlands, Scandinavia, the United Kingdom, and the USA (Health Insurance Plan study). A large variability was reported, with an overall point estimate of the proportional incidence of 18.5% from all randomized trials and 27.3% from service screening programmes, corresponding to episode sensitivity estimates of 91.5% for the randomized trials and 72.7% for service screening.

A pooled analysis in the service screening centres of six European countries (Törnberg et al., 2010) reported a large variation in screening sensitivity and performance, with a proportional incidence of 46% (episode sensitivity, 54%) in the 24 months after screening. The European standards (Perry et al., 2006) were 30% and 50% for the proportional incidence at the prevalent screen and at subsequent screenings, respectively, corresponding to recommended episode sensitivities of 70% and 50%, respectively.

(iv) Interval cancers

Note that all three measures discussed above require an estimation of interval cancer incidence. This illustrates the crucial nature of interval cancers in programme evaluation. Whereas screen detection rates are important, the future cancer risk in those screened negative is at least equally informative about the programme’s ability to detect cancer in the preclinical phase.

Bennett et al. (2011) noted the complexity of the evaluation of interval cancers on a large scale. They analysed 26 475 interval cancers in the NHS Breast Screening Programme (England, Wales, and Northern Ireland) and found a large variability in the regional estimates, with an estimate of a higher level than expected on the basis of the randomized trial experience. The conclusion was that comparison of different programmes is possible only if the methodology used is very thorough and guidelines are agreed upon in advance, with accurate follow-up and homogeneous reporting.

Table 5.1 includes standards for maximum interval cancer rates, that is, rates of symptomatic cancers that are diagnosed after a screen with negative findings and before the next scheduled screen (usually a period of 1–3 years). Together with prompt and nearly complete cancer registration, the interval cancer rate can be a powerful indicator of screening quality (Bennett et al., 2011). The observation that interval cancer rates were very high in the early years of the United Kingdom programme in the East of England prompted a radiological audit, which consisted of re-reading previous screening mammograms, both of interval cancers and of non-cancers, without knowledge of the diagnostic result (Day et al., 1995). This identified issues of sensitivity, which were later remedied, and served...
as a learning resource for quality improvement in other regions of England (Duncan & Wallis, 1995). Interval cancer rates are now considerably lower in the East of England and similar to those in the rest of the United Kingdom (Bennett et al., 2011; Offman & Duffy, 2012). The radiological audit of advanced disease may be suggested in health-care settings where cancer registration systems do not sufficiently identify interval cancers.

Interval cancer rates can also yield inferences about the effect of changes to the screening regimen. The policy of two-view mammography for incident screens was shown first to increase detection rates (Blanks et al., 2005) and subsequently to reduce interval cancer rates by almost exactly the same absolute numbers (Dibden et al., 2014). The concomitant reduction in interval cancer rates gave some assurance that the increased detection capability was not an over-diagnosis phenomenon.

Estimates and characteristics of interval cancers in national and regional screening programmes have been published, confirming the need for surveillance and improvement of service screening (Ganry et al., 2001; Wang et al., 2001; Hofvind et al., 2006; Bucchi et al., 2008; Domingo et al., 2013a; Carbonaro et al., 2014; Dibden et al., 2014; José Bento et al., 2014; Renart-Vicens et al., 2014).

The relationship between detection modality and tumour characteristics of breast cancers has been investigated ever since the first randomized trials (Duffy et al., 1991). Recently, the renewed interest in interval cases and their radiological classification (Houssami et al., 2006) has enabled the analysis of tumour characteristics by detection mode and interval type in terms of new biomolecular classifications and mammographic breast density at screening. Such analyses, along with recent findings with respect to genetic predisposition, have raised interest in personalized screening (Hall & Easton, 2013). Although personalized screening is not simple to incorporate into existing programmes (Paci & Giorgi Rossi, 2010), such interest does indicate that investigation of interval cancers can inform hypotheses to potentially improve screening policy.

(c) Breast cancer mortality

As noted above, the most telling indicator of the effectiveness of a screening programme is its effect on breast cancer mortality. However, estimating this effect is not straightforward (Duffy et al., 2007; Otten et al., 2008; Broeders et al., 2012; Independent UK Panel on Breast Cancer Screening, 2012). Temporal and geographical comparisons are potentially confounded with other parameters that influence breast cancer mortality; simultaneous temporal and geographical control yields more directly interpretable results (Otto et al., 2003; Olsen et al., 2005). The introduction of breast screening as in Finland, with date-of-birth clusters randomized to receive screening first, yields results that may be interpretable directly as estimates of the efficacy of the programme (Hakama et al., 1997). It is worth noting that such designs do not obviate the need for sufficient follow-up. In absolute terms, in the early years of a programme the adverse effects are enumerable, but the benefits in terms of numbers of breast cancer deaths avoided are not.

Arguably the most important issue for observational evaluation of screening and breast cancer mortality is the diagnostic period. Because of the generally good breast cancer survival rates, unrefined mortality (used hereafter to denote breast cancer mortality regardless of the time of diagnosis) in the epoch of screening will be contaminated by a substantial numbers of deaths from cancers diagnosed before screening was initiated (Duffy et al., 2007). This will tend to bias results against screening. The bias can be avoided by using refined or incidence-based mortality (IBM), where mortality is ascertained specific to the diagnostic period (Olsen et al., 2005; Swedish Organised Service Screening Evaluation Group, 2006).
Breast cancer screening

2006a, b). Alternatively, the bias can be minimized by estimating the mortality effect in a period beginning some years after the start of screening, albeit with some qualifications on interpretation (Duffy et al., 2010).

Epoch of diagnosis also has implications for treatment and management of breast cancer, so that the before–after comparisons of mortality are almost invariably confounded with changes in treatment, as with the expansion in use of adjuvant systemic therapies in the 1980s and 1990s. This is considered further in Section 5.1.2.

Concerns have been expressed with respect to ascertainment of cause of death (Gøtzsche & Jørgensen, 2013). Results suggest that this is not a serious cause of bias (Goldoni et al., 2009; Holmberget al., 2009), partly because the number of women with advanced breast cancer who do not die of breast cancer is limited (de Koning et al., 1992). In any case, it can be addressed by estimating the effect of screening on excess mortality in breast cancer cases, which does not require individual determination of cause of death (Jonsson et al., 2007).

Methods and results in terms of breast cancer screening and mortality are dealt with in more detail in Section 5.1.2, and possible surrogate indicators of breast cancer mortality are considered in Section 5.1.3.

5.1.2 Study designs to assess the effectiveness of screening

(a) General principles

Attempts to estimate exact proportions of recent reductions in breast cancer mortality are subject to difficulties in modelling and interpreting the dynamism of incidence, behaviour, screening policy, treatment policy, and the correlations among these. In addition, there are always difficulties in interpreting directions of causality in changes, particularly in breast cancer incidence.

The main observational methods to assess the effect of screening are: (i) analysis of temporal trends in unrefined breast cancer mortality, reporting annual percentage changes in screening and pre-screening periods and change points when trends are estimated to change in magnitude or direction; (ii) comparison of unrefined mortality rates in screening or invited exposed populations with temporal, geographical, or other demographic control; (iii) the same comparison using IBM; and (iv) case–control studies where women who have died of breast cancer are compared with women who have not, with respect to screening histories before diagnosis of the case. In addition, modelling studies can provide information on outcomes beyond the limits of observational studies. This section outlines the principles and practice of each method, illustrating them with published results. First, two commonly occurring biases, and possible methods for their correction, are described.

(i) Self-selection for screening

Any estimate of the effect of being screened might be biased by factors influencing self-selection, such as the risk of death from breast cancer. In the Swedish breast screening trials, women not attending screening had a 36% higher risk of death from breast cancer compared with the uninvited control group (Duffy et al., 2002a). This was a combination of a lower incidence of breast cancer and a considerably higher case fatality rate (Duffy et al., 1991). A difference of this nature would induce a bias in favour of screening if not addressed by design or analysis.

Cuzick et al. (1997) developed a method to correct for this bias in randomized controlled trials (RCTs), assuming a latent non-attender population in the control group. Duffy et al. (2002a) adapted this for case–control studies and later for other designs of observational studies (Swedish Organised Service Screening Evaluation Group, 2006a). The correction depends crucially
on an estimate of the relative risk of breast cancer death in non-attenders compared with an uninvited population. Although this can be readily estimated within a given trial, in observational studies this is not generally the case. In the past, observational studies have relied on a relative risk estimate of 1.36 from the Swedish trials (e.g. Allgood et al., 2008) and, more recently, on estimates from the target population (Paap et al., 2011). Paap et al. (2011) noted that in the Netherlands, the non-participant population had, if anything, a lower a priori risk of breast cancer death compared with the participant population. Table 5.3 shows the odds ratios (with and without correction for self-selection bias) for breast cancer mortality associated with screening, and the relative risks for non-participants in screening, in five regions of the Netherlands. Those regions with a non-participant relative risk greater than 1 had a corrected odds ratio that was less extreme than the uncorrected one, whereas those regions with a non-participant relative risk less than 1 had a more extreme corrected odds ratio. This leads to the observation that in the organized screening in the Netherlands, self-selection bias appeared to have only a minor effect (Otto et al., 2012a).

Differences in prognosis between attenders and non-attenders could be explained by: a different underlying risk of disease; different help-seeking habits for symptoms, which lead, in turn, to differences in stage at presentation; varying compliance with treatment; or different comorbidities, which have a bearing on outcome (Aarts et al., 2011). Socioeconomic status has been suggested as the major confounder of both outcome and participation in screening (Palli et al., 1986; Aarts et al., 2011), although adjustment for it made almost no difference to the estimated effect of attending screening (Palli et al., 1986).

There is greater uncertainty about the appropriate correction in observational studies with respect to randomized trials when estimating the effect of actually being screened. However, Duffy et al. (2002a) illustrated that the relative risk of breast cancer death may differ a priori between attenders and non-attenders, in ways that are not related to screening and thus completely annul the benefit observed among the screened population. The authors first considered a Swedish case–control study with an uncorrected relative risk of 0.50 for being screened, and then calculated that the a priori risk of breast cancer death among non-attenders would have to be 1.53 to be entirely due to self-selection bias, in a programme with 70% attendance. For a true (i.e. often suggested by trials' meta-analyses) relative risk of 0.80 associated with invitation to screening, the relative

---

**Table 5.3 Odds ratios, with and without correction for self-selection bias, for breast cancer mortality associated with screening in five regions of the Netherlands**

<table>
<thead>
<tr>
<th>Region</th>
<th>Uncorrected OR (95% CI)</th>
<th>RR, non-participants/uninvited (95% CI)</th>
<th>OR corrected for self-selection bias (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.67 (0.42–1.08)</td>
<td>0.64 (0.46–0.90)</td>
<td>0.40 (0.22–0.74)</td>
</tr>
<tr>
<td>2</td>
<td>0.52 (0.38–0.73)</td>
<td>0.77 (0.63–0.93)</td>
<td>0.38 (0.25–0.57)</td>
</tr>
<tr>
<td>3</td>
<td>0.27 (0.12–0.62)</td>
<td>0.92 (0.65–1.30)</td>
<td>0.24 (0.10–0.62)</td>
</tr>
<tr>
<td>4</td>
<td>0.44 (0.32–0.60)</td>
<td>1.08 (0.82–1.43)</td>
<td>0.49 (0.30–0.78)</td>
</tr>
<tr>
<td>5</td>
<td>0.46 (0.30–0.72)</td>
<td>1.08 (0.85–1.37)</td>
<td>0.51 (0.30–0.87)</td>
</tr>
</tbody>
</table>

a Region 1: Bevolksonderzoek Noord-Nederland; region 2: IKA; region 3: Limburg; region 4: Bevolksonderzoek Borstkanker Zuidwest Nederland; region 5: Vroege Opsporing Kanker Oost-Nederland.

b Region-specific estimates of the relative risk of breast cancer death in non-participants compared with uninvited women. CI, confidence interval; OR, odds ratio; RR, relative risk.

Adapted from *Breast*, Volume 23, issue 4, Paap et al. (2014), Breast cancer screening halves the risk of breast cancer death: a case-referent study, pages 439–444, Copyright (2014), with permission from Elsevier.
Breast cancer screening

risk would have to be 1.23. Such reverse calculation of the required size of the bias to annul the result, or to give a result consistent with the trials, may provide some assistance in interpreting the results of observational estimates of the effect of actually being screened.

(ii) Screening opportunity bias

Screening opportunity bias pertains particularly to case–control studies, where controls can only be exposed to screening if they attended their last screen, whereas cases can be exposed to screening if they attended their last screen or were screen-detected (Walter, 2003). This means that if the screens at which any screen-detected cases were detected are included as exposure, there is a bias against screening, and if they are excluded, there is a bias in favour of screening. Duffy et al. (2008) developed a method that estimates the additional opportunity for screening exposure among the cases and yields a correction to the odds ratio for this, obtaining an estimate that lies between the odds ratios including and excluding the detection screen.

(b) Prospective or retrospective cohort analysis of unrefined mortality

A common evaluation technique consists of comparing rates of unrefined mortality (i.e. regardless of time of diagnosis) in a screened versus an unscreened population (whether historical or contemporaneous or both). An early but very clear example of this approach is the estimation of the effect of the NHS Breast Screening Programme in England and Wales by Blanks et al. (2000). The authors fitted age-cohort models to breast cancer mortality data recorded over the period 1971–1989, before the advent of substantial screening coverage, and projected these to estimate the expected mortality in the absence of screening for the period 1990–1998, in which the screening programme was achieving high coverage. The authors compared the observed reductions in mortality with expected rates for the age groups 55–69, 50–54, and 75–79 years. The observed reductions in breast cancer mortality were 21.3% in the age group 55–69 years and 14.9% in the age groups 50–54 years and 75–79 years, age groups that might reasonably be expected to be unaffected by breast screening. The estimated reduction in breast cancer mortality associated with the NHS Breast Screening Programme was 6.4%. The authors noted that the inclusion of deaths from cancers diagnosed before the screening started would dilute the observed benefit of screening. Duffy et al. (2002b) subsequently showed that more than half of the breast cancer deaths in a given 10-year period are from cancers diagnosed before screening started, and consequently that the effect on mortality from cancers diagnosed in the screening epoch is likely to be twice as high as the 6.4% mortality reduction estimated. For this and other reasons, the full effect of the screening programme was unlikely to be seen until between 2005 and 2010.

As with any temporal comparison, the issue of confounding with treatment arises. Although the age groups above the screening range might not have benefited fully from the therapeutic changes, it is reasonable to suppose that the age groups below the screening range would have done so. The greater mortality reduction in 1998 in the age group 50–54 years compared with the age group 75–79 years (17.0% vs 12.8%) appears to bear this out.

(c) Prospective or retrospective cohort analysis of incidence-based mortality

Incidence-based mortality studies are cohort studies in which the incidence-based mortality from breast cancer diagnosed after the first invitation to screening is compared with an estimate of expected breast cancer mortality in the absence of screening. The breast cancer mortality expected in a situation without screening can be estimated using breast cancer mortality rates in a cohort not (yet) invited to screening, or
using historical data on breast cancer mortality patterns from the same region. Ideally, historical and current data on breast cancer mortality from a region in which screening is absent are included, to account for possible temporal changes that affect breast cancer mortality (e.g. improvements in breast cancer treatment). Incidence-based mortality studies have several methodological advantages, including avoidance of lead-time bias and achieving appropriate correspondence in time of the breast cancer incidence and mortality between the study and control cohorts.

Suppose a screening programme started in 1990, in a stable target population of 100 000 women aged 50–69 years. One might have available data to compare breast cancer mortality in the 1 000 000 person–years of eligible follow-up in 1990–1999 with the same mortality in the corresponding 1 000 000 person–years of observation in 1980–1989, before the screening was initiated. However, such a comparison of deaths from breast cancer regardless of time of diagnosis would include in 1990–1999 deaths from breast cancers diagnosed before 1990 and so with no potential for exposure to screening. The IBM approach would include only deaths from cancers diagnosed at ages 50–69 years during either 1990–1999 or 1980–1989. Although this approach may incur some conservative bias due to lead time, this would be outweighed by the correct classification of exposure to invitation to screening (Swedish Organised Service Screening Evaluation Group, 2006a). Since the risk of breast cancer death may change with time since diagnosis, it is desirable that the observation periods with and without screening be of equal duration.

A real instance of this approach is now considered. The study of Olsen et al. (2005) compared changes in incidence-based breast cancer mortality in the period 1991–2001 in the Copenhagen screening programme with changes in the rest of Denmark (which was without a screening programme and was consequently taken as the national control group). Incidence-based breast cancer mortality rates declined from 69 per 100 000 in the pre-screening period to 52 per 100 000 in the screening period in the Copenhagen area, and almost no change (from 52 to 53 per 100 000) was observed in the national control group. This observation led to an estimated relative risk of breast cancer death of 0.75 (95% confidence interval [CI], 0.63–0.89). Any changes in therapy in the Copenhagen area over the period would also have been seen in the national control group, given the standardization of treatment performed in accordance with the Danish Breast Cancer Cooperative Group (Fischerman & Mouridsen, 1988). Since the only deaths included were those from cancers diagnosed during the relevant periods, there was no dilution of the effect of the screening due to deaths from cancers diagnosed before screening started.

(d) Case–control studies

In a case–control study, exposure to screening (history of breast cancer screening attendance) is compared between women who died of breast cancer (cases) and women who did not die of breast cancer (controls). Potentially important biases associated with case–control studies include selection bias and information bias related to the time at which exposure is defined. Because screening attendance is used as the exposure measure, selection bias plays an important role, as women attending screening might be more health-conscious than women not attending screening. Selection bias influences the estimated effect of the study in favour of screening but may be corrected, at least partially, using statistical methods (adaptation by Duffy et al., 2002a of the correction of Cuzick et al., 1997 for RCTs). For a correct estimate of selection bias, it is crucial to have data available on the variables that influence breast cancer mortality, or on breast cancer mortality between attenders and non-attenders (Paap et al., 2014).
Generally speaking, the definition of exposure to screening can lead to bias both in favour of screening and against screening. If exposure is defined as “ever screened” versus “never screened”, bias will occur in favour of screening. Because all cases have died of breast cancer and were therefore very likely to have been diagnosed with breast cancer some time before death, most will have stopped being invited to screening some time before death. In contrast, controls (most of whom were not diagnosed with breast cancer) would have continued to be invited to screening up to near the time of their death, and would thus have been more likely to be exposed to screening. This difference in the probability of having been screened would lead to bias in favour of screening. This bias in favour of screening is eliminated if exposure is defined as screening attendance to the time of the case’s breast cancer diagnosis, so that exposure stops simultaneously for cases and controls. Although in this design the bias in favour of screening is eliminated, bias against screening is likely to occur because a case is eligible to be screened until cancer is detected either clinically or by screening, whereas controls matched to a case with a cancer detected by screening are eligible to be screened only until the cancer of their matched case is detected by screening. This bias can be corrected by defining exposure for controls matched to cases with a screen-detected cancer to the time at which cases with a screen-detected cancer would have been clinically diagnosed (in the absence of screening), but this requires an estimate of the screening lead time for each case (Connor et al., 2000). Exposure of controls matched to cases with a clinical diagnosis remains unchanged.

Essential elements in performing case–control studies are: (i) sampling cases and controls from the same population (i.e. controls that would have had the same probability of becoming cases); (ii) qualitatively equal information on the primary outcome measure; and (iii) correct definition of (population-based) mammography screening exposure. In countries with complete population registries and full coverage of cancer registries and vital statistics, such case–control studies approximate nested case–control studies. Examples of this type of study are the case–control studies done in the Netherlands (e.g. Paap et al., 2014).

Case–control studies consistently report a greater breast cancer mortality reduction associated with screening (up to 50%) compared with the RCTs (Walter, 2003; Broeders et al., 2012). Only a small part of this difference in breast cancer mortality reduction can be explained by differences in study design. RCTs compare breast cancer mortality in women offered screening with that in women not offered screening. The estimated effect is influenced by the participation rate (women who decline the invitation to screening are included in the screened group) and by contamination of the control group. In contrast, most case–control studies estimate breast cancer mortality reduction in women who are screened compared with women who are not screened, thereby excluding women who decline the invitation to screening from the case group and avoiding contamination of the control group. Therefore, the effect estimate assessed in case–control studies can be expected to be stronger, even if adjusted for selection effects.

The independent United Kingdom panel on breast cancer screening reviewed the usefulness of case–control studies in estimating breast cancer mortality reduction associated with screening and considered that bias could inflate the estimate of benefit and that the RCTs provide more reliable evidence for mortality reduction (Marmot et al., 2013). However, the number of screens performed in current screening programmes outnumbers the women screened in the RCTs by hundreds of millions. Therefore, studies conducted in high-quality organized invitation systems, which have almost complete follow-up data and high acceptance rates, can best estimate whether currently implemented
programmes are of benefit to women invited (effectiveness).

The case–control approach is a relatively quick and inexpensive one, based on the principle that if the screening is reducing mortality, women who have died of breast cancer will be characterized by lesser screening histories than those who have not. It does have specific complexities and risks of bias (Walter, 2003; Duffy, 2007; Verbeek & Broeders, 2010). However, these can to some extent be addressed by design and analytical tactics. Within opportunistic, rather than organized, screening, the case–control approach is one of the few evaluation options available. In some health-care environments, it may not be possible to link screening and mortality records, in which case the advanced disease status might be used to define cases (with the possibility to be interviewed with respect to screening status in the absence of screening records).

A notable feature of the case–control evaluation is that its primary comparison is made between participants and non-participants in the screening programme, and this option thus introduces the possibility of self-selection bias. Duffy et al. (2002a) developed a correction for this bias that requires a reliable estimate of the relative risk of breast cancer death in non-attenders versus those not invited to screening. This may be difficult to estimate; however, the method also provides an estimate of how large this relative risk would have to be for the observed benefit to be entirely due to self-selection bias.

An example of a case–control evaluation is the study of the effect of participation in the BreastScreen Australia programme, which has been inviting women aged 50–69 years to 2-yearly mammography since the mid-1990s (Nickson et al., 2012). The 427 breast cancer deaths occurring at some time during 1995–2006 were compared with 3650 controls who were alive. A variable number of controls, selected by incidence density sampling, were matched by month and year of birth to cases (Greenland & Thomas, 1982). In each case–control matched set, a date of first diagnosis of breast cancer (in the majority, the date of diagnosis of the case) was defined as the reference date. The primary definition of exposure to screening was having had a mammogram between the woman’s 50th birthday and the case–control set reference date. Exposure to screening was less common in cases than in controls (39% vs 56%). The odds ratio associated with screening, adjusted for remoteness of residence and socioeconomic status, was 0.48 (95% CI, 0.38–0.59). A series of sensitivity analyses yielded a range of 0.44 to 0.52.

This result may be affected by self-selection bias, despite the adjustment for socioeconomic status and the various sensitivity analyses performed. However, to be entirely due to self-selection bias, the a priori risk of breast cancer death in non-participants compared with uninvited women would have to be at least 1.80, which seems unlikely given the evidence that participants are at a higher risk of breast cancer than non-participants (Thompson et al., 1994; van Schoor et al., 2010; Beckmann et al., 2013). Clearly, the self-selection bias can act in either direction. However, the results do indicate that case–control evaluations appear to be less conservative compared with prospective evaluation approaches.

(e) Ecological studies

An ecological study makes use of aggregated data for exposure or outcome identification, or both, rather than individual-level assessment of the association of the exposure with the outcome.

Ecological studies are generally accorded a lower status than randomized trials or studies using individual data, such as case–control and cohort studies. However, there may be cases where a well-conducted ecological study is more pertinent than a poorly conducted cohort or case–control study. In fact, for population interventions such as mammography breast cancer screening, the distinctions between these study
Breast cancer screening

types may be blurred, making it more important to consider the studies on a case-by-case basis, or at least according to a finer subdivision of types.

Two factors limit the ability to interpret findings in ecological studies. First, the ecological fallacy relates to the uncertain relationship between the mean and the median of characteristics of individuals in cells of aggregated data. Thus, the average use of screening in region A may be higher than that in region B, but if this average is due to very intensive use by a small number of women, one would not expect to see an overall mortality advantage for the women in region A. Second, differences in outcomes may be explained by other risk factors that differ between two regions. These may not be adjusted for, because they are unknown, are unmeasured, or are measured only on average (which returns one to the ecological fallacy). Adequate treatment of these two issues is a necessary condition for considering an ecological study as informative with respect to the effectiveness of mammography screening.

Ecological studies for breast cancer mortality compare data in countries or areas before and after the introduction of screening (interrupted time series), or concurrently between areas with and without screening (geographical comparisons). In the first type of study, extrapolation of time trends means that decisions must be made, for example about the linearity or otherwise of the trend, the choice of time periods considered as “before” and “after” screening, and the age groups included. In the second type of study, choices must be made about the areas to include, the time period considered, and the age groups included. Such decisions, which can appear to have been made rather arbitrarily, can have a profound impact on the estimates obtained. Lack of comparability and different time trends in the groups being contrasted could lead to substantial bias.

Ecological studies that use temporal trends fit regression models to national or regional published mortality data, commonly to estimate annual rates of change in mortality over time and to assess whether and to what extent breast cancer screening affects them. The change points are either dictated by the date of introduction of screening programmes or estimated from the data using joinpoint regression models (Mukhtar et al., 2013). Studies comparing the levels of mortality rates between screening and non-screening periods are not included in this definition (please refer to Sections 5.1.2b and c).

Mukhtar et al. (2013) analysed unrefined breast cancer mortality data (i.e. regardless of epoch of diagnosis) from 1971 or 1979 to 2009 in England, using log-linear models with joinpoint regression. They estimated similar contemporaneous downward trends in mortality during the screening epoch for women younger than 50 years and for those older than 50 years, the lower age limit for screening in Oxford. The joinpoint regression estimated no changes in trends for women aged 64 years or younger but significant changes in the late 1980s in older women. In England as a whole, the authors estimated the largest decreasing relative trend in women younger than 40 years. Years of peak mortality were observed in the mid- to late 1980s, before an effect of screening would be expected.

The authors concluded that screening was unlikely to have affected breast cancer mortality. Problems with this interpretation include the following. (i) The greatest mortality reduction in the most recent period was observed for the youngest age group. Rates were rising in the screening age group until the mid-1980s and falling thereafter. (ii) Because of the methodology’s choice of discontinuities at different ages, the calendar periods comparing the screening and non-screening age groups are not the same. (iii) Screening was mostly confined to ages 50–64 years, and the effect on mortality would be quite substantial in the late sixties and early seventies rather than in the early fifties. (iv) The emphasis on individual years of peak mortality
and year-to-year trends loses sight of the more stable mortality estimates as a whole. The level of mortality was considerably lower in the screening epoch than in the pre-screening epoch, and this difference was most pronounced in the screening age group. (v) The maximum number of change points allowed should be specified. This will also affect their estimated occurrence.

Usually, it is most difficult to anticipate the occurrence of a change point, or its magnitude, based on year-to-year trends in unrefined mortality. This may influence the subjective decision about the number of joinpoints and about whether trends of decreasing mortality would have continued unabated in the absence of screening. Nevertheless, despite the significant complexities of analysis and interpretations, trend studies can be informative, such as the Otto et al. (2003) study.

(f) Modelling studies

Formally, RCTs answer one specific outcome question, namely whether mammography screening reduces breast cancer mortality, given the exact design features, like fixed interval, starting age, and stopping age, and given the background situation of the control group to compare with. Modelling studies are generally intended to predict outcomes beyond the (limited) end of the trial follow-up, and for different schedules of screening. They seek to avoid possible overestimation of the effect of screening on breast cancer mortality, due to lead-time and length bias, by modelling the breast cancer process more directly. The essence of modelling is simulating the natural history of disease, based on the best available data. This is realized by incorporating variables associated with the disease process and with detection and treatment of breast cancer, including the mean duration of the preclinical detectable phase, the probability of transition to the next tumour stage, age- and stage-specific sensitivity of mammography, and stage-specific response to treatment (Berry et al., 2005; Groenewoud et al., 2007). As an example, the number and the time frame of interval cancers being diagnosed give estimates of sensitivity, whereas the detection rates (by stage, age, calendar year, etc.) and interval cancers together give information on the sojourn times of disease (duration of period when cases are screen-detectable). Modelling produces estimates of these unobservable phenomena, and thus there is sometimes scepticism about the evidence coming from modelling studies. Modelling tries to incorporate all available screen and non-screen data and to give the best estimate of the natural history of disease and of what would have happened if no screening had been implemented. In the evaluation of screening, when it is already being introduced, such model predictions are valuable to evaluate and steer the programme, and they are also advisable before implementation for estimating the optimal programme of screening with its benefits and harms as well as its cost-effectiveness. With good estimates, especially of the screen-detectable period, overdiagnosis can be estimated (van Ravesteyn et al., 2015).

However, all good modelling analyses that predict the consequences of treating earlier in the natural history of disease are dependent on efficacy measures, from RCTs or high-quality observational studies, to estimate such results. Therefore, high-quality models are calibrated to such high-quality data (de Koning et al., 1995). The advantage is that differences in protocol, for example attendance and referral rates, and in follow-up period can specifically be taken into account.

In such modelling, the natural history of breast cancer in the absence of screening is first modelled. Some women in the simulated population may develop breast cancer, which develops from a small preclinical lesion to a symptomatic cancer, possibly leading to breast cancer death. In each stage, a lesion may grow to the next stage, regress, or be clinically diagnosed because of symptoms. The natural course of the disease
may be interrupted by screening, at which a preclinical lesion can become screen-detected. Screen detection can result in the detection of smaller tumours, which may entail a survival benefit. Each screen-detected or clinically diagnosed tumour may be treated with adjuvant systemic therapy, which may also improve survival. Critical components of such models are the assumed natural history component, the effects of interrupting by screening or treatment, and extrapolating lifetime harms and benefits (Heijnsdijk et al., 2012). In principle, such elements are calibrated and validated against data from trials and observational studies, and criteria to evaluate models have been proposed (Habbema et al., 2014).

5.1.3 Surrogate indicators of effect on mortality

As noted above, although in principle the main indicator of the effectiveness of a screening programme is its impact on breast cancer mortality, to estimate this impact in practice can be complicated. The population incidence of advanced-stage disease (Smith et al., 2004; Autier et al., 2011) or predicted mortality from the stage of disease diagnosed have been suggested as surrogates for mortality. Randomized trials show that screening that results in a reduction in the incidence of node-positive breast cancer is also accompanied by a reduction in mortality (Smith et al., 2004). A review confirmed this strong inverse association of exposure to screening and of screen detection with nodal status and tumour size (Nagtegaal & Duffy, 2013). To consider potential confounding, the incidence of disease should be compared before and after the introduction of screening, to account for changes in treatment as well as more complete pathological staging and reporting (e.g. the implementation of sentinel node biopsy) in the screening epoch. This gives rise to further complexities of analysis and interpretation of data (Swedish Organised Service Screening Evaluation Group, 2007).

Another possible confounder is the increase in breast cancer incidence recorded in almost all parts of the world in the second half of the 20th century, which is related to mortality and incidence of advanced disease as well as to the introduction of screening. Thus, there are methodological problems when trying to estimate the expected incidence of disease by stage in the absence of screening.

Despite these problems, the rates of advanced-stage disease are still a very direct measure of the impact of early detection by screening, as several studies have reported. To estimate the potential beneficial effect, not simply the proportion of cases with advanced-stage disease but also the reduction in the absolute rate of advanced-stage disease should be reported.

Thus, the incidence of advanced-stage disease might be used as a surrogate for the effect of screening on mortality, but the above-mentioned limitations should be considered. Other indicators include the detection rate of interval cancers and of small tumours, which are necessary but not sufficient indicators of the success of screening (Day et al., 1989, 1995; Tabár et al., 1992). Although they are less direct, these indicators are often more generally observable than the absolute population incidence of advanced-stage disease.

5.2 Preventive effects of mammography

5.2.1 Incidence-based cohort mortality studies

IBM studies are the most methodologically robust studies for evaluating the effectiveness of service mammography in reducing breast cancer mortality (see Section 5.1.1). They are cohort studies usually conducted in association with a population-based mammography
screening programme. Their defining feature is the observation of deaths from breast cancer in women diagnosed after their first invitation to (or attendance to) mammography screening, that is, at a time when their risk of breast cancer death could have been affected by screening. The expected number of breast cancer deaths is estimated in women diagnosed with breast cancer but not invited to screening compared with a matching cohort of women over a similar period of time.

The screening and non-screening cohorts can be fixed or dynamic, most commonly dynamic. For those invited to screening, the date of first invitation is taken from screening records or is estimated from the cohort member’s residence location and the history of the roll-out of screening in the study area and period. For those not invited to screening, the date of first invitation may be allocated to correspond in age and time to those invited, or at about the midpoint of the first screening round for those invited. The two cohorts’ age distributions are usually matched, as are the periods over which their breast cancer experience is recorded. In most cases, incident breast cancers during the accrual period for the study (which begins at the date of first invitation to screening for each woman) and the associated breast cancer deaths are identified in a population-based cancer registry, and deaths from other causes in a regional or national death register. In some studies, one or both cohorts have also been identified in national registers and individual women tracked into and out of the cohorts for accurate estimation of person–years of experience; otherwise, the person–years are estimated using aggregated population data.

This description of the results of IBM studies is based on studies correctly characterized as IBM studies, mostly covered by two recent systematic reviews. The first of these, the Euroscreen review, systematically searched for relevant studies published up to February 2011 in women aged 50–69 years covered by European population-based screening mammography programmes (Broeders et al., 2012; Njor et al., 2012). The second had a similar search strategy to the Euroscreen review but without age restriction or limitation to European populations, and included studies published up to January 2013 (Irvin & Kaplan, 2014). Additional IBM studies were found in an unrestricted systematic search that covered literature published between March 2011 and 22 July 2014. One study published after July 2014 (Coldman et al., 2014) and two early studies not identified in the searches (Morrison et al., 1988 and Thompson et al., 1994) were also known to the Working Group.

Four analyses that were excluded from the Njor et al. (2012) review report were also excluded by the Working Group, on the grounds that they were based exclusively on some or all of the data used for previous reports. However, there remains significant overlap among several studies, which is detailed below.

In almost all instances, the studies reviewed were conducted in areas where population-based service mammography screening had been implemented. There is, in principle, no reason for not conducting such studies within a population exposed only to opportunistic screening, but they are more readily conducted in areas of population-based screening and the Working Group knew of no IBM studies that had been conducted in an area with exclusively opportunistic screening.

The following summary of results of IBM studies is organized into two broad sections: studies that report on breast cancer mortality reduction after mammography screening of women in age groups that include most or all of the age range 50–69 years, and studies that report on mortality reduction from screening in an age group that lies mainly below or above that age range (i.e. women younger than 50 years or older than 69 years).
(a) Women aged 50–69 years invited to screening

The results of studies of mammography screening mainly in women aged 50–69 years are summarized in Table 5.4 and Table 5.5. Table 5.4 covers estimates of relative risk of breast cancer death in women invited to mammography screening relative to women not invited. Table 5.5 does the same for women who were invited and attended screening relative to women who were invited but did not attend. Studies are ordered in the table by the country in which they were conducted (with countries in the order in which their mammography screening programmes were first introduced) and within each country by the earliest date of mammography screening that was included in the analysis.

All analyses reviewed here included women in the age group 50–69 years, with the exception of four analyses in which the women invited or otherwise targeted for screening were aged up to 59 or 64 years and one in which only women from age 55 years were invited. Eight analyses included women invited to screening before age 50 years, and five analyses included women invited to screening beyond age 69 years.

(i) Sweden

The six reports based on population-based mammography screening in Sweden have multiple overlaps in space and time; that is, they drew on geographical mammography experience for more than a year that overlapped with that drawn on by at least one other study. The experiences in the reports of Duffy et al. (2002a, b) are almost completely a subset within that of the Swedish Organised Service Screening Evaluation Group (2006a, b) reports; however, the reports of Duffy et al. (2002a, b) provide valuable additional results and so are included separately in Table 5.4 and below. The whole mammography experience of Jonsson et al. (2007) is also included in that of Swedish Organised Service Screening Evaluation Group (2006a, b), but it does provide some independent information since it uses contemporary and not historical control areas. Most of the screening experience in two of the seven screening areas of Jonsson et al. (2001) overlaps with that in Swedish Organised Service Screening Evaluation Group (2006a, b), and two of the control counties overlap more than 50% of the time with the control counties in Jonsson et al. (2007). The screening experience of the one screening county in Jonsson et al. (2003a) overlaps by 2 years that of Duffy et al. (2002a, b) and by 1 year that of Swedish Organised Service Screening Evaluation Group (2006a, b). The screening experience of one of the two counties included in Tabár et al. (2001) is also included in Duffy et al. (2002a, b) and Swedish Organised Service Screening Evaluation Group (2006a, b).

Sweden’s first population-based mammography screening programme was introduced in 1974 to cover women aged 40–64 years in Gävleborg County. Jonsson et al. (2003a) primarily compared IBM in Gävleborg County with an age-matched control population from four neighbouring counties without mammography screening programmes. Cohorts of women were defined in Gävleborg County according to the date at which invitation to screening began in their district, and corresponding cohorts were created in the control counties. Incident breast cancers and their dates of diagnosis were identified, and their date and cause of death obtained from the Swedish Cancer Registry; aggregated population data were used to estimate person–years at risk. The study also included a reference period (1964–1973), in which any pre-existing difference in breast cancer mortality between Gävleborg County and the control counties could be estimated and adjusted for in the analysis. Incident breast cancers were accrued for 10 years, and the follow-up period for breast cancer mortality was 22 years; cases were accrued only in the age group 40–64 years, and follow-up extended to age 79 years. [These differences in accrual and follow-up periods and age
<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sweden</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jonsson et al. (2003a)</td>
<td>Gävleborg County and 4 other counties 1974 40–64 yr average, 38 mo (earlier) and 23 mo (later)</td>
<td>Invited 885 000 Not invited 2 581 000</td>
<td>10 yr</td>
<td>1974–1986 (max 10 yr) Same + 15 yr</td>
<td>40–64 yr Same + 15 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Same period; different counties</td>
<td>Yes</td>
<td>Age, follow-up time, county, period (study or reference)</td>
<td>0.86 (0.71–1.05) RR, 0.82 adjusted for lead-time bias; adjustment for inclusion bias(^d) did not change RR</td>
<td></td>
</tr>
<tr>
<td>Fig. 5.4) (^\text{a})</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tabár et al. (2001)</td>
<td>2 counties 1978 40–69 yr 1.5–2 yr</td>
<td>Invited 1 100 931 Not invited 1 213 136</td>
<td>≤ 9 yr</td>
<td>1988–1996 Same</td>
<td>40–69 yr Not stated</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different periods; same areas</td>
<td>No (screening period was 1 yr shorter than non-screening period)</td>
<td>Selection bias</td>
<td>0.52 (0.43–0.63) 0.64 (0.30–1.36)</td>
<td>It is uncertain whether there is lead-time bias</td>
</tr>
</tbody>
</table>
### Table 5.4 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duffy et al. (2002b)</td>
<td>7 counties, 1978–1994 40 or 50 yr to 69 or 74 yr 1.5–2.75 yr</td>
<td>Invited 3 815 330</td>
<td>5–20 yr</td>
<td>1978–1997 to 1994–1998</td>
<td>40–69 yr (6 counties), 50–59 yr (1 county) Same</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different periods; same areas</td>
<td>Yes</td>
<td>Lead-time bias, time trend in breast cancer mortality 0.74 (0.68–0.81)&lt;sup&gt;f&lt;/sup&gt; ≤ 10 yr of screening: 0.82 (0.72–0.94) &gt; 10 yr of screening: 0.68 (0.60–0.77) Analyses in 5 counties based on ≤ 10 yr of screening, in 2 counties based on &gt; 10 yr. Substantial overlap with Swedish Organised Service Screening Evaluation Group (2006a)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish Organised Service Screening Evaluation Group (2006a)</td>
<td>13 areas, 1980–1990, depending on area 40 or 50 yr to 69 yr, depending on area probably mostly 2 yr</td>
<td>Invited 7 542 833</td>
<td>11–22 yr, depending on area</td>
<td>1980–2001 to 1990–2001</td>
<td>40–69 yr (8 areas) or 50–69 yr (5 areas) Maximum follow-up age not stated</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different periods; same areas</td>
<td>Yes</td>
<td>Time trend in breast cancer mortality 0.73 (0.69–0.77) Updated and expanded analysis incorporating almost all data used for Duffy et al. (2002b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jonsson et al. (2001)</td>
<td>12 counties, 1986 50–69 yr 2 yr</td>
<td>Invited 2 036 000</td>
<td>7 yr</td>
<td>1986–1994</td>
<td>50–69 yr Same + 10 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Same period; different counties</td>
<td>Yes</td>
<td>Age, year of follow-up, area, period 0.90 (0.74–1.10) RR, 0.87 adjusted for inclusion bias.&lt;sup&gt;d&lt;/sup&gt; Lead-time bias estimated to be −0.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Areas, earliest year of programme screening, screening age, screening interval</td>
<td>Person–years</td>
<td>Duration of screening</td>
<td>Accrual and follow-up periods</td>
<td>Diagnosis and death age ranges</td>
<td>Individual or aggregate data</td>
<td>Temporal and geographical similarity of comparison group</td>
<td>Time-balanced follow-up periods?</td>
<td>Adjustments</td>
<td>Breast cancer mortality RR (95% CI)</td>
<td>Comments</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-------------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>--------------------------------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Jonsson et al. (2007)</td>
<td>4 counties 1989 40–74 yr average, 20–22 mo</td>
<td>Invited 1 223 346 Not invited 915 948</td>
<td>7 yr</td>
<td>1989–1996 Same + 10 yr</td>
<td>50–69 yr Individual for breast cancer cases; aggregate, all other women</td>
<td>Different period (accrual 1989–1996 for study group, 1988–1994 for control group); different areas</td>
<td>No (study group follow-up to 2001, control group to 1998)</td>
<td>Age</td>
<td>50–69 yr: 0.86 (0.86–1.17) 40–74 yr: 0.74 (0.58–0.94)</td>
<td>Lead-time bias estimated to be ~2% at ages 50–69 yr and 40–74 yr ~85% of invited women screened</td>
<td></td>
</tr>
</tbody>
</table>

**The Netherlands**

| Peer et al. (1995) | 2 cities 1975 35–64 yr 2 yr | Invited 166 307 Not invited 154 103 | 15 yr | 1975–1990 Same | 35–64 yr Individual for breast cancer cases; aggregate, all other women | Same period; different cities | Yes | None stated | 0.94 (0.68–1.29) | Study followed for 15 yr a cohort aged 35–49 yr at first invitation. Cities may differ in underlying breast cancer mortality trends |

**United Kingdom**

| UK Trial of Early Detection of Breast Cancer Group (1999)a | England and Scotland, 6 health service areas 1979 45–64 yr 2 yr | Invited 793 288 Not invited 2 346 328 | 7 yr | 1979–1995 Same | 45–80 yr Individual | Same period; different health service areas | Yes | Age, pre-trial breast cancer mortality | 0.73 (0.63–0.84) Screening included annual CBE 65% of invited women screened |
### Table 5.4 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years(^a)</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)(^c)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hakama et al. (1997)</strong></td>
<td>84% of municipalities 1987 50–64 yr 2 yr</td>
<td>Invited 400 804 Not invited 299 228</td>
<td>≤ 6 yr</td>
<td>1987–1992 Same</td>
<td>50–64 yr Same</td>
<td>Individual for all women</td>
<td>Same period; same areas</td>
<td>Yes</td>
<td>Age</td>
<td>0.76 (0.53–1.09)</td>
<td>Approximately 1/6 of women invited to 1 screening round, 1/3 to 2 rounds, and 1/2 to 3 rounds 85% of invited women screened</td>
</tr>
<tr>
<td><strong>Anttila et al. (2002)</strong></td>
<td>Helsinki 1986 50–59 yr 2 yr</td>
<td>Invited 161 400 Uninvited 155 400</td>
<td>0.5–10.5 yr; 1–5 screening rounds</td>
<td>1986–1997 Uncertain</td>
<td>50–59 yr Uncertain</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different period; same area</td>
<td>Yes</td>
<td>Age at death, time trend in breast cancer mortality at age 40–49 yr in screened and unscreened cohorts</td>
<td>0.81 (0.62–1.05)</td>
<td>Possible difference in age of case accrual and follow-up, and therefore lead-time bias</td>
</tr>
<tr>
<td><strong>Parvinen et al. (2006)</strong></td>
<td>Turku 1987 55–74 yr 2 yr</td>
<td>Invited 204 896 Not invited 199 329</td>
<td>11 yr</td>
<td>1987–1997 Same + 4 yr</td>
<td>55–74 yr Same + 10 yr</td>
<td>Individual for invited women; aggregate for not invited women</td>
<td>Different periods; same area</td>
<td>Yes</td>
<td>Age, time trend in mortality extrapolated from 1970 to 1986</td>
<td>55–74 yr: 0.58 (0.41–0.83) 65–69 yr: 0.42 (0.21–0.84)</td>
<td>Some lead-time bias</td>
</tr>
</tbody>
</table>
### Table 5.4 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years$^b$</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)$^c$</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anttila et al. (2008)</td>
<td>410 municipalities 1987 50–69 yr 2 yr</td>
<td>Invited 1 822 900 Not invited no estimate provided</td>
<td>≤ 5 yr</td>
<td>1992–1996 Same + 3 yr</td>
<td>50–69 yr Same + 10 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different period; same area</td>
<td>No</td>
<td>Age at diagnosis, cohort, year</td>
<td>0.89 (0.81–98)</td>
<td>Some lead-time bias. Breast cancer mortality in the absence of screening was extrapolated from statistical models of breast cancer mortality from 1971 to 1986 and in age groups 40–49 yr and 65–69 yr up to 1991</td>
</tr>
</tbody>
</table>
### Table 5.4 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person–years*</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)*</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sarkeala et al. (2008a, b)</strong></td>
<td>260 municipalities 1987 50–69 yr (up to 74 yr in some municipalities) 2 yr</td>
<td>Invited 2 330 266 Not invited 401 002</td>
<td>≤ 12 yr 1992–2003 Same</td>
<td>50–69 yr Same</td>
<td>Individual for invited women; aggregate for not invited women</td>
<td>Different period; same area</td>
<td>Yes</td>
<td>Age at death, centre recall categories, period, calendar year within period, interaction between calendar year and age 0.78 (0.70–0.87) 50–59 yr: 1.04 (0.81–1.31) 50–59 yr (up to 69 yr): 0.84 (0.75–0.92) 50–69 yr (up to 74 yr): 0.72 (0.51–0.97)</td>
<td>Time trend in breast cancer mortality taken account of by modelled adjustment for calendar period. 87% of invited women screened. The material was grouped by screening policy of the municipality</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Italy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Paci et al. (2002)</strong></td>
<td>Florence 1990 50–69 yr 2 yr</td>
<td>Invited 254 890 Not invited not stated</td>
<td>≤ 7 yr 1990–1996 Same + 3 yr</td>
<td>50–76 yr Same + 3 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different period; same area</td>
<td>Yes</td>
<td>None stated 0.81 (0.64–1.01)</td>
<td>Some lead-time bias</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Person–years: Number of person–years at risk or observed | Duration of screening: Time period during which observations were made | Accrual and follow-up periods: Periods during which accrual and follow-up occurred | Diagnosis and death age ranges: Age ranges for diagnosis and death | Individual or aggregate data: Data collected for individual or aggregate | Temporal and geographical similarity of comparison group: Similarity of temporal and geographical features | Time-balanced follow-up periods: Periods during which follow-up was balanced | Adjustments: Adjustments made to account for confounding factors | Breast cancer mortality RR: Relative risk of breast cancer mortality | Comments: Additional comments on study design or results.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascunce et al. (2007)</td>
<td>Navarre 1990 50–69 yr 2 yr</td>
<td>Invited 293 000 Not invited 289 000&lt;sup&gt;h&lt;/sup&gt;</td>
<td>5 yr</td>
<td>1997–2001 Same</td>
<td>50–69 yr Uncertain</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different period; same area</td>
<td>Yes</td>
<td>Age</td>
<td>0.58 (0.44–0.75)</td>
<td>Lead-time bias is possible. RR not adjusted for trend in breast cancer mortality; RR for age 30–44 yr was 1.07, for age ≥ 75 yr was 1.03</td>
</tr>
<tr>
<td><strong>Denmark</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olsen et al. (2005)</td>
<td>Copenhagen 1991 50–69 yr 2 yr</td>
<td>Invited 430 823 Not invited 634 224</td>
<td>≤ 10 yr</td>
<td>1991–2001 Same</td>
<td>50–69 yr, mainly 50–79 yr</td>
<td>Individual for all women</td>
<td>Different period; same city</td>
<td>Yes</td>
<td>Age, exposure, period, region, period*region</td>
<td>0.75 (0.63–0.89)</td>
<td>Some lead-time bias. Adjusted for underlying mortality trend and difference between regions by including period*region term in model</td>
</tr>
</tbody>
</table>
### Table 5.4 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Norway</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalager et al. (2010)</td>
<td>1996–2005, depending on area 50–69 yr 2 yr</td>
<td>Invited 2 337 323</td>
<td>10 yr in 1 region; 2–6 yr in 5 regions</td>
<td>1996–2005 Same</td>
<td>50–69 yr Same + 9 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different period; same area</td>
<td>No</td>
<td>Age [0.88&lt;sup&gt;g&lt;/sup&gt; (0.73–1.05)]</td>
<td>Some lead-time bias. Widespread opportunistic screening before programme began</td>
<td></td>
</tr>
<tr>
<td>Olsen et al. (2013)</td>
<td>1996–2005, depending on area 50–69 yr 2 yr</td>
<td>Invited 1 182 747</td>
<td>≤ 6 yr</td>
<td>1996–2001 or 2002</td>
<td>50–69 yr 50–69 or 50–81&lt;sup&gt;i&lt;/sup&gt;</td>
<td>Individual for all women</td>
<td>Different period; same counties</td>
<td>Yes</td>
<td>Age at death, breast cancer mortality trend in reference region&lt;sup&gt;j&lt;/sup&gt;</td>
<td>0.89 (0.71–1.12) from the &quot;evaluation&quot; model</td>
<td>Some lead-time bias. Study group screened 1–3 times in the population-based programme. Widespread opportunistic screening before programme began</td>
</tr>
</tbody>
</table>
### Table 5.4 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person–years&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weedon-Fekjær et al. (2014)</td>
<td>1996–2005, depending on area 50–69 yr 2 yr</td>
<td>Invited 2 407 709 Not invited 12 785 325</td>
<td>1–15 yr median, 4.5 yr</td>
<td>1986–2009 Same</td>
<td>50–79 yr Same</td>
<td>Individual for all women</td>
<td>Partly different period (1986–2009 for all women, 1995–2009 for invited women); whole country</td>
<td>No</td>
<td>Age, period, cohort, county, lead-time bias</td>
<td>0.72 (0.64–0.79)</td>
<td>Bulk of “not invited” follow-up was in 1986–1995. Widespread opportunistic screening before programme began</td>
</tr>
</tbody>
</table>

<sup>a</sup> One study evaluated invitation to mammography plus CBE.

<sup>b</sup> Person–years: number of women or number of breast cancer deaths.

<sup>c</sup> All RRs are for breast cancer as the underlying cause of death.

<sup>d</sup> Bias from inclusion of deaths from breast cancers that were diagnosed in the period between becoming eligible for screening (either by start of screening or by reaching a certain age) and being invited to be screened.

<sup>e</sup> Estimated trend-adjusted, obtained by the Working Group by dividing the authors’ estimate by the incidence-based mortality RR comparing women aged 40–69 years not invited to screening in 1988–1996 with women aged 40–69 years in 1968–1977.

<sup>f</sup> Estimated by combining RRs for ≤ 10 yr screening and > 10 yr screening using a fixed effects meta-analytic method.

<sup>g</sup> RR and 95% CI adjusted for trend or geographical difference in underlying mortality were calculated as ratio of the authors’ estimated RRs comparing screening area with control period or area; 95% CI of ratio estimated using method in Altman & Bland (2003) as implemented in [http://www.hutchon.net/CompareRR.htm](http://www.hutchon.net/CompareRR.htm).

<sup>h</sup> Estimated from number of breast cancer deaths and breast cancer mortality rate in Table 4 of the article.

<sup>i</sup> Alternative dates applied to two different birth cohorts.


CBE, clinical breast examination; CI, confidence interval; mo, month or months; RR, relative risk; yr, year or years.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person–years</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sweden</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tabár et al. (2001)</td>
<td>2 counties 1978 40–69 yr 1.5–2 yr</td>
<td>Screened 932 229 Not screened 168 702</td>
<td>≤ 9 yr</td>
<td>1988–1996 Same</td>
<td>40–69 yr Not stated</td>
<td>Individual for breast cancer cases; aggregate, all other women (including participation in screening)</td>
<td>No (screening period was 1 yr shorter than non-screening period)</td>
<td>None</td>
<td>Time trend [0.46 (0.21–0.97)]</td>
<td>0.37 (0.30–0.46)</td>
<td>Uncertain whether there is lead-time bias. Not adjusted for self-selection bias</td>
</tr>
<tr>
<td>Duffy et al. (2002b)</td>
<td>7 counties 1978–1994 40 or 50 yr to 69 or 74 yr 1.5–2.75 yr</td>
<td>Screened 2 687 855 Not screened 628 681</td>
<td>5–20 yr</td>
<td>1978–1997 to 1994–1998 Same</td>
<td>40–69 yr (6 counties), 50–59 yr (1 county) Same</td>
<td>Individual for breast cancer cases; aggregate, all other women (including participation in screening)</td>
<td>Different periods; same areas</td>
<td>Yes</td>
<td>Lead-time bias, self-selection bias</td>
<td>0.61 (0.55–0.68)</td>
<td>See also Swedish Organised Service Screening Evaluation Group (2006a). Adjusted for self-selection bias using method of Duffy et al. (2002b)</td>
</tr>
<tr>
<td>Reference</td>
<td>Areas, earliest year of programme screening, screening age, screening interval</td>
<td>Person-years</td>
<td>Duration of screening</td>
<td>Accrual and follow-up periods</td>
<td>Diagnosis and death age ranges</td>
<td>Individual or aggregate data</td>
<td>Temporal and geographical similarity of comparison group</td>
<td>Time-balanced follow-up periods?</td>
<td>Adjustments</td>
<td>Breast cancer mortality RR (95% CI)</td>
<td>Comments</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------</td>
<td>-------------</td>
<td>-------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Swedish Organised Service Screening Evaluation Group (2006a)</strong></td>
<td>13 areas 1980–1990, depending on area 40 or 50 yr to 69 yr, depending on area probably mostly 2 yr</td>
<td>Screened 5,612,312  Not screened 1,930,521</td>
<td>11–22 yr, depending on area</td>
<td>1980–2001 or 1990–2001</td>
<td>40–69 yr (8 areas) or 50–69 yr (5 areas) Maximum follow-up age not stated</td>
<td>Individual for breast cancer cases; aggregate, all other women (including participation in screening)</td>
<td>Different periods; same areas</td>
<td>Yes</td>
<td>Lead-time bias, self-selection bias using method of Duffy et al. (2002b)</td>
<td>0.57 (0.53–0.62)</td>
<td>Updated and expanded analysis based on analysis in Duffy et al. (2002b)</td>
</tr>
<tr>
<td><strong>Jonsson et al. (2007)</strong></td>
<td>4 counties 1989–1997, average, 20–22 mo</td>
<td>Invited 1,223,346  Not invited 915,948 (Only 9% of breast cancer cases were in women who did not attend screening)</td>
<td>7 yr</td>
<td>1989–1997  Same + 4 yr</td>
<td>50–69 yr 50–79 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different period (accrual 1989–1997 for study group, 1988–1994 for control group); different counties</td>
<td>No (study group follow-up to 2001, control group to 1998)</td>
<td>Age, difference in breast cancer mortality between study group and control group in preceding 7 yr, self-selection for screening</td>
<td>0.70 (0.57–0.86)</td>
<td>Lead-time adjustment was estimated to be −2% −85% of invited women screened Adjusted for self-selection bias using method of Cuzick et al. (1997)</td>
</tr>
</tbody>
</table>
### Table 5.5 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Finland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sarkeala et al. (2008b)</td>
<td>260 municipalities 1987 50–59 yr (invited); 60–69 yr (optional) 2 yr</td>
<td>Screened 1,023,598 Not screened 1,365,177 (&quot;screened&quot; = screened after first invitation; &quot;not screened&quot; includes not invited and invited but not screened)</td>
<td>≤ 12 yr</td>
<td>1992–2003 Same</td>
<td>50–79 yr 60–79 yr</td>
<td>Individual for screened women; aggregate for unscreened women</td>
<td>Different period; same area</td>
<td>Yes</td>
<td>Age at death, screening policy category, calendar period. Adjusted for self-selection bias using method of Cuzick et al. (1997)</td>
<td>0.63 (0.53–0.75)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>Time trend in breast cancer mortality taken account of by modelled adjustment for calendar period</td>
</tr>
<tr>
<td><strong>Italy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puliti &amp; Zappa (2012)</td>
<td>Florence 1991 50–69 yr 2 yr</td>
<td>Screened 466,205 Not screened 248,182</td>
<td>1–16 yr</td>
<td>1992–2007 Same + 1 yr</td>
<td>50–85 yr 50–86 yr</td>
<td>Individual for all women</td>
<td>Same period; same population</td>
<td>Yes</td>
<td>Age at entry, marital status, deprivation index. No additional adjustment for self-selection bias</td>
<td>0.51 (0.40–0.66)</td>
<td>Some lead-time bias</td>
</tr>
</tbody>
</table>
### Table 5.5 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Canada</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goldman et al. (2014)</td>
<td>7 provinces 1990 Most ≥ 40 yr Most 40–49 yr 1 yr ≥ 50 yr 2 yr</td>
<td>Screened and not screened 20 200 000</td>
<td>1–20 yr</td>
<td>1990–2009 Same</td>
<td>40–99 yr Same</td>
<td>Individual for screened women; aggregate for unscreened women</td>
<td>Same period; same population</td>
<td>Yes</td>
<td>Age 0.60 (0.52–0.67)</td>
<td>Self-selection bias estimated for British Columbia women aged 40–49 yr at entry using an ad hoc approach: unadjusted RR, [0.43 (0.28–0.61)]; adjusted RR, [0.39 (0.19–0.91)]</td>
<td></td>
</tr>
</tbody>
</table>
### Table 5.5 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person–years&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Denmark</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Olsen et al. (2005)</strong></td>
<td>Copenhagen 1991 50–69 yr 2 yr</td>
<td>Invited 430 823 Not invited 634 224 (Not separately estimated for screened and not screened women)</td>
<td>≤ 10 yr 1991–2001, mainly 50–79 yr</td>
<td>Individual for all women</td>
<td>Different period; same city</td>
<td>Yes</td>
<td>Age, exposure, period, region, period<em>region. Adjusted for underlying mortality trend by including period</em>region term in model</td>
<td>0.60 (0.49–0.74)</td>
<td>0.63 adjusted for self-selection bias using an ad hoc approach. ~71% participation; widespread opportunistic screening before programme began</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Norway</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Hofvind et al. (2013)</strong></td>
<td>Norway 1996–2005, depending on area 50–69 yr 2 yr</td>
<td>Screened 4 814 060 Not screened 988 641</td>
<td>1–15 yr median, 4.5 yr 1996–2009, mainly 50–85 yr</td>
<td>Individual for all women</td>
<td>Same period; same population</td>
<td>Yes</td>
<td>Age, calendar period, time in screened or unscreened cohort, self-selection bias using method of Cuzick et al. (1997)</td>
<td>0.57 (0.51–0.64)</td>
<td>Some lead-time bias. Estimated RR in those invited, 0.64. Widespread opportunistic screening before programme began</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Areas, earliest year of programme screening, screening age, screening interval</td>
<td>Person-years</td>
<td>Duration of screening</td>
<td>Accrual and follow-up periods</td>
<td>Diagnosis and death age ranges</td>
<td>Individual or aggregate data</td>
<td>Temporal and geographical similarity of comparison group</td>
<td>Time-balanced follow-up periods?</td>
<td>Adjustments</td>
<td>Breast cancer mortality RR (95% CI)</td>
<td>Comments</td>
</tr>
<tr>
<td>-----------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>--------------------------------</td>
<td>-------------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------</td>
<td>------------</td>
<td>----------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>USA</td>
<td>BCDDP (29 centres) 1973–1977 35–74 yr 1 yr</td>
<td>Screened 55 053 White women</td>
<td>5 yr</td>
<td>1–9 yr after first screen [1973–1986]</td>
<td>35–83 yr 35–83 yr</td>
<td>Individual for screened women; aggregate for unscreened women</td>
<td>Same period; comparison derives from SEER</td>
<td>Yes</td>
<td>Age, calendar period, lead-time bias</td>
<td>0.80</td>
<td>Age at entry: 35–49 yr: 0.89 50–59 yr: 0.76 60–74 yr: 0.74</td>
</tr>
</tbody>
</table>

**Table 5.5 (continued)**
Table 5.5  (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person–years&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Comments</th>
</tr>
</thead>
</table>
| Thompson et al. (1994)<sup>a</sup> | Western Washington State 1985  
≥ 40 yr  
1–3 yr | Whole cohort: 94 656 women  
Subcohort: 2242, including 5 breast cancer deaths | ≤ 3.5 yr in programme  
< 5 yr including opportunistic | 1982–1988 Same  
Same | ≥ 40 yr Same  
Same for all women  
Same period; same area | Yes | Age, mother’s history of breast cancer, nulliparity, history of breast biopsy | ≥ 40 yr: 0.80  
(0.34–1.85)  
≥ 50 yr: 0.61  
(0.23–1.62) | Screening included CBE.  
Unadjusted RR, 1.09  
(0.58–2.07) | Two studies evaluated invitation to mammography plus CBE.  
Person–years: number of women or number of breast cancer deaths.  
RRs are for breast cancer as the underlying cause of death when alternative estimates (e.g. excess mortality) are also provided.  
Estimated by combining RRs and 95% CIs, using a fixed effects model, across the three screening policy categories in Table 3 of the article. In an earlier analysis of similar data (Sarkeala et al., 2008a), the authors reported an RR for screening of 0.66 (95% CI, 0.58–0.75) in women aged 50–69 years in follow-up, which, when adjusted for self-selection, became 0.72 (95% CI, 0.56–0.88).  
BCDDP, Breast Cancer Detection Demonstration Project; CI, confidence interval; mo, month or months; RR, relative risk; SEER, Surveillance, Epidemiology, and End Results; yr, year or years. |
groups created the possibility of lead-time bias in the results. Also, bias due to inclusion of some cases of breast cancer that occurred early in the roll-out of screening and before the first invitation to screening (inclusion bias) was possible. The estimated IBM relative risk for death from breast cancer was 0.86 (95% CI, 0.71–1.05) based on breast cancer deaths ascertained as the underlying cause of death from the death certificate and adjusted for age, follow-up time, county, and period (study or reference). Corresponding relative risks were 0.82 (no CI stated) after adjustment for lead-time and inclusion biases, 0.82 (95% CI, 0.65–1.03) when based on an estimate of excess mortality due to breast cancer, which does not require use of the certified underlying cause of death, and 0.93 (95% CI, 0.77–1.11) when based on the “rest of Sweden” as the control group.

The relative risk of 0.86 (95% CI, 0.71–1.05) was chosen from the alternatives listed above to be reported in the table. This choice was made a priori on the grounds that: (i) the relative risk was based on the underlying cause of death (the excess mortality measure is not consistently reported in the studies reviewed); (ii) it was the most fully adjusted relative risk that also included its 95% confidence interval; (iii) the Working Group considered four neighbouring counties to be a more nearly similar control group for the study group than the whole of the rest of Sweden; and (iv) this study overlapped the least with other Swedish studies.

Jonsson et al. (2001) and Jonsson et al. (2007) had fundamentally the same design as Jonsson et al. (2003a), except that Jonsson et al. (2007) made historical rather than geographical comparisons of breast cancer mortality in women invited to screening in a later period with that in women in the same population not invited to screening in an earlier period. Jonsson et al. (2007) is the weakest of the three, because of its overlaps with Jonsson et al. (2001) and Swedish Organised Service Screening Evaluation Group (2006a, b) and because of the difference in the length of the follow-up periods in women invited and not invited to screening. The IBM relative risk in invited women aged 50–69 years was 0.90 (95% CI, 0.74–1.10) (0.87 adjusted for inclusion bias, and with lead-time bias estimated to be −0.4%) in Jonsson et al. (2001) and 0.86 (95% CI, 0.86–1.17) (lead-time bias estimated to be −2%) in Jonsson et al. (2007).

Tabár et al. (2001) estimated post-RCT effectiveness of mammography screening in the Swedish Two-County study by comparing post-RCT experience with a balanced period of pre-RCT experience. [The reporting of this analysis is limited; there is uncertainty as to whether the result may be affected by lead-time bias and whether there is any statistical adjustment of the relative risks.] To obtain the IBM relative risk for breast cancer mortality in women invited to screening, the authors first estimated the IBM relative risk for attendance to screening (by comparing breast cancer mortality in women aged 40–69 years who attended screening in 1988–1996 with that in women aged 40–69 years in 1968–1977, before any screening) and then adjusted this for self-selection bias to obtain an adjusted relative risk for invitation to screening of 0.52 (95% CI, 0.43–0.63). [However, this estimate was not adjusted for the underlying trend in breast cancer mortality between 1968–1977 and 1988–1996.]

In a similar historical control-design IBM study based in seven Swedish counties, Duffy et al. (2002a, b) estimated an IBM relative risk of 0.74 (95% CI, 0.68–0.81) for screening in women aged 40–69 years based on 5–20 years of screening and follow-up until 1997 or 1998, and adjusted for lead-time bias and the underlying time trend in breast cancer mortality. For counties with 10 years or less of screening, the estimated relative risk was 0.82 (95% CI, 0.72–0.94), and for counties with more than 10 years of screening, it was 0.68 (95% CI, 0.60–0.77).

The Swedish Organised Service Screening Evaluation Group (2006a, b) analysis was of
Breast cancer screening

(a) Sweden

A similar design but expanded to 13 areas of Sweden and had 11–22 years of screening experience of women aged 40–69 years or 50–69 years and followed up until 2001. The IBM relative risk for screening at age 40–69 years was 0.73 (95% CI, 0.69–0.77) after adjustment for the underlying trend in breast cancer mortality.

(ii) The Netherlands

Peer et al. (1995) compared breast cancer mortality in women born in 1925–1939 who were resident in Nijmegen and were offered mammography screening every 2 years from 1975 until the end of 1990 with that of age-matched women resident in Arnhem and not offered screening. Cause of death was ascertained from clinical records and was considered to be breast cancer if metastases had been diagnosed and other causes of death could be ruled out. The IBM relative risk for breast cancer mortality in Nijmegen women relative to Arnhem women was 0.94 (95% CI, 0.68–1.29). Breast cancer mortality in women aged 35–64 years had been reported to be lower in Nijmegen than that in Arnhem in 1970–1974. This difference was observed not to persist in the period 1975–1979. No adjustment was made for possible differences or trends in underlying breast cancer mortality rates.

(iii) United Kingdom

The United Kingdom Trial of Early Detection of Breast Cancer (UK Trial of Early Detection of Breast Cancer Group, 1999) was a non-randomized trial that began in 1979 and preceded population-based mammography screening in the United Kingdom by 10 years. IBM to 16 years of follow-up was compared between two health service areas in which women aged 45–64 years were invited to be screened by mammography and clinical breast examination (CBE) every 2 years for four rounds, with CBE only in the intervening years, and two areas in which women received the usual care. The relative risk was 0.73 (95% CI, 0.63–0.84).

(iv) Finland

Five studies have reported IBM analyses of mammography screening in Finland. [Overlaps are not accurately identifiable from published reports but seem likely.] The study of Hakama et al. (1997) overlaps minimally with the studies of Anttila et al. (2008) and Sarkeala et al. (2008a, b) because Hakama et al. (1997) covered screening in 1987–1992 and the other three covered screening from 1992 to 2002 or 2003. Anttila et al. (2008) and Sarkeala et al. (2008a, b), which cover 410 and 260 municipalities, respectively, appear to overlap substantially; each of these two studies also overlaps with that of Parvinen et al. (2006), in which the intervention group primarily covered the “entry” cohort in the city of Turku in 1987. The study of Anttila et al. (2002), which included screening in Helsinki in the period 1986–1997, does not overlap with that of Hakama et al. (1997) or with that of Sarkeala et al. (2008a, b) but is assumed to overlap with that of Anttila et al. (2008) in the period 1992–1997. On these bases, it appears that Hakama et al. (1997), Anttila et al. (2002), and Sarkeala et al. (2008a, b) give nearly complete coverage of screening in Finland from 1986 to 2003 with minimal overlap.

Hakama et al. (1997) compared IBM in women aged 50–64 years invited and not invited to mammography screening in 84% of municipalities in 1987–1992, the first 6 years of nationwide screening in Finland. Individual year-of-birth cohorts of women were progressively invited for the first time during this period and experienced up to three screening rounds. The estimated relative risk of breast cancer death was 0.76 (95% CI, 0.53–1.09). The analysis of Anttila et al. (2002) of screening of women in Helsinki over the period 1986–1997 compared IBM in women born in 1935–1939, who had been invited to screening, with that in women born in 1930–1934, who had not. The estimated relative risk of breast cancer death was 0.81 (95% CI, 0.62–1.05) after adjustment for age at death and the estimated...
trend in breast cancer mortality from the trend across the two cohorts at age 40–49 years. [There may be lead-time bias in this result.] Using data from 260 Finnish municipalities and modelling the time trend in breast cancer mortality in the absence of screening, with mortality data from 1974–1985 providing estimated pre-screening mortality, Sarkeala et al. (2008a) estimated an IBM relative risk for invitation to screening in 1992–2003 of 0.78 (95% CI, 0.70–0.87) at age 50–69 years. All municipalities regularly invited only women aged 50–59 years. In those municipalities that had regularly invited women aged 50–69 years (and up to 74 years in some of these) throughout the study period, the corresponding IBM relative risk was 0.72 (95% CI, 0.51–0.97). Incidence and death were measured at age 60–79 years, whereas no impact was observed in municipalities that had stopped screening at age 59 years (Sarkeala et al., 2008b). Studies with variable screening policies provided no clear evidence for a difference in the relative risk for screening between the first 5 years (Hakama et al., 1997) and the next 10 years (Anttila et al., 2002; Sarkeala et al., 2008a, b). In addition, the results of Parvinen et al. (2006) demonstrated a significant effect in women screened regularly at age 55–74 years since 1987 in the “entry” cohort of the screening programme in the municipality of Turku (Table 5.4).

(v) Italy

Paci et al. (2002) estimated the IBM relative risk for women aged 50–69 years invited to screening in the first 7 years of population-based mammography screening in Florence over the period 1990–1999. The expected number of deaths in the absence of invitation to screening was estimated from the expected number of incident breast cancers in women not yet invited to screening in each half-year of the period 1990–1996 and the estimated number of breast cancer deaths to 1999 (from estimated case fatality rates for up to 9.5 years after diagnosis) in women expected to be diagnosed with breast cancer in each of these half-year cohorts. The estimated relative risk was 0.81 (95% CI, 0.64–1.01). [The nature of 13 breast cancer deaths classified as “other” (neither invited nor not invited, and treated as not invited in the analysis) is unclear. If they had been treated as invited, the relative risk would have been 0.83.]

(vi) Spain

Based on a population-based mammography screening programme targeting women aged 45–64 years in Navarre, Ascunce et al. (2007) reported an IBM relative risk of 0.58 (95% CI, 0.44–0.75) for invitation to screening of women aged 50–69 years in 1997–2001. There was no adjustment for the overall trend in breast cancer mortality; the corresponding relative risk was 1.07 (95% CI, 0.66–1.74) in women aged 30–44 years and 1.03 (95% CI, 0.77–1.37) in those aged 75 years and older (outside the target age group). The relative risk adjusted for the average of these two trends was 0.56 (95% CI, 0.39–0.80).

(vii) Denmark

Based on linked screening registry, cancer registry, cause of death registry, and population register data for individual women, Olsen et al. (2005) analysed IBM for invitation to screening in the first 10 years (1991–2001) of population-based mammography screening offered every 2 years to women aged 50–69 years in Copenhagen. Three comparison groups, Copenhagen in 1981–1991 and Denmark (except Copenhagen and two other areas with population-based screening before 2001) in 1991–2001 and 1981–1991 (secondary control groups to provide data on the underlying trend in breast cancer mortality), were constructed from women’s individual records in the population register, and the women were allocated pseudo-dates of first invitation. In all cases, women with prevalent breast cancer before their real date or pseudo-date of invitation were excluded. Analysis was done by way of a Poisson
Breast cancer screening

A regression model of breast cancer mortality with age, whether invited or not, period, region, and interaction between period and region as covariates, thus adjusting the estimate of effect of invitation for differences in age, place, and time between invited and not invited women. The estimated IBM relative risk for invitation to screening was 0.75 (95% CI, 0.63–0.89).

(viii) Norway

A population-based programme that offers mammography screening every 2 years to women aged 50–69 years began as a pilot programme in four of the 19 Norwegian counties in 1996; roll-out to the rest of the country began 2 years later and was completed in 2005 (Hofvind et al., 2013). Population-based screening was preceded by widespread opportunistic screening, to the extent that 38% of women who had their first mammogram within the programme in 1996–2006 had received a mammogram within the preceding 3 years, and 64% had ever had a mammogram (Hofvind et al., 2013). Also, importantly, the roll-out of population-based screening in Norway was accompanied by or preceded by the establishment of multidisciplinary breast cancer care units in each county, in which all women being investigated or treated for breast cancer (whether screen-detected or not) were managed (Kalager, 2011).

Three studies have reported on IBM in women invited to screening in the Norwegian population-based programme. One included population-based screening experience accumulated to 2001–2002 in women in the four pilot study counties (Olsen et al., 2013). The second included the experience in the whole country to the end of 2005 (Kalager et al., 2010), thus fully with overlapping the first. The third included the experience in the whole country to the end of 2009 (Weedon-Fekjær et al., 2014), thus fully overlapping with both of the others.

Olsen et al. (2013) compared mortality from breast cancer diagnosed after screening began in women in the four pilot screening counties with the corresponding mortality in women in these counties over the 6 years before screening began. They adjusted their comparison for the underlying trend in breast cancer mortality by estimating it in five non-screening counties in similar periods before and after the beginning of 1996. The authors linked individual data obtained from the central population register, cancer registry, and cause of death registry for all women within the scope of their analysis; aggregated data were not required. However, they did not have individual screening data, so women in the screening counties during the screening period were allocated the date of first invitation to screening in their municipality as their first invitation date. Women included in the 6-year control period for the screening counties were allocated pseudo-dates of invitation 6 years before those in the screening period. The maximum period of screening was 6 years. The authors estimated the IBM relative risk for invitation to screening to be 0.89 (95% CI, 0.71–1.12). [This relative risk includes lead-time bias. Also, the underlying downtrend may have been greater in screening counties than in non-screening counties, due to the introduction of multidisciplinary breast cancer care units along with screening.]

The analysis of Kalager et al. (2010) used a similar approach to that of Olsen et al. (2013) except that it covered mammography screening in the period 1996–2005 and had individual data only for women who had been diagnosed with breast cancer. To address effects of the underlying trends in breast cancer mortality, comparisons were made between women invited to screening in 1996–2005 and corresponding women not invited to screening in 1986–1995, and vice versa. The comparisons were made primarily in women aged 50–69 years at diagnosis of breast cancer. [Balanced breast cancer accrual and follow-up periods and age groups avoided lead-time bias. However, as a consequence of the manner of roll-out of population-based screening in
Norway, the group invited to screening and its historical comparison were concentrated in the second halves of the compared periods and the group not invited to screening and its historical comparison were concentrated in the first halves, making the latter a potentially inaccurate estimate of the underlying trend in breast cancer mortality in the group offered screening. The authors estimated the relative risk comparing IBM for the group invited to screening relative to its historical comparison group to be 0.72 (95% CI, 0.63–0.81) and the corresponding relative risk in the group not invited to screening to be 0.82 (95% CI, 0.71–0.93). From these relative risks, the Working Group estimated the IBM for invitation to screening adjusted for the underlying mortality trend to be 0.88 (95% CI, 0.73–1.05). The Working Group noted, in agreement with Olsen et al. (2013), that the mortality trend in areas without screening may not accurately indicate the trend in areas with screening. Weedon-Fekjær et al. (2014) obtained individually linked data for all women, as Olsen et al. (2013) had, and in addition obtained individual dates of screening invitations. Unusually, however, they based their analysis of invitation to screening over the period 1996–2009 on the complete, dynamic population of Norwegian women aged 50–79 years in 1986–2009. Thus, their population of women unexposed to screening included women from 10 years before the implementation of population-based screening; as a result, they drew on nearly 13 million person–years of experience before invitation to screening and only 2.4 million after. The IBM relative risk for invitation to screening was estimated to be 0.72 (95% CI, 0.64–0.79) using a complex Poisson regression modelling approach. The authors noted that they could not exclude possible effects of the establishment of multidisciplinary breast cancer care centres in parallel with the roll-out of the screening programme. The relative risks for invitation to screening of these three, overlapping studies of the Norwegian experience are compatible to the extent that their 95% confidence intervals overlap, although the upper limit for the Weedon-Fekjær et al. (2014) study is less than the point estimates for the other two studies, suggesting that it could be lower. In principle, a lower relative risk in Weedon-Fekjær et al. (2014) would be expected because: it includes a later 4 years of the population-based programme’s experience than the other two studies; it would be based, on average, on longer periods of individual women’s experience in the programme; and it would be less affected by the previous high level of opportunistic screening. It might also, perhaps, be affected by the inclusion of a large volume of pre-screening breast cancer mortality experience, which, in the event of a falling trend in underlying breast cancer mortality, might produce an artificially lower relative risk. There is evidence of such a trend (Kalager et al., 2010), and it could be sufficient to explain the difference between the estimate of Weedon-Fekjær et al. (2014) and those of the other two studies. Although the adjustment for period should have addressed this issue, the statistical dominance of person–years before 1996 may have compromised the effectiveness of this adjustment.

Summary

The IBM relative risks for invitation to screening ranged overall from 0.58 to 0.94, with a median value of 0.78. Lead-time bias was the most common residual bias and would be expected to be conservative. If the Swedish, Finnish, and Norwegian studies that are overlapped substantially or fully by other studies (Duffy et al., 2002b; Jonsson et al., 2007; Anttila et al., 2008; Kalager et al., 2010; Olsen et al., 2013) are removed, the range of the remaining 14 studies is the same as for all 19 studies and the median is little changed, at 0.77. Furthermore, if all Norwegian studies are removed because of the introduction of multidisciplinary breast care centres in parallel with screening, the range remains the same and the
Breast cancer screening

median is 0.76. The United Kingdom Trial of Early Detection of Breast Cancer (relative risk [RR], 0.73; 95% CI, 0.63–0.84) included annual CBE in the intervention.

(b) **Women aged 50–69 years who attended screening**

The design and results of studies reviewed are summarized in Table 5.5. Studies are ordered in the table by the country in which they were conducted (with countries in the order in which their mammography screening programmes were first introduced) and within each country by the earliest date of mammography screening that was included in the analysis.

Most of the studies in Table 5.5 were based on the same mammography experience as was used for analyses of the outcomes of invitation to screening. Self-selection for attendance is an important issue in these analyses because the numerator for the IBM relative risk for breast cancer mortality is based on the experience of women attending screening while the denominator is based on all women in a different era or area who were not invited to screening or on women in the same area and era who chose not to attend screening. Self-selection may bias the IBM relative risk estimate if it creates a difference in the underlying risk of breast cancer death between women attending screening and all women, or women not attending screening.

(i) **Sweden**

Tabár et al. (2001) reported an estimate of the IBM relative risk for women aged 40–64 years attending screening of 0.37 (95% CI, 0.30–0.46). [The estimate appears not to have been adjusted either for self-selection or for the underlying time trend in breast cancer mortality. However, data on this trend in women aged 20–39 years in 1968–1977 or 1988–1996 were reported, and the Working Group used this trend to obtain an adjusted IBM relative risk of 0.46 (95% CI, 0.21–0.97). This relative risk may still be affected by self-selection bias.]

The other three Swedish studies that estimated IBM relative risk for attendance to screening (Duffy et al., 2002a, b; Swedish Organised Service Screening Evaluation Group, 2006a, b; Jonsson et al., 2007) overlapped substantially with one another in their coverage of the screening experience, and Swedish Organised Service Screening Evaluation Group (2006a, b) included the experience of one of the counties analysed in Tabár et al. (2001). These three studies variously covered screening of women aged 40–74 years, but mostly aged 50–69 years, and screening during various parts of the period 1978–2001. The results, each adjusted for self-selection bias, were reasonably similar (Table 5.5): the IBM relative risks were 0.61 (95% CI, 0.55–0.68) for Duffy et al. (2002a, b), 0.57 (95% CI, 0.53–0.62) for Swedish Organised Service Screening Evaluation Group (2006a, b), and 0.70 (95% CI, 0.57–0.86) for Jonsson et al. (2007). The methods of adjustment for self-selection bias were, respectively, that of Duffy et al. (2002a, b), a refinement of that method as reported in Swedish Organised Service Screening Evaluation Group (2006a, b), and the method of Cuzick et al. (1997).

(ii) **Finland**

One Finnish study has estimated the IBM relative risk for attendance to screening (Sarkeala et al., 2008b). Based on the same data set as used for Sarkeala et al. (2008a), this study was designed primarily to assess the effect of different screening centre policies on screening effectiveness. Screened women attended between 1992 and 2003; unscreened women included those residing in the same areas in 1974–1985 and women who were invited in 1992–2003 but did not attend. The IBM relative risk for attendance to screening was 0.63 (95% CI, 0.53–0.75), adjusted for self-selection bias using the method of Cuzick et al. (1997).
(iii) Italy

Based on a similar population of women invited to the first screening round (1991–1993) in Florence described in Paci et al. (2002) (Table 5.4), Puliti & Zappa (2012) followed up women invited to mammography screening every 2 years at age 50–69 years for incidence of breast cancer to 2007 and mortality from breast cancer and other causes to 2008 (Table 5.5). The estimated IBM relative risk for women who had ever been screened relative to those who had never been screened was 0.51 (95% CI, 0.40–0.66). This estimate was adjusted for marital status and small-area deprivation index in the hope of reducing self-selection bias. [There would also have been some lead-time bias because the mortality follow-up period was 1 year longer than the period of incident breast cancer accrual.]

(iv) Canada

Based on data obtained from seven of the 12 provincial mammography screening programmes established in or after 1988 under the Canadian Breast Cancer Screening Initiative, Coldman et al. (2014) reported an IBM relative risk of 0.60 (95% CI, 0.52–0.67) for women who were screened at least once in the period 1990–2009 (Table 5.5). For the seven individual provinces, the relative risk ranged from 0.41 (95% CI, 0.33–0.48) in New Brunswick to 0.73 (95% CI, 0.68–0.78) in Ontario. The analysis was based on 20.2 million person–years of experience. Population data from Statistics Canada indicated that 32.4% (Ontario) to 53.0% (New Brunswick) of women aged 50–69 years attended screening in 2005–2006 and that 56.1% (Manitoba) to 64.3% (Quebec) reported undergoing bilateral mammography during the same period. An ad hoc method (described fully in the authors’ online supplementary methods) was used to adjust the relative risk in British Columbia for self-selection.

(v) Denmark

Olsen et al. (2005), who estimated the IBM relative risk for women invited to screening in the Copenhagen population-based mammography programme (Table 5.4), also estimated the IBM relative risk for women screened relative to those not screened, which was 0.60 (95% CI, 0.52–0.67) unadjusted for self-selection for screening. The relative risk adjusted for self-selection using an ad hoc approach was estimated to be 0.63 (95% CI not reported).

(vi) Norway

In a study of women invited to attend the Norwegian Breast Cancer Screening Program, Hofvind et al. (2013) compared breast cancer mortality in women who accepted the invitation with that in women who did not (Table 5.5). This study is based entirely on linked unit record data of individual women invited to attend a population-based mammography screening programme, which included screening history, cancer registrations, and death records. Women could contribute person–years of experience to both the unscreened and the screened group. Overall, 84% of women attended screening for 1–15 years, with a median of 4.5 years. Accrual of incident breast cancers ended in 2009, and emigration and mortality follow-up continued until the end of 2010. The relative risk of death from breast cancer in screened relative to unscreened women was estimated to be 0.57 (95% CI, 0.51–0.64) adjusted for age at breast cancer diagnosis, calendar year, time since inclusion in the unscreened or screened group, and self-selection bias estimated using the average estimate of the breast cancer mortality relative risk for non-attenders relative to uninvited women (1.36; 95% CI, 1.11–1.67, from Duffy et al. 2002a, b) and the study estimate of attendance in response to a screening invitation. The authors noted that 38% of women first attending the Norwegian Breast Cancer Screening Program in 1996–2006 reported having had a mammogram within the
preceding 3 years, which could have biased the estimate of programme effectiveness. They also noted that the contemporaneous introduction of multidisciplinary breast care centres should not have biased their relative risk estimates because only women who were invited to the programme were included in the analysis. [No adjustment was made for lead-time bias.]

(vii) USA

Morrison et al. (1988) examined breast cancer mortality within the Breast Cancer Detection Demonstration Project, which was initiated in 1973 by the American Cancer Society and the National Cancer Institute to demonstrate the feasibility of large-scale screening for breast cancer (Beahrs et al., 1979; Baker, 1982). Screening was initially with two-view mammography, CBE, and thermography, but in later years thermography was dropped and mammography use was reduced, particularly in women younger than 50 years. Morrison et al. (1988) estimated the ratios (observed to expected) for death from breast cancer to be 0.80 overall and 0.89, 0.76, and 0.74, respectively, for women aged 35–49, 50–59, and 60–74 years at entry. [No confidence intervals or P values were reported.]

A case–cohort study approach was used by Thompson et al. (1994) to evaluate the effect of a mammography screening programme offered from 1985 to eligible members of a health maintenance organization in Washington State. Women aged 40–49 years were offered screening in the programme only if they had a risk factor for breast cancer, and women aged 50 years and older were invited every 1–3 years, depending on their risk factors; all were recommended to have annual CBE. A randomly selected age-stratified sample representing 2.4% of women was selected as a subcohort to represent the experience of all women in the cohort in the analysis. The formal screening programme began in 1985 and included mammography every 1–3 years depending on risk and annual CBE. About 10% of the women had been screened before implementation of the programme. By 1988 (3.5 years after implementation of the programme), about 34–56% of women (depending on age) had been screened. The IBM relative risk adjusted for mother’s history of breast cancer, nulliparity, and history of previous breast biopsy was 0.61 (95% CI, 0.23–1.62) for women aged 50 years and older.

Summary

The IBM relative risks for attendance to screening ranged from 0.51 to 0.80 after adjustment for self-selection. The lower value of 0.46 of Tabár et al. (2001) was not adjusted for selection bias, and it is likely that the value of 0.51 of Puliti & Zappa (2012) was incompletely adjusted for self-selection bias. The relative risks for the remaining studies ranged from 0.57 to 0.80 (median, 0.60) when including only the largest of the substantially overlapping Swedish studies (Swedish Organised Service Screening Evaluation Group, 2006a, b). The two studies in the USA (RR, 0.80 for each) included CBE in the intervention.

(c) Women younger than 50 years or older than 69 years

Only studies designed to separate the effect of screening on breast cancer mortality in a specified age group were considered to be informative. To study effectiveness of screening in women younger than 50 years, the analysis of breast cancer mortality should be limited to deaths in women whose breast cancer was diagnosed when they were younger than 50 years, unless screening was offered only to women while they were younger than 50 years (see Section 4.2.1 for discussion of age creep). Similarly, to study effectiveness of screening in women older than 69 years, the analysis should be limited to women first offered screening when they were older than 69 years and to breast cancer deaths that followed a diagnosis of breast cancer when the women were older than 69 years. Only results of studies that
meet these criteria are included in this section. Studies are not included that presented age-specific results for women younger than 50 years but included deaths from breast cancers diagnosed at later ages (UK Trial of Early Detection of Breast Cancer Group, 1999; Coldman et al., 2014) or for women older than 69 years at death from breast cancer who had not been offered screening (Ascunce et al., 2007; Sarkeala et al., 2008b; Kalager et al., 2010; Weedon-Fekjær et al., 2014) or had not been first offered screening in this age group (Jonsson et al., 2007).

The design and results of studies reviewed for this section are summarized in Table 5.6, by age (younger than 50 years or older than 69 years) and by country (in the order in which their mammography screening programmes were first introduced), and within each country by the earliest date of mammography screening that was included in the analysis.

(i) Women younger than 50 years

Sweden

Jonsson et al. (2000) compared IBM in women with breast cancer diagnosed at age 40–49 years in 14 Swedish study-group areas in which population-based mammography screening was offered from age 40 years and 15 control-group areas in which it was offered from age 50 years. These areas excluded five in which RCTs of screening had been conducted, one in which screening had been introduced very early, and one that offered screening from age 45 years. Women in the study group entered the study when screening started in their area. In both groups, mortality follow-up was to age 59 years, creating the possibility of lead-time bias in the result. A geographically identical, historical reference period (1976–1986) was defined for the study group and for the control group. The estimated IBM relative risk for women invited to screening at age 40 years was 0.91 (95% CI, 0.72–1.15), compared with the geographical areas that started screening at age 50 years, and adjusting for year of follow-up, geographical area, and time period. [Geographical area, as included in the model, was not defined but is likely to have been highly correlated with invitation to screening; therefore, the reported relative risk may be unreliable.]

The mammography screening experience of Jonsson et al. (2007) overlaps almost completely with that analysed by Jonsson et al. (2000), and also compares IBM in women invited and not invited to screening over unbalanced time periods. The IBM relative risk for invitation to screening in women aged 40–49 years was 0.64 (95% CI, 0.43–0.97). [The Working Group estimated the IBM relative risk to be 0.51 (95% CI, 0.29–0.90) after adjustment for the difference in underlying breast cancer mortality with reference to results in the authors’ Table 3. Lead-time bias was estimated to be −5%.

Hellquist et al. (2011) updated the analysis of Jonsson et al. (2000) and extended the period of accrual of breast cancer cases from 1997 to 2005. Women in 34 Swedish counties or screening areas were considered invited to screening if they resided when aged 40–49 years in an area that invited women of this age to screening (the same logic was applied for uninvited women in control areas during 1986–2005, with the same average follow-up time and mid-calendar year of follow-up). Such areas were required to have offered screening to women aged 40–49 years for at least 6 years from 1986 to 2005 (mean, 15.8 years). Only breast cancers incident at age 40–49 years were included. The IBM relative risk adjusted for misclassification of breast cancer cases in women invited to screening was 0.74 (95% CI, 0.66–0.83). Assuming 1 month and 1 year of lead time produced estimates of lead-time bias of −0.01% and −0.05%, respectively. Adjusted relative risks for breast cancer deaths in women diagnosed at ages 40–44 years and 45–49 years were estimated to be 0.83 (95% CI, 0.70–1.00) and 0.68 (95% CI, 0.59–0.78), respectively. Adjusted relative risks in women who
Table 5.6 Incidence-based mortality studies of the effectiveness of invitation to mammography screeninga mainly in women younger than 50 years or older than 69 years

<table>
<thead>
<tr>
<th>Reference Country</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person–yearsb</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)c</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Women younger than 50 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jonsson et al. (2000) Sweden</td>
<td>29 areas 1986–1997, depending on area 40–49 yr 18–22 mo; average, 20 mo</td>
<td>Invited 2 229 000 Not invited 3 383 000</td>
<td>3–10 yr average, 8.0 yr</td>
<td>1986–1996 Same</td>
<td>40–49 yr Same + 10 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Same period; different areas</td>
<td>No (follow-up in study population was from start of screening in each area; in control population, it was from 1987)</td>
<td>Year of follow-up, area, time period</td>
<td>0.91 (0.72–1.15)</td>
<td>Lead-time bias estimated to be −0.4%, and inclusion bias −3% RR was 0.97 after excluding &gt; 8 yr of follow-up from control group</td>
</tr>
<tr>
<td>Jonsson et al. (2007) Sweden</td>
<td>4 counties 1989 40–49 yr average, 20–22 mo</td>
<td>Invited 485 468 Not invited 387 173</td>
<td>7 yr</td>
<td>1989–1996 Same</td>
<td>40–49 yr Same + 10 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Different period (accrual 1989–1996 for study group, 1988–1996 for control group); different areas</td>
<td>No (study group follow-up to 2001, control group to 1998)</td>
<td>Not stated</td>
<td>[0.51 (0.29–0.90)]d</td>
<td>RR adjusted by the Working Group for difference in underlying breast cancer mortality. Lead-time bias estimated to be −5%</td>
</tr>
<tr>
<td>Reference Country</td>
<td>Areas, earliest year of programme screening, screening age, screening interval</td>
<td>Person–years&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Duration of screening</td>
<td>Accrual and follow-up periods</td>
<td>Diagnosis and death age ranges</td>
<td>Individual or aggregate data</td>
<td>Temporal and geographical similarity of comparison group</td>
<td>Time-balanced follow-up periods?</td>
<td>Adjustments</td>
<td>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Comments</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-------------------------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------</td>
<td>--------------------------------</td>
<td>-------------------------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Hellquist et al. (2011)</strong> Sweden</td>
<td>34 areas 1986–1997, depending on area 40–49 yr 18 mo</td>
<td>Invited 6 994 421 Not invited 8 843 852</td>
<td>6–20 yr 1986–2005 Same</td>
<td>40–49 yr 40–68 yr</td>
<td>Individual for women who died of breast cancer; aggregate, all other women</td>
<td>Same period; different areas (3 of 34 areas changed status)</td>
<td>Yes</td>
<td>Breast cancer cases in study-group women known not to have been invited to screening; contamination in control group</td>
<td>Invited to screening: [0.79 (0.67–0.92)] Ever screened: [0.76 (0.64–0.89)&lt;sup&gt;d&lt;/sup&gt;]</td>
<td>RR adjusted for pre-screening differences in breast cancer mortality. Lead-time bias estimated to be −0.01% to −0.05%</td>
<td></td>
</tr>
<tr>
<td><strong>Hakama et al. (1995)</strong> Finland</td>
<td>City of Kotka 1982 40–51 yr 2 yr</td>
<td>Invited to screening 38 220 Attended screening 32 910 Not screened 56 233</td>
<td>8–9 yr 1982–1990 Same + 1 yr</td>
<td>40–54 yr 40–55 yr</td>
<td>Yes</td>
<td>Same period; same area</td>
<td>Yes</td>
<td>Age</td>
<td>Invited to screening: 0.11 (0.00–0.71) Attended screening: 0.10 (0.00–0.53)</td>
<td>Screening included CBE. Lead-time bias possible. RR based on 1 breast cancer death. Programme sensitivity estimated to be 25%</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.6 (continued)

<table>
<thead>
<tr>
<th>Reference Country</th>
<th>Areas, earliest year of programme screening, screening age, screening interval</th>
<th>Person-years</th>
<th>Duration of screening</th>
<th>Accrual and follow-up periods</th>
<th>Diagnosis and death age ranges</th>
<th>Individual or aggregate data</th>
<th>Temporal and geographical similarity of comparison group</th>
<th>Time-balanced follow-up periods?</th>
<th>Adjustments</th>
<th>Breast cancer mortality RR (95% CI)</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Women older than 69 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jonsson et al. (2003b) Sweden</td>
<td>23 areas 1986–1990, depending on area 70–74 yr 22.8 mo</td>
<td>Invited 1 251 000 Not invited 580 000</td>
<td>8–12 yr average, 8.1 yr</td>
<td>1986–1998 Same 70–74 yr; Same + 12 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Same period; different areas</td>
<td>Yes</td>
<td>Age during follow-up, area, time period</td>
<td>Underlying cause of death: 0.96 (0.73–1.25) Excess mortality estimate: 0.84 (0.59–1.19)</td>
<td>RR adjusted for both inclusion bias and lead-time bias was estimated to be 0.93 for underlying cause of death and 0.78 for excess mortality (95% CIs not reported)</td>
<td></td>
</tr>
<tr>
<td>Van Dijck et al. (1997) The Netherlands</td>
<td>2 cities 1977 68–83 yr 2 yr</td>
<td>Invited 60 313 Not invited 61 832</td>
<td>13 yr</td>
<td>1977–1990 Same 68–95 yr 68–95 yr</td>
<td>Individual for breast cancer cases; aggregate, all other women</td>
<td>Same period; different cities</td>
<td>Yes</td>
<td>Difference in underlying risk of breast cancer in the 2 cities</td>
<td>[0.89 (0.56–1.40)]</td>
<td>Women first invited to screening at age 68 yr or older; 46% of invited women screened once or more</td>
<td></td>
</tr>
<tr>
<td>Reference Country</td>
<td>Areas, earliest year of programme screening, screening age, screening interval</td>
<td>Person–years&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Duration of screening</td>
<td>Accrual and follow-up periods</td>
<td>Diagnosis and death age ranges</td>
<td>Individual or aggregate data</td>
<td>Temporal and geographical similarity of comparison group</td>
<td>Time-balanced follow-up periods?</td>
<td>Adjustments</td>
<td>Breast cancer mortality RR (95% CI)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Comments</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>--------------------------------</td>
<td>-----------------------------</td>
<td>---------------------------------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>-----------------------------</td>
<td>----------</td>
</tr>
<tr>
<td><strong>Coldman et al. (2014)</strong></td>
<td>Canada 7 provinces 1990 70–79 yr No recall after 69 yr</td>
<td>Screened and not screened at all ages 20 200 000 (Analysis for screening 70–79 yr based on 4 of 7 provinces)</td>
<td>1–20 yr, all women not known, women 70–79 yr</td>
<td>1990–2009 Same</td>
<td>70–99 yr Same</td>
<td>Individual for screened women; aggregate for unscreened women</td>
<td>Same period; same population</td>
<td>Yes</td>
<td>Age</td>
<td>0.65 (0.56–0.74)</td>
<td>Not adjusted for self-selection bias (see Table 5.5). Analysis based on age at first participation in organized screening; previous opportunistic screening cannot be excluded</td>
</tr>
</tbody>
</table>

<sup>a</sup> Two studies evaluated invitation to mammography plus CBE.

<sup>b</sup> Person–years: number of women or number of breast cancer deaths.

<sup>c</sup> RRs are for breast cancer as the underlying cause of death when alternative estimates (e.g. excess mortality) are also provided.

<sup>d</sup> RRs and 95% CIs adjusted for trend or geographical difference in underlying mortality were calculated as ratio of the authors’ estimated RRs comparing screening area with control period or area; 95% CI of ratio estimated using the method in Altman & Bland (2003) as implemented in http://www.hutchon.net/CompareRR.htm. With reference to Hellquist et al. (2011); see also Weedon-Fekjær et al. (2014). CI, confidence interval; mo, month or months; RR, relative risk; yr, year or years.
Breast cancer screening

attended screening were 0.71 (95% CI, 0.62–0.80), 0.82 (95% CI, 0.67–1.00), and 0.63 (95% CI, 0.54–0.75) for the age groups 40–49, 40–44, and 45–49 years, respectively. These estimates were made by adjusting the estimates for invitation to screening using the method of Cuzick et al. (1997). The above estimates were not adjusted for a pre-screening difference in breast cancer mortality (RR, 0.94; 95% CI, 0.85–1.05) between screening and non-screening areas; taking this into account, the Working Group calculated an IBM relative risk of [0.79 (95% CI, 0.67–0.92)] for invited women and [0.76 (95% CI, 0.64–0.89)] for women who were ever screened using a method developed by Altman & Bland (2003).

Finland

Mammography was initiated on a pilot basis in Finland in the early 1980s. Women born in 1940 or 1942 were invited to attend screening with mammography and CBE in 1982; women born in 1936 or 1938 were invited in 1983, and thus they were aged 40–47 years at entry. They were re-invited every 2 years until 1990 (a total of four or five invitations), and women were considered to be non-attenders if they did not attend the first round. Women born in alternate years from 1935 to 1943 were used as a control cohort. The IBM relative risk was 0.11 (95% CI, 0.00–0.71) for invitation to screening and 0.10 (95% CI, 0.00–0.53) for attendance to screening (Hakama et al., 1995). [The Working Group agreed with the authors’ opinion that an estimated programme sensitivity of 25% was too low for programme effectiveness to be the sole explanation for the very low relative risk.]

Summary

The Swedish study of Hellquist et al. (2011) encompassed the whole screening experience covered by Jonsson et al. (2000) and Jonsson et al. (2007) and provided IBM relative risks of 0.74 (95% CI, 0.66–0.83) for being invited to screening and 0.71 (0.62–0.80) for being ever screened. No weight was given to the very low relative risk that Hakama et al. (1995) observed, because it was based on only one death and appears incompatible with the estimated screening programme sensitivity of 25%.

(ii) Women older than 69 years

Sweden

The results of Jonsson et al. (2003b) are similar to those of Jonsson et al. (2000), except that the analysis was based on first invitation to screening of women aged 65–74 years and covered 23 areas (16 study-group areas and 7 control-group areas) and not 29; the additional exclusions were principally counties in which screening did not begin until after 1990. The mean follow-up time was 10.1 years in the study group (8.1 years if estimated individual date of first screening was used, and not date of start of the screening programme in each area) and 9.3 years in the control group. Breast cancer deaths included in the analysis were only those that followed a diagnosis of breast cancer at age 70–74 years. The IBM relative risk for invitation to screening was 0.96 (95% CI, 0.73–1.25) when breast cancer mortality was based on underlying cause of death and adjusted for the difference in underlying mortality between the study-group and control-group areas. With further adjustment for inclusion bias and lead-time bias, the relative risk was 0.93 (95% CI not reported). The authors argued that the underlying cause of death may have been a particularly inaccurate classifier of mortality due to breast cancer in older women and that an excess mortality estimate would be more accurate. The corresponding excess mortality estimate of the relative risk was 0.84 (95% CI, 0.59–1.19) adjusted for the difference in underlying mortality between the study-group and control-group areas; with further adjustment for inclusion bias and lead-time bias, the relative risk was 0.78.
Jonsson et al. (2007) also reported on IBM associated with invitation to screening at age 70–74 years. However, the two screening counties in this study were also study-group (screening) counties in the Jonsson et al. (2003b) study, and the periods covered by the two studies were nearly the same. Therefore, Jonsson et al. (2007) was not considered to provide independent evidence.

The Netherlands

Van Dijck et al. (1997) reported on IBM in women first invited to mammography screening at age 68–83 years in the city of Nijmegen compared with that in the city of Arnhem over an accrual and follow-up period of 1977–1990. Attendance rates in Nijmegen fell sharply with age, from approximately 70% in women in their late sixties to about 40% in those in their seventies and to less than 20% for the first round and less than 10% for the second and later rounds in women in their eighties and nineties. Screening began in Arnhem in 1989. The IBM relative risk for invitation to screening over the whole study period was estimated to be 0.80 (95% CI, 0.53–1.22), which became [0.89 (95% CI, 0.56–1.40)] when adjusted for the estimated difference in underlying breast cancer mortality between Nijmegen and Arnhem (see Table 5.6). For the period 9–13 years after the start of screening, the IBM relative risk estimate was 0.53 (95% CI, 0.27–1.04), and 0.59 (95% CI, 0.30–1.16) after adjusting for the difference in underlying breast cancer mortality.

Canada

In the Canadian provincial mammography screening programmes (Coldman et al., 2014), the relative risk for women first screened at age 70–79 years was 0.65 (95% CI, 0.56–0.74) in the four provinces that offered screening to women in this age group. The province-specific relative risks varied from 0.63 (95% CI, 0.49–0.76) to 0.84 (95% CI, 0.36–1.31). The authors estimated that self-selection bias was conservative (−9% in an analysis limited to women aged 40–49 years in British Columbia). [This estimate may not be applicable to screening of women aged 70–79 years. Also, opportunistic breast screening before first screening in the provincial programmes could have affected the reported results, particularly in the age group 70–79 years.]

Summary

Three studies reported potentially valid estimates of IBM relative risks for breast cancer mortality in women older than 69 years: one for the age group 68–83 years (Van Dijck et al., 1997), one for 65–74 years (Jonsson et al., 2003b), and one for 70–79 years (Coldman et al., 2014). The reported relative risks, of 0.89 (95% CI, 0.56–1.40) by Van Dijck et al. (1997), 0.96 (95% CI, 0.73–1.25) by Jonsson et al. (2003b), and 0.65 (95% CI, 0.56–0.74) by Coldman et al. (2014), are heterogeneous. However, the heterogeneity is reduced if the excess mortality estimate of the relative risk, 0.84 (95% CI, 0.59–1.19), of Jonsson et al. (2003b) is accepted as the more accurate estimate from that study. Lack of adjustment for self-selection bias and lack of consideration of possible effects of previous opportunistic screening limit the weight that can be given to the result of Coldman et al. (2014).

5.2.2 Case–control studies

The reported case–control studies are presented by country in the text and tables. All case–control studies are based on defined populations, but some of these are specific cohorts, with the methods of analysis being a case–control study nested within the cohort. In many case–control studies, the risk estimates are calculated for women who participated in screening compared with women who had been invited (or to whom screening was otherwise offered) but who did not participate. The non-participating women may have a different risk of death from breast cancer compared with the average population (Cuzick et al., 1997; Duffy et al., 2002a;
Breast cancer screening

Swedish Organised Service Screening Evaluation Group, 2006a; Sarkeala et al., 2008a, b), so this may result in selection bias. If the case–control study is based on systematic historical databases on screening, information bias can be considered minimal. However, in other case–control studies, information bias may be a problem. Rather few case–control studies have assessed screening impact compared with expectation in the absence of screening (or invitation) in the average population, as is usually done in cohort mortality studies. There are further limitations in the reported case–control studies in taking into account full screening histories in the risk estimates, and consequently there is wide variation in the follow-up windows for incidence and mortality after index screening. This potentially affects the magnitude of the estimates, even though these follow-up details are not always reported in connection with the individual studies. Some studies used only age at death in matching, whereas most studies also matched on residence at the time of diagnosis of the case. In addition, since the risk of breast cancer could be different among women who attend screening after receiving an invitation compared with those who are invited but do not attend, selection factors may confound the estimates of efficacy. A potential asset in case–control studies is that an adjustment for sociodemographic factors can also be attempted.

(a) Case–control studies within service screening programmes

See Table 5.7.

(i) United Kingdom

Allgood et al. (2008) performed a case–control study in the East Anglia region. The cases were deaths from breast cancer in women diagnosed between the ages of 50 years and 70 years, after the initiation of the East Anglia Breast Screening Programme in 1989. The controls were women (two per case) who had not died of breast cancer, from the same area, matched by date of birth to the cases. Each control was known to be alive at the date of death of her matched case. All women were known to the breast screening programme and had been invited, at least once, to be screened. The unadjusted odds ratio for risk of death from breast cancer in women who attended at least one routine screen compared with those who did not attend was 0.35 (95% CI, 0.24–0.50), and 0.65 (95% CI, 0.48–0.88) after adjusting for self-selection bias using the more conservative intention-to-treat analysis (Duffy et al., 2002a).

Fielder et al. (2004) conducted a case–control study to estimate the effect of service screening, as provided by the NHS Breast Screening Programme, on breast cancer mortality in Wales. The 419 cases were deaths from breast cancer in women aged 50–75 years at diagnosis who were diagnosed after the start of screening in 1991 and who died after 1998. The 717 controls were women who had not died of breast cancer or any other condition during the study period. The aim was to select one control from the same general practitioner’s practice and another from a different general practitioner’s practice within the same district, matched by year of birth. The unadjusted odds ratio for risk of death from breast cancer in women who attended at least one routine screen compared with those who had never been screened was 0.62 (95% CI, 0.47–0.82), and 0.75 (95% CI, 0.49–1.14) after excluding cases diagnosed before 1995 and adjusting for self-selection bias.

(ii) Iceland

Gabe et al. (2007) conducted a case–control study to evaluate the impact of the Icelandic breast screening programme, which was initiated in November 1987 in Reykjavik and covered the whole country from December 1989, comprising biennial invitation to mammography screening for women aged 40–69 years. The cases were deaths from breast cancer matched by age and
Table 5.7 Case–control studies of the effectiveness of mammography screening within service screening programmes, by country

<table>
<thead>
<tr>
<th>Reference</th>
<th>Area, year programme began, screening age and interval, women included</th>
<th>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</th>
<th>Screening exposure; age of included women</th>
<th>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</th>
<th>Linkage or use of screening, cancer registry, death databases; data items available</th>
<th>Issues or items related to screening history; whether prevalent cases were excluded</th>
<th>Adjustments</th>
<th>Breast cancer mortality OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>United Kingdom</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allgood et al. (2008)</td>
<td>East Anglia 1989 50–70 yr active, ≥ 70 yr allowed 3-yearly Women registered with GP</td>
<td>284 East Anglia cancer registry database 1995–2004 from 1995 16 deaths excluded</td>
<td>At least 1 invitation to breast screening 50–70 yr</td>
<td>568 NHS Exeter system database Same source as cases DOB; most were from same health authority as case Alive at DOD of case</td>
<td>All 3 DOB, date of diagnosis, DOD, screening history (time since last screen, number of screens)</td>
<td>Prevalent cases were minimized by restricting to deaths and diagnoses from 1995, 6 yr after start of programme</td>
<td>SES, self-selection bias using method of Duffy et al. (2002a)</td>
<td>0.65 (0.48–0.88) for at least 1 screen</td>
</tr>
<tr>
<td>Fielder et al. (2004)</td>
<td>Wales 1989 50–75 yr 3-yearly Women registered with GP and identified in health authority registers</td>
<td>419 Breast Test Wales database and “standard death registration” 1998–2001 from 1991 84%</td>
<td>At least 1 invitation before date of diagnosis or pseudo-diagnosis 50–75 yr</td>
<td>717 Database of those eligible for screening in Breast Test Wales Year of birth; 1 control from same GP and 1 from other GP Alive at time of diagnosis of case</td>
<td>Breast Test Wales for screening history and breast cancer diagnoses Year of birth, date of diagnosis, screening history (time since last screen, number of screens)</td>
<td>All cancers diagnosed early in the programme in 1991–1994 excluded; controls with breast cancer diagnosis were eligible</td>
<td>Self-selection bias using method of Duffy et al. (2002a)</td>
<td>0.75 (0.49–1.14) for at least 1 screen</td>
</tr>
</tbody>
</table>
Table 5.7 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Area, year programme began, screening age and interval, women included</th>
<th>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</th>
<th>Screening exposure; age of included women</th>
<th>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</th>
<th>Linkage or use of screening, cancer registry, death databases; data items available</th>
<th>Issues or items related to screening history; whether prevalent cases were excluded</th>
<th>Adjustments</th>
<th>Breast cancer mortality OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Iceland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabe et al. (2007)</td>
<td>1987 40–69 yr 2-yearly All women in age group</td>
<td>226 Source not stated 1990–2002 from start of service screening 7 deaths before 1990 excluded</td>
<td>Ever screened before date of diagnosis or pseudo-diagnosis 40–70+ yr</td>
<td>902 National registry Same source as cases DOB, screening area Alive at DOD of case</td>
<td>Probably the national cancer and screening registries DOB, date of diagnosis, DOD, urban/rural residence, screening history (time since last screen, number of screens)</td>
<td>Excluded 7 deaths before 1990; screening history excluded after diagnosis for controls diagnosed with cancer</td>
<td>Self-selection bias using method of Duffy et al. (2002a), and screening opportunity bias</td>
<td>0.65 (0.39–1.09)</td>
</tr>
<tr>
<td><strong>The Netherlands</strong></td>
<td></td>
<td></td>
<td>At least 1 invitation 50–74 yr</td>
<td>785 Same source population as cases Alive and residing in Nijmegen at DOD of case, invited to participate in the index screening round, free of breast cancer at their index invitation</td>
<td>Data on invitation and participation were kept in the screening registry</td>
<td>Analysis includes only women who attended screening</td>
<td>Age at screening</td>
<td>0.68 (0.33–1.41)</td>
</tr>
<tr>
<td>Broeders et al. (2002)</td>
<td>Nijmegen 1975 50–69 yr until 1997; 50–74 yr thereafter 2-yearly All women</td>
<td>157 Screening registry 1987–1997 Last 10 yr of the programme NR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>By age: 40–49 yr: 0.90 (0.38–2.14) 50–59 yr: 0.71 (0.35–1.46) 60–69 yr: 0.80 (0.42–1.54) 70–79 yr: 1.13 (0.50–2.58) &gt; 79 yr: 2.92 (0.55–15.4)</td>
</tr>
<tr>
<td>Reference</td>
<td>Area, year programme began, screening age and interval, women included</td>
<td>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</td>
<td>Screening exposure; age of included women</td>
<td>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</td>
<td>Linkage or use of screening, cancer registry, death databases; data items available</td>
<td>Issues or items related to screening history; whether prevalent cases were excluded</td>
<td>Adjustments</td>
<td>Breast cancer mortality OR (95% CI)</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>van Schoor et al. (2011)</td>
<td>Nijmegen 1975 Invitations sent to women aged ≥ 35 yr</td>
<td>282 Women invited to the screening programme in Nijmegen NR 1975–2008 191 cases were screened and 91 not screened</td>
<td>Screening invitation during a 4-yr period before breast cancer diagnosis of the case (biennial screening schedule including 2 consecutive invitations) 50–69 yr</td>
<td>1410 Same source as cases Eligible for screening, not having breast cancer at the time of invitation, and living in Nijmegen at DOD of case; 5 per case randomly sampled</td>
<td>Separate registry on all breast cancer patients in Nijmegen diagnosed within and outside the screening programme Vital status from the Municipal Personal Records Database Assessments of causes of death by a committee of physicians unaware of the screening history</td>
<td>Including an interaction term, the combination of screening and calendar year, in the logistic regression model; corrected for the confounding influence of age at index invitation by stratification into 5-yr age groups</td>
<td>By calendar period: 1975–2008: 0.65 (0.49–0.87) 1975–1991: 0.72 (0.47–1.09) 1992–2008: 0.35 (0.19–0.64)</td>
<td>0.65 (0.49–0.87) 1975–2008: 0.65 (0.49–0.87) 1975–1991: 0.72 (0.47–1.09) 1992–2008: 0.35 (0.19–0.64)</td>
</tr>
<tr>
<td>Reference</td>
<td>Area, year programme began, screening age and interval, women included</td>
<td>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</td>
<td>Screening exposure; age of included women</td>
<td>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</td>
<td>Linkage or use of screening, cancer registry, death databases; data items available</td>
<td>Issues or items related to screening history; whether prevalent cases were excluded</td>
<td>Adjustments</td>
<td>Breast cancer mortality OR (95% CI)</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
<td>--------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Paap et al. (2010)</td>
<td>Limburg Province 1989 50–75 yr every 2 yr Women aged 50–75 yr who received at least 1 invitation to screening in the region</td>
<td>118 Women invited to screening in IKL region Deaths between 2004 and 2005 Years of diagnosis NR Proportion of eligible cases included NR</td>
<td>Received at least 1 invitation to the service screening programme 50–75 yr</td>
<td>118 Same source population as cases Matched for year of birth and area of residence Alive at DOD of case</td>
<td>IKL includes a screening registry and a cancer registry Cause of death was determined by linkage to Statistics Netherlands For cases, DOD, DOB, date of diagnosis</td>
<td>For cases and controls, complete screening history was obtained from the screening registry. Controls with breast cancer diagnosis at time of invitation to screening were excluded</td>
<td>Self-selection bias</td>
<td>0.24 (0.10–0.58)</td>
</tr>
<tr>
<td>Reference</td>
<td>Area, year programme began, screening age and interval, women included</td>
<td>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</td>
<td>Screening exposure; age of included women</td>
<td>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</td>
<td>Linkage or use of screening, cancer registry, death databases; data items available</td>
<td>Issues or items related to screening history; whether prevalent cases were excluded</td>
<td>Adjustments</td>
<td>Breast cancer mortality OR (95% CI)</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Paap et al. (2014)</td>
<td>(5 of 9 screening regions) 1990 50–74 yr 2-yearly All women</td>
<td>1233 Netherlands Cancer Registry 2004 or 2005 from start of service screening Proportion NR</td>
<td>Screened at index invitation (most recent before diagnosis of case) or the preceding screening round 50–75 yr</td>
<td>2090 Women in 5 regions with at least 1 screening invitation Same source as cases Year of birth, area of residence, screening invitation in same round as case index invitation Alive at DOD of case</td>
<td>All 3 DOB, date of diagnosis, DOD, screening history (time since last screen, number of screens)</td>
<td>Screening participation restricted to maximum 2 rounds</td>
<td>Self-selection bias using correction factor for each region based on IBM method (Paap et al., 2011), and for screening opportunity bias (control matched to screening round of index invitation of case)</td>
<td>0.42 (0.33–0.53)</td>
</tr>
</tbody>
</table>
## Table 5.7 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Area, year programme began, screening age and interval, women included</th>
<th>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</th>
<th>Screening exposure; age of included women</th>
<th>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</th>
<th>Linkage or use of screening, cancer registry, death databases; data items available</th>
<th>Issues or items related to screening history; whether prevalent cases were excluded</th>
<th>Adjustments</th>
<th>Breast cancer mortality OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Otto et al. (2012b)</strong></td>
<td>South-western region 1990 50–69 yr (extended to 75 yr in 1998) 24.5 mo  All female residents</td>
<td>755 Cohort of women invited by the screening organization in south-western Netherlands 1995–2003 1990–2003 98.6%</td>
<td>Index period: time period from index invitation backward to a maximum of 2 invitations before the index invitation; total number of invitations varied from 1 to 3 per case–control set 50–75 yr</td>
<td>3739 Same source as cases 5 controls per case, matched on year of birth, year of first invitation, and number of invitations before diagnosis of case</td>
<td>Linkage with cause of death registry and cancer registry, Comprehensive Cancer Centre Rotterdam, and Statistics Netherlands</td>
<td>Screening histories for all women ever invited to a mammography screening examination were systematically retrieved from the same database</td>
<td>Self-selection bias using method of Duffy et al. (2002a) and own correction factor</td>
<td>49–75 yr: 0.51 (0.40–0.66) 50–69 yr: 0.61 (0.47–0.79) 50–75 yr: 0.52 (0.41–0.67) 70–75 yr: 0.16 (0.09–0.29)</td>
</tr>
<tr>
<td><strong>Italy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Puliti et al. (2008)</strong></td>
<td>Northern and central Italy, 5 regions 1990 50–69 yr 2-yearly</td>
<td>1750 Regional mortality registers 1988–2002 from year before start of service screening to end of 2001 Proportion NR</td>
<td>Any service screen before date of diagnosis or pseudo-diagnosis 50–74 yr</td>
<td>7000 All women 50–69 yr resident in the selected areas for any period of time Same source as cases DOB and resident in the municipality in year of death of subject</td>
<td>IMPACT database used cancer, screening, and mortality registers DOB, screening history (screening in 3 yr before diagnosis of case, number of screens)</td>
<td>Not-yet-invited women included in unscreened; free of breast cancer diagnosis before diagnosis date of case</td>
<td>Self-selection bias using method of Duffy et al. (2002a) and own correction factor</td>
<td>0.55 (0.36–0.85)</td>
</tr>
</tbody>
</table>
### Table 5.7  (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Area, year programme began, screening age and interval, women included</th>
<th>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</th>
<th>Screening exposure; age of included women</th>
<th>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</th>
<th>Linkage or use of screening, cancer registry, death databases; data items available</th>
<th>Issues or items related to screening history; whether prevalent cases were excluded</th>
<th>Adjustments</th>
<th>Breast cancer mortality OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Roder et al. (2008)</strong></td>
<td>South Australia 1989 50–69 yr active; 40–49 yr and ≥ 70 yr allowed 2-yearly</td>
<td>491 South Australia Cancer Registry 2002–2005 from 1994 94%</td>
<td>BreastScreen attendance before date of diagnosis or pseudo-diagnosis 45–80 yr</td>
<td>1473 Electoral roll Same source as cases</td>
<td>All 3 DOB, screening history (number of screens)</td>
<td>Date of breast cancer diagnosis for case; only if date of diagnosis in controls later than in case</td>
<td>SES, remoteness, access (ARIA)</td>
<td>0.59 (0.47–0.74)</td>
</tr>
<tr>
<td><strong>Nickson et al. (2012)</strong></td>
<td>Western Australia mid-1990s 50–69 yr active; 40–49 yr allowed 2-yearly</td>
<td>427 Western Australia Cancer Registry 1995–2006 from 1995 Proportion NR</td>
<td>Receiving a screening mammogram between age 50 yr and reference date 50–69 yr</td>
<td>Average 8.5 controls per case Electoral roll 1995–2006 Same source population as cases Month and year of birth of case; Western Australia resident at time of diagnosis of cases Alive at DOD of case</td>
<td>All 3 DOB, date of any cancer diagnosis, DOD, screening history (year of first screen)</td>
<td>Earliest breast cancer diagnosis in case–control set; women were excluded if they had a screen before age 50 yr</td>
<td>SES, remoteness, HRT use, family history of breast cancer</td>
<td>0.48 (0.38–0.59)</td>
</tr>
</tbody>
</table>

ARIA, Accessibility/Remoteness Index of Australia; CI, confidence interval; DOB, date of birth; DOD, date of death; GP, general practitioner; HRT, hormone replacement therapy; IBM, incidence-based mortality; IKL, Comprehensive Cancer Centre Limburg; mo, month or months; NHS, National Health System; NR, not reported; OR, odds ratio; SES, socioeconomic status; yr, year or years.
Breast cancer screening

screening area to population-based controls. The unadjusted odds ratio for risk of death from breast cancer in women who attended at least one screen compared with those who had never been screened was 0.59 (95% CI, 0.41–0.84), and 0.65 (95% CI, 0.39–1.09) after correction for both self-selection bias and screening opportunity bias.

(iii) The Netherlands

Broeders et al. (2002) conducted a case–control study to describe the effect of population-based mammography screening in Nijmegen on breast cancer mortality, based on a 20-year follow-up period. The risk of death from breast cancer was calculated per 10-year moving age group for women who had attended the index screening (the screening immediately before diagnosis of breast cancer) versus those who had not. Odds ratios were presented by age group for both participation in index screening (see Table 5.7) and participation in either the index screening or the previous screening, or both; none showed a statistically significant effect. The youngest 10-year age group that showed an effect was women aged 45–54 years at their index screening; the odds ratio in women aged 45–49 years was 0.56 (95% CI, 0.20–1.61). The odds ratios for women aged 40–49 years were 0.90 (95% CI, 0.38–2.14) for participation in the index screening and 0.84 (95% CI, 0.30–2.29) for participation in the index screening and the previous screening. The corresponding odds ratios for women aged 70–79 years were 1.13 (95% CI, 0.50–2.58) and 0.70 (95% CI, 0.32–1.54). There was no limitation in these analyses as to age at first attendance to screening. [This analysis overlaps partly with that of van Schoor et al. (2010) (see Section 5.2.2b).]

By 2008, 55,529 women had received an invitation to screening in Nijmegen, and another case–control study was performed (van Schoor et al., 2011). The odds ratio for breast cancer death in the screened group over the complete period was 0.65 (95% CI, 0.49–0.87). Analyses were also performed by calendar period of index invitation to screening (see Table 5.7). [It is unclear why the numbers analysed for the two screening periods are so much less than the overall total of cases and controls included in this study.]

Paap et al. (2010) designed a case–control study to investigate the effect of mammography screening at the individual level. The study population included all women aged 50–75 years in Limburg Province who had been invited to the screening programme in 1989–2006. The unadjusted odds ratio for the screened versus the unscreened women was 0.30 (95% CI, 0.14–0.63), and 0.24 (95% CI, 0.10–0.58) after adjustment for self-selection. [This analysis includes only deaths in the most recent screening years. Deaths in the period from inception of the programme in 1989 until 2003 were not included.]

Paap et al. (2014) estimated the effect of the Dutch screening programme on breast cancer mortality by means of a large multiregion case–control study. They identified all breast cancer deaths in 2004 and 2005 in women aged 50–75 years who had received at least one invitation to the service screening programme in five participating screening regions. Cases were individually matched to controls from the population invited to screening. Conditional logistic regression was used to estimate the odds ratio of breast cancer death according to individual screening history. The unadjusted odds ratio for breast cancer death in screened versus unscreened women was 0.48 (95% CI, 0.40–0.58), and 0.42 (95% CI, 0.33–0.53) after adjustment for self-selection bias using regional correction factors for the difference in the baseline risk of breast cancer death between screened and unscreened women.

Otto et al. (2012b) conducted a case–control study in the south-western region of the Netherlands for the period 1995–2003, including women aged 49–75 years. There was no restriction with respect to age at first invitation. The all-age odds ratio for the association between attending screening at the index invitation and
risk of breast cancer death was 0.56 (95% CI, 0.44–0.71), and 0.51 (95% CI, 0.40–0.66) for women attending any of the three screening examinations (for analyses by age at the index invitation, see Table 5.7).

(iv) Italy

Puliti et al. (2008) conducted a case–control study to evaluate the impact of service screening programmes on breast cancer mortality in five regions of Italy. The odds ratio for invited women compared with not-yet-invited women was 0.75 (95% CI, 0.62–0.92). When the analyses were restricted to invited women, the odds ratio for screened women compared with never-respondent women, corrected for self-selection bias, was 0.55 (95% CI, 0.36–0.85).

(v) Australia

Roder et al. (2008) conducted a case–control study of women in South Australia aged 45–80 years during 2002–2005 (diagnosed after the start of BreastScreen Australia) and live controls (three per death) randomly selected from the state electoral roll after date-of-birth matching. The programme has provided biennial screening, with two-view mammography and double reading, since its inception. It actively targets women aged 50–69 years and allows access to women aged 40–49 years and those aged 70 years and older. The odds ratio for breast cancer death in all BreastScreen participants compared with non-participants was 0.59 (95% CI, 0.47–0.74). The corresponding odds ratio in women younger than 50 years at diagnosis was 1.18 (95% CI, 0.70–1.98) and in those aged 70 years and older at diagnosis was 0.43 (95% CI, 0.25–0.72). Compared with non-participants, the odds ratio was 0.70 (95% CI, 0.47–1.05) for women last screened through BreastScreen more than 3 years before diagnosis of the index case, and 0.57 (95% CI, 0.44–0.72) for women screened more recently.

Nickson et al. (2012) conducted another case–control study within BreastScreen Australia, in which women aged 50–69 years on the electoral roll (98.9% of the eligible population) are invited to attend screening. Eligible women were those aged 50 years and older on the Western Australian electoral roll between 1995 and 2006. The cases were women from this population who died of breast cancer between 1995 and 2006. Controls (10 per case) were selected by incidence density sampling from the source population (those with a breast cancer diagnosis were not excluded). Exposure to screening was defined as receipt of a screening mammogram from BreastScreen at any point between the woman’s 50th birthday and the case–control set reference date (the date of earliest breast cancer diagnosis for that set; for 89%, this was the date of diagnosis of the case); 56% of controls and 39% of cases attended screening. The odds ratio from the primary analyses (adjusted for remoteness and relative socioeconomic disadvantage) was 0.48 (95% CI, 0.38–0.59). The odds ratio was found to vary little by reference age group or year of death and was robust to sensitivity analyses.

(b) Other case–control studies

See Table 5.8.

(i) The Netherlands

In 1974, de Waard et al. (1984a) set up a population-based study of periodic screening by xeromammography of women aged 50–64 years in Utrecht; 72% of invited women attended the first of four rounds. The effect of the programme on breast cancer mortality was evaluated in a nested case–control study, which showed an odds ratio for breast cancer mortality in women who had ever been screened of 0.30 (95% CI, 0.13–0.70) compared with those who had never been screened (Collette et al., 1984). The odds ratios for women aged 50–54, 55–59, 60–64, and 65–69 years at diagnosis were 1.13, 0.31, 0, and 0.10, respectively. [These estimates were based on
<table>
<thead>
<tr>
<th>Reference</th>
<th>Area, year programme began, screening age and interval, women included</th>
<th>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</th>
<th>Screening exposure; age of included women</th>
<th>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</th>
<th>Linkage or use of screening, cancer registry, death databases; data items available</th>
<th>Issues or items related to screening history; whether prevalent cases were excluded</th>
<th>Adjustments</th>
<th>Breast cancer mortality OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Netherlands</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.30 (0.13–0.70)</td>
</tr>
<tr>
<td>Collette et al. (1984)</td>
<td>Utrecht 1974 50–64 yr at the start of the project  All women born in 1911–1925 (72% attended screening)</td>
<td>46 Birth cohort under study 1974–1981 Screening at the first visit and after 12, 18, and 24 mo 50% screened</td>
<td>Screening at first visit and after 12, 18, and 24 mo 50–64 yr</td>
<td>138 Birth cohort under study, same source 3 controls for each case, lived in Utrecht when the case died and same year of birth as case 43% screened</td>
<td>All breast cancer patients included in breast cancer registry; dates of diagnosis checked with general practitioners’ registries</td>
<td>Screening history for the time up to and including date of diagnosis of case</td>
<td>Stratification by birth cohort or age</td>
<td></td>
</tr>
</tbody>
</table>
| Miltenburg et al. (1998) | Utrecht 1974–1975 ≤ 2 yr  All women born in 1911–1925 | 177 Birth cohort under study 1975–1992 NR | At 1, 1.5, 2, and 4 yr 50–64 yr | 531 Birth cohort under study, same source 3 per case, same birth year, living in Utrecht in 1974, selected from the screening intervention file | Linkage to DOM project breast cancer registry; causes of death provided by general practitioners or hospitals | Screening history for the time up to and including date of diagnosis; 17 yr of follow-up of screening programme; for both cases and controls, participation was low; exclusion of cases with follow-up of < 1 yr | Stratification by birth cohort | 0.54 (0.37–0.79)  
By birth cohort: 1911–1915: 0.40 (0.21–0.75)  
1916–1920: 0.57 (0.31–1.04)  
1921–1925: 0.71 (0.34–1.48) |
<table>
<thead>
<tr>
<th>Reference</th>
<th>Area, year programme began, screening age and interval, women included</th>
<th>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</th>
<th>Screening exposure; age of included women</th>
<th>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</th>
<th>Linkage or use of screening, cancer registry, death databases; data items available</th>
<th>Issues or items related to screening history; whether prevalent cases were excluded</th>
<th>Adjustments</th>
<th>Breast cancer mortality OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Verbeek et al. (1985)</strong></td>
<td>Nijmegen Reference to Verbeek et al. (1984)</td>
<td>62 residents 1975–1982 NR</td>
<td>Diagnosed after first screening invitation; stratification by age at first invitation</td>
<td>310 Birth cohort under study, same source 5 per case, same year of birth as case, and same invitation history</td>
<td>NR</td>
<td>NR</td>
<td>Residential district and marital status</td>
<td>0.51 (0.26–0.99) By age: 35–49 yr: 1.2 (0.31–4.8) 50–64 yr: 0.26 (0.10–0.67) ≥ 65 yr: 0.81 (0.23–2.8)</td>
</tr>
<tr>
<td><strong>Van Dijck et al. (1996)</strong></td>
<td>Nijmegen 1975 35–64 yr (since 1977, also older women) 2-yearly Women invited to participate at age ≥ 65 yr and free of breast cancer at first screening invitation</td>
<td>82 Nijmegen population of invited women, before 1 January 1994 NR</td>
<td>Index round: most recent invitation before diagnosis of primary breast cancer 65–92 yr</td>
<td>410 Age-matched population in Nijmegen, invited to screening at same index round as the case</td>
<td>Cause of death classified by a panel of physicians unaware of the screening history</td>
<td>Patients with advanced breast cancer who died of other, unrelated causes not included as cases</td>
<td>NR</td>
<td>By age: ≥ 65 yr: 0.56 (0.28–1.13) 65–74 yr: 0.45 (0.20–1.02) ≥ 75 yr: 1.05 (0.27–4.14)</td>
</tr>
<tr>
<td>Reference</td>
<td>Area, year programme began, screening age and interval, women included</td>
<td>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</td>
<td>Screening exposure; age of included women</td>
<td>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</td>
<td>Linkage or use of screening, cancer registry, death databases; data items available</td>
<td>Issues or items related to screening history; whether prevalent cases were excluded</td>
<td>Adjustments</td>
<td>Breast cancer mortality OR (95% CI)</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>van Schoor et al. (2010)</td>
<td>Nijmegen 1975 40–69 yr 2-yearly Women invited to the screening 272 Women invited to screening programme in Nijmegen NR 1975–1990 NR</td>
<td>272 Women invited to screening programme in Nijmegen NR 1975–1990 NR 1975–1990 40–69 yr at invitation</td>
<td>1360 Same source Risk sets of controls from which 5 controls were randomly sampled for each case, eligible for screening, and living in Nijmegen at date of death of case</td>
<td>NR</td>
<td>Linkage to vital status from the Municipal Personal Records Database Assessments of causes of death made by a committee of physicians</td>
<td>NR</td>
<td>For differences in age at index invitation between the comparison groups by stratification; thereafter, combination of screening and age as an interaction term to the logistic model Sensitivity analysis for obesity, socioeconomic group, nulliparity, late age at menopause, early age at menarche, and family history</td>
<td>By age:</td>
</tr>
<tr>
<td>Italy</td>
<td>Italy Palli et al. (1989) Florence 1970 40–70 yr Invitation every 30 mo All residents 103 death certificates 1977–1987 After at least a first invitation to the programme and within 3 yr of the last invitation NR</td>
<td>103 death certificates 1977–1987 After at least a first invitation to the programme and within 3 yr of the last invitation NR</td>
<td>515 Same source Selected for year of birth and town of residence 5 per case</td>
<td>Form completed for each woman, with clinical and demographic information</td>
<td>Screening history until date of diagnosis from the Centre for the Study and Prevention of Oncological Diseases</td>
<td>Number of children, age at first birth, civil status, years of education, occupation, place of birth, family history, screening history for cervical cancer, self-referred to breast clinic for mammography</td>
<td>By age at diagnosis:</td>
<td>40–49 yr: 0.63 (0.24–1.6) ≥ 50 yr: 0.51 (0.29–0.89)</td>
</tr>
</tbody>
</table>
Table 5.8 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Area, year programme began, screening age and interval, women included</th>
<th>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</th>
<th>Screening exposure; age of included women</th>
<th>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</th>
<th>Linkage or use of screening, cancer registry, death databases; data items available</th>
<th>Issues or items related to screening history; whether prevalent cases were excluded</th>
<th>Adjustments</th>
<th>Breast cancer mortality OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>California, Massachusetts, Minnesota, Oregon, and Washington</td>
<td>1351 deaths from breast cancer or causes possibly related to breast cancer 1983–1998 1983–1993 100% from 4 of 6 sites, 25% from 1 site, 33% from 1 site</td>
<td>3 yr up to and including the index date: the date of first symptom or suspicion of cancer (in the breast where the cancer was later identified); same date allocated to matched controls 40–49 yr 50–69 yr</td>
<td>2501 Same source as cases Matched on health plan, age, and level of risk for breast cancer, who were alive on the date that the matched case subject had died, and were active health plan members at the time of the matched case subject’s breast cancer diagnosis</td>
<td>Health plan information linked to SEER cancer registries or other cancer registries, and medical chart review</td>
<td>Screening history for 3 yr before index date (mammography and CBE) extracted from medical record review; diagnosis of breast cancer before 1983 was excluded</td>
<td>Race, comorbidity, and age at first birth</td>
<td>By age at screening by CBE or mammography: 40–65 yr: 0.91 (0.78–1.07) 40–49 yr: 0.92 (0.76–1.13) 50–65 yr: 0.87 (0.68–1.12) By age at screening by mammography: 40–65 yr: 0.92 (0.79–1.08) 40–49 yr: 0.85 (0.69–1.05) 50–65 yr: 1.04 (0.82–1.33)</td>
</tr>
<tr>
<td>Elmore et al. (2005)</td>
<td>USA</td>
<td>3 yr up to and including the index date: the date of first symptom or suspicion of cancer (in the breast where the cancer was later identified); same date allocated to matched controls 40–49 yr 50–69 yr</td>
<td>2501 Same source as cases Matched on health plan, age, and level of risk for breast cancer, who were alive on the date that the matched case subject had died, and were active health plan members at the time of the matched case subject’s breast cancer diagnosis</td>
<td>Health plan information linked to SEER cancer registries or other cancer registries, and medical chart review</td>
<td>Screening history for 3 yr before index date (mammography and CBE) extracted from medical record review; diagnosis of breast cancer before 1983 was excluded</td>
<td>Race, comorbidity, and age at first birth</td>
<td>By age at screening by CBE or mammography: 40–65 yr: 0.91 (0.78–1.07) 40–49 yr: 0.92 (0.76–1.13) 50–65 yr: 0.87 (0.68–1.12) By age at screening by mammography: 40–65 yr: 0.92 (0.79–1.08) 40–49 yr: 0.85 (0.69–1.05) 50–65 yr: 1.04 (0.82–1.33)</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Area, year programme began, screening age and interval, women included</td>
<td>No. of breast cancer deaths, source, time period for breast cancer deaths, years of diagnoses; proportion of eligible women included</td>
<td>Screening exposure; age of included women</td>
<td>No. of controls, source, whether same source population as cases, matching variables, alive at date of death or diagnosis of case</td>
<td>Linkage or use of screening, cancer registry, death databases; data items available</td>
<td>Issues or items related to screening history; whether prevalent cases were excluded</td>
<td>Adjustments</td>
<td>Breast cancer mortality OR (95% CI)</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>Norman et al. (2007)</td>
<td>CARE multicentre study NR 40–64 yr White women and Black women in metropolitan Atlanta, Georgia; Detroit, Michigan; Los Angeles, California; Philadelphia, Pennsylvania; and Seattle, Washington 553 Women with a new diagnosis of invasive breast cancer in 1994–1998 who died NR 1994–1998 NR</td>
<td>At least 1 screening mammogram in the 2 yr before the reference date (month and year of initial diagnosis for cases) 40–64 yr 4016 Women identified by random-digit dialling who had never been diagnosed with cancer</td>
<td>Standard SEER follow-up procedures used, primarily passive linkage with state death records; for the Pennsylvania site, state death records used</td>
<td>Screening histories from population screening registries or medical records</td>
<td>BMI, family history, education, marital status, parity, alcohol consumption in year before reference date, smoking status, number of pre-existing medical conditions, use of oral contraceptive, use of combined estrogen–progestin hormone replacement therapy, use of estrogen therapy, and less than twice the federal poverty threshold for household income. Model with stratification by age was further adjusted for menopausal status</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BMI, body mass index; CARE, Contraceptive and Reproductive Experiences; CBE, clinical breast examination; CI, confidence interval; mo, month or months; NR, not reported; OR, odds ratio; RR, relative risk; SEER, Surveillance, Epidemiology, and End Results; yr, year or years.
small numbers, and no confidence intervals were given.]

An updated case–control analysis 17 years after the initiation of this project was reported by Miltenburg et al. (1998). Controls (three for each case) were defined as women with the same year of birth as the case, living in the city of Utrecht at the time the case died, and having had the opportunity to be screened in the DOM project. The odds ratio for breast cancer mortality for screening in the period 1975–1992 was 0.54 (95% CI, 0.37–0.79). Stratification by birth cohort is given in Table 5.8.

In 1975, a population-based screening programme was set up in Nijmegen, a city with about 150 000 inhabitants (Peeters et al., 1989a). The first screening round, in 1975–1976, involved 23 000 women born in 1910–1939, who were thus aged 35–64 years. In the subsequent screening rounds, the same birth cohort was invited, as well as 7700 women born before 1910. The odds ratio for death from breast cancer estimated in a case–control analysis was 1.2 (95% CI, 0.31–4.8) for women aged 35–49 years, 0.26 (95% CI, 0.10–0.67) for those aged 50–64 years, and 0.81 (95% CI, 0.23–2.8) for those aged 65 years and older (Verbeek et al., 1985).

In a further case–control study based on the Nijmegen population, Van Dijck et al. (1996) selected women who were 65 years or older when first invited to screening. The rate ratio of breast cancer mortality in women who had participated regularly (i.e. in the two most recent screening rounds before diagnosis) compared with those who had not participated in screening was 0.56 (95% CI, 0.28–1.13). The rate ratio for women aged 65–74 years at the most recent invitation was 0.45 (95% CI, 0.20–1.02), and for women aged 75 years and older it was 1.05 (95% CI, 0.27–4.14). [The Working Group estimated rate ratios for women who had ever been screened by combining, using fixed effects meta-analysis, reported relative risks for women who had been screened regularly and women who had been screened “otherwise” relative to women who had not been screened. The estimates rate ratios were 0.68 (95% CI, 0.44–1.05) for all ages, 0.54 (95% CI, 0.31–0.95) for ages 65–74 years, and 0.94 (95% CI, 0.45–1.88) for ages 75 years and older. Forty of the 82 deaths from breast cancer included in this study were included in a separate IBM analysis of effectiveness of screening in women aged 68–83 years at entry into the Nijmegen screening programme (Van Dijck et al., 1997).]

van Schoor et al. (2010) designed a case–control study to investigate the effect of biennial mammography screening on breast cancer mortality in women aged 40–69 years between 1975 and 1990 in Nijmegen. In women aged 40–49 years at their index screening (in cases, the last screening before diagnosis of breast cancer), the odds ratio for screening was 0.50 (95% CI, 0.30–0.82). Similarly, an odds ratio of 0.54 (95% CI, 0.35–0.85) was reported for women aged 50–59 years, and an odds ratio of 0.65 (95% CI, 0.38–1.13) for those aged 60–69 years.

(ii) Italy

Between 1970 and 1980, women aged 40–70 years living in 24 municipalities in Florence were invited to mammography screening with craniocaudal and mediolateral oblique views every 2.5 years. In 1989, the screening area was extended to include the city of Florence. Palli et al. (1986, 1989) conducted a case–control study within this population to estimate the impact on breast cancer mortality. The odds ratios for women aged 40–49 years and for those aged 50 years and older at diagnosis of breast cancer were estimated to be 0.63 (95% CI, 0.24–1.6) and 0.51 (95% CI, 0.29–0.89), respectively.

(iii) USA

Elmore et al. (2005) conducted a matched case–control study among women enrolled in six health plans in the states of California, Massachusetts, Minnesota, Oregon, and Washington and examined the efficacy of
Breast cancer screening by mammography and/or CBE among women in two age cohorts (40–49 years and 50–65 years) and in two levels of breast cancer risk (in women at average risk and women with a family history and/or previous breast biopsy) until 1983–1998. The effect of screening with mammography, or of screening with mammography and CBE, during the 3 years before the index date (defined as the date of first suspicion of breast abnormalities in case subjects, with the same date used for matched control subjects) was evaluated. For women aged 40–49 years at diagnosis of breast cancer, the odds ratio was 0.85 (95% CI, 0.65–1.23), and for women aged 50–65 years, the odds ratio was 0.47 (95% CI, 0.35–0.63) for screening with mammography alone. The odds ratio for women at an increased risk was 0.74 (95% CI, 0.50–1.03) and for women at average risk was 0.96 (95% CI, 0.80–1.14); however, the difference was not statistically significant ($P = 0.17$).

Norman et al. (2007) used data from a subset of the Women’s Contraceptive and Reproductive Experiences (CARE) Study, a population-based multicentre case–control study of risk factors for breast cancer among White and Black women conducted in metropolitan Atlanta, Georgia; Detroit, Michigan; Los Angeles, California; Philadelphia, Pennsylvania; and Seattle, Washington, to estimate the relative mortality rates from invasive breast cancer among women with at least one screening mammogram in the 2 years before a baseline reference date compared with unscreened women, adjusting for potential confounding. The odds ratio for breast cancer death within 5 years after diagnosis was 0.89 (95% CI, 0.65–1.23) for ages 40–49 years at diagnosis and 0.47 (95% CI, 0.35–0.63) for ages 50–64 years at diagnosis.

A meta-analysis was performed of some of the earlier case–control studies (Demissie et al., 1998), and Broeders et al. (2012) conducted a meta-analysis of seven more recent case–control studies. The combined unadjusted odds ratio in women who were screened versus those who were not screened was 0.46 (95% CI, 0.40–0.54), and 0.52 (95% CI, 0.42–0.65) when adjusted for self-selection using the method of Duffy et al. (2002a). The crude odds ratio for breast cancer mortality reduction, translated to intention-to-treat estimates for women who were invited versus those who were not invited was 0.69 (95% CI, 0.57–0.83).

(c) Specific age groups

Several of the case–control studies summarized above reported results in several age groups, including those that lie below or above the age range 50–69 years. Such results can be validly used to infer the effectiveness, or otherwise, of screening women younger than 50 years, provided they are based only on deaths from breast cancer of women whose breast cancer was diagnosed when they were younger than 50 years. The results that permit this inference are those of Palli et al. (1989), Broeders et al. (2002), Elmore et al. (2005), Norman et al. (2007), and Roder et al. (2008) (see Table 5.8).

The use of results from case–control studies to infer effectiveness at ages older than 69 years is less straightforward because, even if they are based only on deaths from breast cancer of women whose breast cancer was diagnosed when they were older than 69 years, the relative risk of death calculated will have been influenced by screening at age 69 years and younger, assuming screening effectiveness (Otto et al., 2012b). This influence can only be removed by limiting the analysis to women first offered screening after age 69 years. No case–control study has been done in a context in which this limitation could be applied; however, that of Van Dijck et al. (1996) was limited to women first offered screening from age 65 years.
5.2.3 Ecological studies

In assessing the effectiveness of breast cancer screening, the Working Group considered that accurate information on standards of breast cancer treatment in different regions analysed and careful matching of regions by treatment standards or adjustment for differences between regions in treatment standards are minimum criteria for validity of ecological studies. Therefore, simple comparisons of trends between unmatched regions or without potentially effective statistical adjustment, or in a single region over time, were excluded.

Correcting for differences in underlying incidence is a challenge. Differences in incidence between regions, or across time, may indicate an important difference in baseline risk that must be adjusted for, or they may indicate overdiagnosis and should not be adjusted for. These studies were therefore excluded, as were any that measured differences in survival, due to the well-recognized issue of lead time.

Studies of population-based screening in Europe were reviewed to assess the value of trend analyses in population breast cancer mortality (Moss et al., 2012). A literature review identified 17 reports, of which 12 provided quantitative estimates of the impact of screening. Due to differences in comparisons and outcome measures, no pooled estimate of effectiveness was calculated. Overall, this approach proved to be of limited value for assessment of screening impact.

For the purpose of selecting studies to review, the Working Group defined the following subcategories:

Category 1: Single-country or single-region studies that consider time trends in total incidence or total mortality, or that use, at best, different age groups to standardize treatment effects. These studies were excluded because of the impossibility of disentangling temporal changes in incidence, overdiagnosis, lead-time effect, and changes in treatment.

Category 2: Studies that measure proportional distribution of breast cancers by stage, proportional or relative survival, or post-diagnosis survival time over time or between countries with different screening protocols. These studies were excluded because of the potential bias due to overdiagnosis or the clear bias due to earlier diagnosis in screened women (lead-time bias).

Category 3: Studies of incidence of advanced-stage breast cancers over time between matched regions. These studies were included, subject to appropriate care having been taken to match or otherwise account for differences in risk factors or treatment. It is also necessary to account for differing completeness or reliability of staging. The advantage of such studies is that they should minimize the effects of overdiagnosis (which would generate mostly early-stage cancers) and differences in treatment. Correction is still required for a changing underlying rate of breast cancer incidence. This correction is generally based on the assumption that this change is driven by lifestyle changes, which change progressively, and in a similar manner in matched regions. Hence, smooth temporal trends are used to model the underlying rate, whereas effects of screening should manifest both by more rapid changes and by contrasts between regions that introduced screening on different dates.

Category 4: Studies of breast cancer mortality over time in matched regions. These studies raise the same issues as those of advanced-stage breast cancers, with the further complication of potential or real differences in treatment. This may include the availability of systemic hormone treatments or the organization of health-care systems.

A total of 87 studies were identified by the Working Group through literature searches and were reviewed for initial categorization according
to the above criteria. After the initial exclusion of studies in categories 1 \((n = 25)\) or 2 \((n = 20)\), studies of other designs (9 case–control studies, 4 cohort studies, and 3 studies based on RCTs), and studies with other limitations \((n = 12)\), 14 studies were further considered. Eight of these were then identified as IBM studies (Tabár et al., 2001; Duffy et al., 2002b; Jonsson et al., 2003a, b; Parvinen et al., 2006; Anttila et al., 2008; Sarkeala et al., 2008b; Kalager et al., 2010) and were therefore excluded. Of the remaining six ecological studies, two were judged to be uninformative: Das et al. (2005) used correlation as the measure of association, and Autier et al. (2011) may have been biased by the evolution of staging data over the study period; the remaining four studies were found to be informative. One additional informative study was identified separately (Otto et al., 2003) and was included in the review.

Otto et al. (2003) reviewed mortality trends in the Netherlands from 1980 to 1998, using clustered municipality-level data in 1-month bands, including the progressive introduction of screening from 1989 until 1997. Four age bands were compared to detect changes in treatment effectiveness: 45–54, 55–64, 65–74, and 75–84 years. Rates of change and cumulative changes were estimated in both the pre-screening and screening eras. Analysis was via linear splines (i.e. a single joinpoint). There was a downturn in mortality for the middle two age bands (55–64 years and 65–74 years) coincident with the introduction of screening, with an accumulated mortality reduction by 1999 estimated to be 19.1%. The annual rate of decline (annual percentage change) was 1.7% (95% CI, 1–2.4%) in these two age groups combined and 1.2% (95% CI, 0.1–2.4%) in the younger age group (45–54 years). There was no significant change in the older age group (75–84 years). Before screening, the trend was upward at 0.3% per year.

Törnberg et al. (2006) compared time trends in breast cancer incidence and mortality after the introduction of mammography screening in Copenhagen, Helsinki, Stockholm, and Oslo. In Helsinki, screening was offered to women aged 50–59 years, starting in 1986, and in the other three capitals, screening was offered to women aged 50–69 years, starting between 1989 and 1996. Peaks in breast cancer incidence depended on the age groups covered by the screening, the length of the implementation of screening, and the extent of background opportunistic screening. No mortality reduction after the introduction of screening was visible after 7–12 years of screening in any of the capitals. [No visible effect on mortality reduction was expected in Oslo, due to too short an observation period.]

Jørgensen et al. (2010) compared breast cancer mortality trends in Denmark, between Copenhagen (where screening was introduced in 1991) and Funen County (where screening started in 1993) and the rest of Denmark (which served as an unscreened control group). Unscreened age groups were used to further control for effects of changing treatment. Screening was offered to women aged 55–74 years, and mortality was evaluated in three age bands: 35–54, 55–74, and 75–84 years. The pre-screening period was 1982–1991, and the post-screening period was restricted to 1997–2006, to allow for a lag in benefit. The annual percentage change in breast cancer mortality was evaluated by Poisson regression. For the likely-to-benefit age band (55–74 years), the annual percentage change changed from +1 to −1% in the screening areas and from +2 to −2% in the non-screening areas. For the younger age band (35–54 years), the annual percentage change changed from +2% to −5% in the screening areas and from 0% to −6% in the non-screening areas. No significant changes were observed in the older age band.

The mortality benefit of attending screening was estimated using a Markov model of disease progression based on three regions in France (Uhry et al., 2011). Attempts were made to correct for opportunistic screening, and overdiagnosis was included as an explicit assumption, at either
10% or 20%. The corresponding estimates of mortality reduction were 23% (95% CI, 4% to 38%) and 19% (95% CI, −3% to 35%). [Problems of model fit were reported.]

Poisson regression was used in a study reanalyzing population data from the era of Swedish screening trials (Haukka et al., 2011). [The data used were from NORDCAN (Engholm et al., 2010), which had variable levels of agreement with trial data where it could be compared.] The model assumed a delayed step change due to screening after the staggered introduction by region, with different lead times tested for best fit. Using the 3-year lead time estimate, breast cancer mortality decreased by 16% (RR, 0.84; 95% CI, 0.78–0.91) in the screening age group 40–69 years and by 11% (RR, 0.89; 95% CI, 0.80–0.98) in the age group 70–79 years.

5.2.4 Other measures of screening performance

See Table 5.9.

(a) Studies reporting on tumour size and nodal status in women aged 50–69 years

Hofvind et al. (2012c) compared incidence of advanced breast cancer cases diagnosed among screened and unscreened women aged 50–69 years in Norway. A total of 11 569 breast tumours (1670 ductal carcinoma in situ [DCIS] and 9899 invasive cancer) were diagnosed among 640 347 women who were invited to the screening programme during the study period. Participants in the screening programme accounted for 9726 breast tumours (1517 DCIS and 8209 invasive cancer) and non-participants accounted for 1843 breast tumours (153 DCIS and 1690 invasive cancer). When cases were compared between participants and non-participants, a significant reduction was observed in stage III (RR, 0.5; 95% CI, 0.4–0.7) and stage IV (RR, 0.3; 95% CI, 0.2–0.4) cancers, in tumours larger than 50 mm (RR, 0.4; 95% CI, 0.4–0.6), and in distant metastasis (RR, 0.3; 95% CI, 0.2–0.4). Distributions by stage, size, and nodal status were similar in women who did not attend screening and those who were not invited.

Domigo et al. (2013b) analysed data on invitation to organized screening programmes in Copenhagen (first eight invitations rounds, 1991–2008) and in Funen (first six invitation rounds, 1993–2005) (Table 5.10). Both programmes offered biennial screening to women aged 50–69 years. The Working Group calculated the rate ratios and 95% confidence intervals for tumour size and nodal status of screen-detected breast cancers versus those diagnosed in women who were not screened, for Copenhagen and Funen together. Among screen-detected cancers, a significant increase in detection of tumours of size 0–10 mm [RR, 2.91; 95% CI, 2.47–3.44] and 11–20 mm [RR, 1.27; 95% CI, 1.14–1.41] and a reduction in detection of tumours of size 21–30 mm [RR, 0.47; 95% CI, 0.40–0.55] and larger than 30 mm [RR, 0.26; 95% CI, 0.21–0.33] and in node–positive cancers [RR, 0.61; 95% CI, 0.54–0.67] were estimated. The rates of large screen-detected cancers were significantly lower, and screen-detected cancers were significantly less frequently lymph node-positive.

(b) Studies reporting incidence rates since the beginning of the screening period

Foca et al. (2013) analysed data from 700 municipalities in Italy, with a total population of 692 824 women aged 55–74 years targeted by organized mammography screening from 1991 to 2005. The effect of the screening was evaluated from year 1 (the year screening started at the municipal level) to year 8 (based on the decreasing number of available municipalities). The study was based on a total of 14 447 incident breast cancers. The observed 2-year, age-standardized (Europe) incidence rate ratio (ratio of the incidence rate to the expected rate) was calculated. Expected rates were estimated assuming that the incidence of breast cancer was stable and
<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, age, interval</th>
<th>Duration of screening</th>
<th>Compared groups: contemporary or historical, period(s) covered, nature of groups</th>
<th>Denominators for rate/proportions calculations</th>
<th>Period of observation for screened and not screened</th>
<th>Adjustments</th>
<th>RR (95% CI) unless otherwise stated</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hofvind et al. (2012c)</td>
<td>Norway 1996 50–69 yr 2 yr</td>
<td>1–12 yr Individual</td>
<td>Contemporary 1996–2007 Invited and screened, invited but not screened</td>
<td>Invitations to screening Screened: 1 475 978 (9726) Not screened: 449 747 (1843)</td>
<td>2 yr after each invitation to screening</td>
<td>None</td>
<td>Stage:&lt;sup&gt;h&lt;/sup&gt; 0: 3.0 (2.6–3.6) I: 2.0 (1.8–2.2) II: 1.2 (1.1–1.3) III: 0.5 (0.4–0.7) IV: 0.3 (0.2–0.4) Tumour size: &gt; 50 mm: 0.4 (0.4–0.6) Node-positive: No: 2.0 (1.8–2.1) Yes: 1.1 (1.0–1.2) Distant metastasis: No: 1.8 (1.7–1.9) Yes: 0.3 (0.2–0.4)</td>
<td>Distributions by stage, size, and nodal status were similar between not attending and not invited women</td>
</tr>
<tr>
<td>Domingo et al. (2013b)</td>
<td>Denmark (Copenhagen and Funen) 50–69 yr 2 yr</td>
<td>Copenhagen: 8 biennial screening rounds Funen: 6 biennial screening rounds Individual</td>
<td>Same years of observation</td>
<td>Copenhagen: Participants: 214 088 Not screened: 139 461 Funen: Participants: 486 722 Not screened: 230 153</td>
<td>Copenhagen: 1991–2008 Funen: 1993–2005</td>
<td>Rate ratios: Tumour size: ≤ 10 mm: [2.91 (2.47–3.44)] 11–20 mm: [1.27 (1.14–1.41)] 21–30 mm: [0.47 (0.40–0.55)] &gt; 30 mm: [0.26 (0.21–0.33)] Node-positive: No: [1.61 (1.47–1.77)] Yes: [0.61 (0.54–0.67)] See Table 5.10 for original data</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Areas, earliest year of programme screening, age, interval</td>
<td>Duration of screening Individual or aggregate data</td>
<td>Compared groups: contemporary or historical, period(s) covered, nature of groups</td>
<td>Denominators for rate/proportions calculations</td>
<td>Period of observation for screened and not screened</td>
<td>Adjustments</td>
<td>RR (95% CI) unless otherwise stated</td>
<td>Comments</td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------</td>
<td>-----------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Foca et al. (2013)</td>
<td>Italy (700 municipalities) 1991–2005 55–74 yr</td>
<td>1991–2005 Individual</td>
<td>(Analysis from year 1 to year 8) year 1: 692 824 women year 8: 300 859 women Total number of eligible cancer cases: 14 447 Advanced cancers analysed: 4036 (28%) pT2–pT4 cancers</td>
<td>Study endpoints: total incidence of breast cancer incidence of pT2–pT4 breast cancer</td>
<td>1991–2005 (analysis from year 1 to year 8)</td>
<td>1–2 yr after introduction of screening: Total breast cancer: 1.35 (1.03–1.41) pT2–pT4: 0.97 (0.90–1.04) 5–6 yr after introduction of screening: Total breast cancer: 1.14 (1.08–1.20) pT2–pT4: 0.79 (0.73–0.87) 7–8 yr after the introduction of screening: Total breast cancer: 1.14 (1.08–1.21) pT2–pT4: 0.71 (0.64–0.79)</td>
<td>Excluded women aged 50–54 yr Restricted to municipalities in which the proportion of total incident cancers detected by screening reached 30% within year 2 Annual incidence expected in the absence of screening assumed stable and equivalent to that observed in the past 3 yr before year 1 Effect evaluated based on the decreasing number of available municipalities Supplementary analysis of the subgroup of municipalities that had a complete 8-yr period of observation</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Areas, earliest year of programme screening, age, interval</td>
<td>Duration of screening Individual or aggregate data</td>
<td>Compared groups: contemporary or historical, period(s) covered, nature of groups</td>
<td>Denominators for rate/ proportions calculations</td>
<td>Period of observation for screened and not screened</td>
<td>Adjustments</td>
<td>RR (95% CI) unless otherwise stated</td>
<td>Comments</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------------------------------------------</td>
<td>--------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------</td>
<td>-----------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Nederend et al. (2012)</td>
<td>Netherlands 50–75 yr 2 yr Women aged 50–69 yr (75 yr in 1998)</td>
<td>1997–2008</td>
<td>351 009 consecutive screens of 85 274 women</td>
<td>Age, family history of breast cancer, previous breast surgery, use of HRT, initial screen, interval between 2 latest screens, breast density at latest screening mammogram, mammographic abnormality, tumour histology of invasive cancers</td>
<td>Rate per 1000 (95% CI)</td>
<td>Age, family history of breast cancer, previous breast surgery, use of HRT, initial screen, interval between 2 latest screens, breast density at latest screening mammogram, mammographic abnormality, tumour histology of invasive cancers</td>
<td>Advanced cancers: 1997–1998: 1.5 (1.2–1.9) 1999–2000: 1.6 (1.3–2.0) 2001–2002: 1.6 (1.3–2.0) 2003–2004: 1.6 (1.3–1.9) 2005–2006: 1.5 (1.2–1.8) 2007–2008: 1.9 (1.5–2.2) Total: 1.6 (1.5–1.8) Non-advanced cancers: 1997–1998: 3.0 (2.5–3.5) 1999–2000: 3.3 (2.8–3.8) 2001–2002: 3.0 (2.5–3.5) 2003–2004: 3.9 (3.4–4.4) 2005–2006: 3.3 (2.9–3.7) 2007–2008: 3.3 (2.9–3.7) Total: 3.3 (3.1–3.5)</td>
<td>At multivariate analysis, women with a ≥ 30-mo interval between the latest two screens had an increased risk of screen-detected advanced breast cancer (OR, 1.63; 95% CI, 1.07–2.48)</td>
</tr>
</tbody>
</table>
### Table 5.9  (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, age, interval</th>
<th>Duration of screening Individual or aggregate data</th>
<th>Compared groups: contemporary or historical, period(s) covered, nature of groups</th>
<th>Denominators for rate/proportions calculations</th>
<th>Period of observation for screened and not screened</th>
<th>Adjustments</th>
<th>RR (95% CI) unless otherwise stated</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autier &amp; Boniol (2012)</td>
<td>West Midlands, United Kingdom 1988 50–64 yr</td>
<td>1988–2004 Aggregate</td>
<td>No comparison APC of the incidence rates of lymph node-positive/ negative and of tumours &gt; 50 mm reported for the screening period</td>
<td>First procedure based on CI5plus (Ferlay et al., 2014) and on proportions derived from Nagtegaal et al. (2011), for distinguishing cancers found in women attending and not attending screening</td>
<td>Data reported for the screening period 1989–2004 only</td>
<td>APC</td>
<td>See Fig. 5.1</td>
<td>The &gt; 50 mm cut-off is not appropriate to study changes in incidence rates of advanced cancers in a country with a high level of awareness, as United Kingdom Sources for estimation of incidence trends of advanced breast cancer NR</td>
</tr>
<tr>
<td>Reference</td>
<td>Areas, earliest year of programme screening, age, interval</td>
<td>Duration of screening</td>
<td>Compared groups: contemporary or historical, period(s) covered, nature of groups</td>
<td>Denominators for rate/proportions calculations</td>
<td>Period of observation for screened and not screened</td>
<td>Adjustments</td>
<td>RR (95% CI) unless otherwise stated</td>
<td>Comments</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------------------------------------------------</td>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------</td>
<td>-----------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Elting et al. (2009)</td>
<td>Texas, USA &gt; 40 yr</td>
<td>2002–2004 Individual</td>
<td>Incident breast cancer cases diagnosed among women aged &gt; 40 yr in 2004 Total of 12,469 women</td>
<td>Risk of invasive breast cancer and DCIS in Texas Counties with facility compared with counties without facilities</td>
<td>2004 Age, race, ethnicity, higher probabilities of advanced disease among African-American and Hispanic women</td>
<td>Stage at diagnosis: DCIS: 1.27 (1.07–1.5) Regional nodes: 1.12 (0.98–1.27) Locally advanced or distant disease: 0.81 (0.66–0.98) Factors associated with diagnosis of DCIS compared with local disease: In-county facility 1.32 (0.98–1.77) Factors associated with diagnosis of locally advanced or disseminated disease compared with local disease: In-county facility 0.36 (0.26–0.51) ( P &lt; 0.001 )</td>
<td>Significant associations between the absence of in-county mammography facilities and both low odds of screening and high odds of diagnosis at a late stage of breast cancer</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.9 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Areas, earliest year of programme screening, age, interval</th>
<th>Duration of screening Individual or aggregate data</th>
<th>Compared groups: contemporary or historical, period(s) covered, nature of groups</th>
<th>Denominators for rate/proportions calculations</th>
<th>Period of observation for screened and not screened</th>
<th>Adjustments</th>
<th>RR (95% CI) unless otherwise stated</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Helvie et al. (2014)</strong></td>
<td>USA, 18 SEER geographical areas, which captured cancer data from 27.8% of the United States population &gt; 40 yr</td>
<td>2007–2009</td>
<td>Trend</td>
<td>Before mammography (1977–1979) Mammography screening period (2007–2009) Underlying temporal trends</td>
<td>Late-stage breast cancer incidence decreased by 37%, with a reciprocal increase in early-stage rates Late-stage breast cancer incidence decreased by from 21% to 48% Total invasive breast cancer incidence decreased by 9%</td>
<td>Projected incidence stage-specific values were compared with actual observed values in 2007–2009. Used different APC estimates</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Areas, earliest year of programme screening, age, interval</td>
<td>Duration of screening</td>
<td>Compared groups: contemporary or historical, period(s) covered, nature of groups</td>
<td>Denominators for rate/proportions calculations</td>
<td>Period of observation for screened and not screened</td>
<td>Adjustments</td>
<td>RR (95% CI) unless otherwise stated</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------</td>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------</td>
<td>-------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>Hou &amp; Huo (2013)</td>
<td>USA, 18 SEER registries No data on screening</td>
<td>2000–2009</td>
<td>Trend Breast cancer incidence rates from 2000 to 2009</td>
<td>Incidence rates of in situ, localized, regional, distant (per 100,000)</td>
<td>None</td>
<td>DCIS (all racial groups): APC, 2.3–3.0% ($P &lt; 0.005$)</td>
<td>Localized breast cancer: non-Hispanic Black women: APC, 1.3% ($P = 0.004$) Asian women: APC, 1.2% ($P = 0.03$) Regional and distant cancers: non-Hispanic White women: APC, −2.5% ($P = 0.02$) Hispanic women: APC, −1.1% ($P = 0.006$)</td>
<td>It is unlikely that the overall trends of incidence rates are due to changes in mammography screening rate, since mammography use did not change substantially from 2000 to 2010</td>
</tr>
<tr>
<td>Reference</td>
<td>Areas, earliest year of programme screening, age, interval</td>
<td>Duration of screening Individual or aggregate data</td>
<td>Compared groups: contemporary or historical, period(s) covered, nature of groups</td>
<td>Denominators for rate/proportions calculations</td>
<td>Period of observation for screened and not screened</td>
<td>Adjustments</td>
<td>RR (95% CI) unless otherwise stated&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>-----------------------------------------------------------</td>
<td>---------------------------------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DeSantis et al. (2014)</td>
<td>USA, SEER Program, SEER 9 registries</td>
<td>NR Aggregate</td>
<td>Historical</td>
<td>Data about incidence, probabilities of developing cancer, and cause-specific survival obtained from the SEER Program</td>
<td>Prevalence data on mammography by age and state obtained from the 2010 and 2012 Behavioral Risk Factor Surveillance System</td>
<td>1975–2010</td>
<td>Rates age-adjusted to the 2000 United States standard population within each age group</td>
<td>Correlation between mammography screening prevalence in 2010 and breast cancer stage at diagnosis (2006–2010): Non-Hispanic White women: in situ stage, ( r = 0.62 \quad (P &lt; 0.001) ) late stage, ( r = -0.51 \quad (P &lt; 0.001) ) African-American women: in situ stage, ( r = 0.47 \quad (P &lt; 0.006) ) late stage, NS</td>
</tr>
</tbody>
</table>

<sup>a</sup> Comparing screened and unscreened.

<sup>b</sup> Calculated using COMPARE2 in WinPepi V11.39 (http://www.brixtonhealth.com/pepi4windows.html).

<sup>c</sup> Calculated using Stata/SE 13.1.

APC, annual percentage change; CI, confidence interval; DCIS, ductal carcinoma in situ; HRT, hormone replacement therapy; IRR, incidence rate ratio; NR, not reported; NS, not significant; OR, odds ratio; RR, relative risk; SEER, Surveillance, Epidemiology, and End Results; yr, year or years.
## Table 5.10 Number of breast cancers (invasive and carcinoma in situ) detected at screening in participants, diagnosed as interval cancers in participants, or diagnosed in unscreened women (Copenhagen and Funen screening programmes, Denmark)

<table>
<thead>
<tr>
<th>Invitation round</th>
<th>Screened women</th>
<th>Unscreened women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Participants</td>
<td>Screen-detected cancers (of which CIS)</td>
</tr>
<tr>
<td>Copenhagen screening programme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>30 388</td>
<td>361 (44)</td>
</tr>
<tr>
<td>2</td>
<td>26 109</td>
<td>164 (17)</td>
</tr>
<tr>
<td>3</td>
<td>25 153</td>
<td>156 (18)</td>
</tr>
<tr>
<td>4</td>
<td>25 427</td>
<td>147 (18)</td>
</tr>
<tr>
<td>5</td>
<td>25 059</td>
<td>145 (22)</td>
</tr>
<tr>
<td>6</td>
<td>25 271</td>
<td>180 (42)</td>
</tr>
<tr>
<td>7</td>
<td>26 205</td>
<td>227 (40)</td>
</tr>
<tr>
<td>8</td>
<td>30 476</td>
<td>242 (47)</td>
</tr>
<tr>
<td>Total (1–8)</td>
<td>214 088</td>
<td>1622 (248)</td>
</tr>
<tr>
<td>Funen screening programme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>41 519</td>
<td>401 (59)</td>
</tr>
<tr>
<td>2</td>
<td>44 117</td>
<td>236 (35)</td>
</tr>
<tr>
<td>3</td>
<td>44 892</td>
<td>216 (21)</td>
</tr>
<tr>
<td>4</td>
<td>45 817</td>
<td>273 (35)</td>
</tr>
<tr>
<td>5</td>
<td>47 458</td>
<td>257 (19)</td>
</tr>
<tr>
<td>6</td>
<td>48 831</td>
<td>285 (31)</td>
</tr>
<tr>
<td>Total (1–6)</td>
<td>272 634</td>
<td>1668 (200)</td>
</tr>
<tr>
<td>Total</td>
<td>486 722</td>
<td>3290 (448)</td>
</tr>
</tbody>
</table>

\(^a\) Proportion per 1000 women, and rate per 1000 person–years.

\(^b\) Person–years at risk to develop a screen-detected cancer were estimated as number of participants multiplied by length of invitation round.

\(^c\) Person–years at risk to develop an interval cancer were estimated as number of participants, minus participants with screen-detected cancers, multiplied by 2.

\(^d\) Person–years at risk to develop a cancer outside screening were estimated as number of unscreened women multiplied by length of invitation round.

\(^e\) For simplicity, for each invitation round based on the total of screen-detected cancers, interval cancers, and cancers in unscreened women, although part of the interval cancers were diagnosed during the next invitation round.

CIS, carcinoma in situ.

equivalent to that in the 3 years before year 1. The incidence rate ratio for pT2–pT4 breast cancers was 0.97 (95% CI, 0.90–1.04) in years 1 and 2, 0.81 (95% CI, 0.75–0.88) in years 3 and 4, 0.79 (95% CI, 0.73–0.87) in years 5 and 6, and 0.71 (95% CI, 0.64–0.79) in years 7 and 8. A significant and stable decrease in the incidence of late-stage breast cancer was observed from the third year of screening onward.

Nederend et al. (2012) analysed a consecutive series of 351 009 screening mammograms of 85 274 women aged 50–75 years, who underwent biennial screening in a breast screening region in the Netherlands in 1997–2008. A total of 1771 screen-detected cancers and 669 interval cancers were diagnosed in 2440 women. The authors observed, as expected, no decline in detection rates of advanced breast cancer during each round of 12 years of biennial screening mammography in the screened population.

In the source population (data from a cancer registry), no decline in advanced breast cancer has been reported.

Autier & Boniol (2012) estimated incidence trends in advanced breast cancer from 1989 to 2004 in the West Midlands (United Kingdom), where breast screening started in 1988 for women aged 50–64 years (Fig. 5.1). The authors extracted numbers of breast cancer cases from the Cancer Incidence in Five Continents database (Ferlay et al., 2014). They used published data (Lawrence et al., 2009; Nagtegaal et al., 2011) for the annual percentage change (APC) in the incidence rates of lymph node-positive/node-negative breast cancer and of tumours larger than 50 mm for the screening period. According to their analysis, the incidence rates of node-positive breast cancer increased from 1989 to 1992 and then decreased below the pre-screening level in 1993–1995 but returned to pre-screening levels in 1996–2000 and then stabilized. From 1989 to 2004, the APC...
Breast cancer screening

was 2.2% (95% CI, 1.2% to 3.1%) for node-negative cancers and −0.7% (95% CI, −1.8% to 0.3%) for node-positive cancers. The incidence of tumours larger than 50 mm remained stable from 1989 to 2004 (APC, 0.2%; 95% CI, −2.2% to 2.7%).

Eisemann et al. (2013) reported data from 2008–2009 in Germany, where breast cancer screening started in 2005, biennially, for women aged 50–69 years. From 2002 to 2007, the absolute number of breast cancer diagnoses (including in situ cases) increased markedly, by 15%: for in situ tumours, by +94%; for T1 tumours, by +18%; for T2 tumours, by +11%; for T3 tumours, by +14%; and for tumours of unknown stage, by +24%. A decrease of about −10% was observed for T4 tumours. [No comparison of rates of advanced cancers was reported in the screened or invited population versus the population not screened or not invited.]

Elting et al. (2009) assessed the association between in-county mammography facilities (in 2002–2004) and mammography screening and breast cancer diagnosis at a late stage among women in Texas older than 40 years. Half of the 254 counties had no mammography facility. In 2004, a total of 12 469 of the 4 639 842 women in Texas older than 40 years were diagnosed with either invasive breast cancer or DCIS (risk per 10 000 women aged > 40 years, 26.87; 95% CI, 26.4–27.3). The risk of diagnosis at early and late stages varied significantly between counties with and without mammography facilities. After accounting for confounding by age, race, and ethnicity, multivariate analysis showed that women who lived in counties with facilities were more likely to be diagnosed with DCIS (odds ratio [OR], 1.32; 95% CI, 0.98–1.77; P = 0.06) and significantly less likely to be diagnosed at an advanced stage (OR, 0.36; 95% CI, 0.26–0.51; P < 0.001) than their counterparts who lived in counties without a facility. These differences were observed despite adjustment for higher probabilities of advanced disease among African-American and Hispanic women.

(c) Studies reporting incidence rates using SEER data

Bleyer & Welch (2012) used data from the Surveillance, Epidemiology, and End Results (SEER) Program of the United States National Cancer Institute to examine trends from 1976 to 2008 in the incidence of early-stage and late-stage breast cancer among women aged 40 years and older. The 3-year period 1976–1978 was chosen to obtain the estimate of the baseline incidence of breast cancer detected without mammography. During this period, the incidence of breast cancer was stable and few cases of DCIS were detected (findings compatible with the very limited use of screening mammography). The estimate of the current incidence of breast cancer was based on the 3-year period 2006–2008. To eliminate the effect of use of hormone replacement therapy, the observed incidence was truncated if it was higher than the estimate of the current incidence (the annual incidence per 100 000 women of DCIS was not allowed to exceed 56.5 cases, of localized disease to exceed 177.5 cases, of regional disease to exceed 77.6 cases, and of distant disease to exceed 16.6 cases, during the period 1990–2005). A substantial increase in the use of screening mammography during the 1980s and early 1990s among women aged 40 years and older in the USA, a substantial concomitant increase in the incidence of early-stage breast cancer among these women, and a small decrease in the incidence of late-stage breast cancer were observed. A large increase in cases of early-stage cancer (absolute increase of 122 cases per 100 000 women) and a small decrease in cases of late-stage cancer (absolute decrease of 8 cases per 100 000 women) were observed. The trends in regional and distant late-stage breast cancer showed that the variable pattern in late-stage cancer (which includes the excess diagnoses associated with use of hormone replacement therapy in the late 1990s and early 2000s) was almost entirely attributable to changes in the incidence of regional (largely
node-positive) disease. However, the incidence of distant (metastatic) disease remained unchanged (95% CI for the APC, −0.19% to 0.14%). The SEER data did not distinguish between women who were screened and those who were not screened. Helvie et al. (2014), similarly to Bleyer & Welch (2012), compared the SEER breast cancer incidence and stage for the pre-mammography period (1977–1979) and the mammography screening period (2007–2009) in women older than 40 years. The authors estimated pre-screening temporal trends using several measures of APC. Stage-specific incidence values for 1977–1979 (baseline) were adjusted using APC values of 0.5%, 1.0%, 1.3%, and 2.0% and then compared with observed stage-specific incidence in 2007–2009. Pre-screening APC temporal trend estimates ranged from 0.8% to 2.3%. The joinpoint estimate of 1.3% for women older than 40 years approximated the four-decade-long APC trend of 1.2% noted in the Connecticut Tumor Registry. At an APC of 1.3%, late-stage breast cancer incidence decreased by 37% (56 cases per 100 000 women), with a reciprocal increase in early-stage rates noted from 1977–1979 to 2007–2009. The resulting late-stage breast cancer incidence decreased by 21% at an APC of 0.5% and by 48% at an APC of 2.0%. Total invasive breast cancer incidence decreased by 9% (27 cases per 100 000 women) at an APC of 1.3%. According to the authors, a substantial reduction in late-stage breast cancer has occurred in the mammography era when appropriate adjustments are made for pre-screening temporal trends.

Hou & Huo (2013) analysed the SEER age-standardized breast cancer incidence rates from 2000 to 2009, for 677 774 women aged 20 years and older. This study represents a descriptive analysis of population-based cancer incidence rates from 18 SEER registries with high-quality data, representing 28% of the United States population. Since 2004, incidence rates in women aged 40–49 years increased significantly for most racial/ethnic groups (overall APC, 1.1%; \( P = 0.001 \)). The incidence rate of DCIS increased significantly in all racial/ethnic groups, with an APC range from 2.3% to 3.0% (\( P < 0.005 \)). The incidence rate of localized breast cancer increased significantly in non-Hispanic Black women (APC, 1.3%; \( P = 0.004 \)) and Asian women (APC, 1.2%; \( P = 0.03 \)). The incidence rates of regional and distant cancers decreased significantly in non-Hispanic White women from 2000 to 2004 (APC, −2.5%; \( P = 0.02 \)) and in Hispanic women from 2000 to 2009 (APC, −1.1%; \( P = 0.006 \)). It is possible that the changes in incidence rates are due in part to improvements in cancer screening methods and, therefore, advances in early detection. It is unlikely that the overall trends of incidence rates are due to changes in the mammography screening rate, since mammography use did not change substantially from 2000 to 2010, although it increased by large magnitudes in small groups with growing populations, such as new immigrants and Asian-Americans.

DeSantis et al. (2014) obtained data on incidence, probability of developing cancer, and cause-specific survival from SEER, and data on the prevalence of mammography by age from the 2010 and 2012 Behavioral Risk Factor Surveillance System, to assess the relationship between mammography screening rates in 2010 and breast cancer stage at diagnosis in 2006–2010. Among non-Hispanic White women, state-level mammography screening prevalence was positively correlated with the percentage of breast cancers diagnosed at the in situ stage (correlation coefficient, \( r = 0.62; P < 0.001 \)) and negatively correlated with the percentage of breast cancers diagnosed at late stages (\( r = −0.51; P < 0.001 \)).

(d) Modifying effects of breast density

Given that increased mammographic breast density is associated with lower sensitivity and higher interval cancer rates (Mandelson et al., 2000), its potential role as an effect modifier of mammography screening effectiveness is
Breast cancer screening

of interest. The effect of breast density on case fatality rate, or breast density as a modifier, has been investigated in several studies. Only one of these has examined differences in survival of women with interval cancers in those with dense versus non-dense breasts. This study in Sweden found that women with interval cancers had worse survival than women with screen-detected cancers (hazard ratio [HR], 1.69; 95% CI, 1.03–2.76, overall) and that interval-cancer survival was poorer in those with non-dense breasts (HR, 1.76; 95% CI, 1.01–3.09) than in those with dense breasts (HR, 1.26; 95% CI, 0.47–3.38) (Eriksson et al., 2013). These effects were observed after adjustment for tumour size and lymph-node metastasis at diagnosis. [Before adjustment, hazard ratios were stronger.]

The remaining studies examined the impact of breast density on survival or mortality rates within populations where screening is available, but they did not differentiate between interval and screen-detected cancers. In a cohort in Denmark participating in biennial mammography at ages 50–69 years, during 1991–2001, the case fatality rate was lower in women with mixed/dense breasts than in those with fatty breasts (HR, 0.60; 95% CI, 0.43–0.84) (Olsen et al., 2009). [Although the case fatality rate is lower for women with dense breasts, it should be noted that because more women with dense breasts develop breast cancer, more women with dense breasts die from breast cancer overall.] In the USA, a study using the Carolina Mammography Registry (22 597 breast cancers) showed no difference in breast cancer mortality between women with dense breasts and those with fatty breasts, after adjusting for incidence differences (HR, 0.908; \( P = 0.12 \)) (stage-adjusted) (Zhang et al., 2013). Similarly, the American College of Radiology Breast Imaging Reporting and Data System (BI-RADS) density score was not associated with breast cancer survival (HR for breast cancer death, 0.92; 95% CI, 0.71–1.19) in the United States Breast Cancer Surveillance Consortium (Gierach et al., 2012), except for an increased risk of breast cancer death among women with low breast density (BI-RADS 1) who were obese or had tumours larger than 20 mm. The Kopparberg RCT, in Sweden, suggested that women with dense breasts have higher breast cancer incidence rates (multivariate RR, 1.57; 95% CI, 1.23–2.01) and breast cancer mortality (RR, 1.91; 95% CI, 1.26–2.91), but that there was no clear difference in survival between women with dense breasts and those with non-dense breasts (HR, 1.41; 95% CI, 0.92–2.14) (not adjusted for tumour characteristics) (Chiu et al., 2010).

One study found poor survival in women with dense breasts compared with those with fatty breasts in women diagnosed at the first screening round but not in those diagnosed at later rounds (rounds 5–10) (van Gils et al., 1998). [The Working Group noted that although breast cancers occurring in dense breasts are more likely to be interval cancers, there is no indication that breast cancer survival rates are poorer for these cancers (despite a shorter lead-time bias). In addition, the studies were performed with screen-film mammography, so it is difficult to extrapolate the results to digital methods.]

(e) Effects of population-based mammography screening in the presence of adjuvant systemic therapy

RCTs of mammography screening, mostly performed in the 1980s or earlier, have reported reductions in breast cancer mortality in women aged 50–69 years. However, the present-day relevance of these trials has been debated because the management and treatment of breast cancer has changed considerably in the past decades (Gøtzsche & Nielsen, 2009; Kalager et al., 2010; Paci & EUROSCREEN Working Group, 2012; Marmot et al., 2013). Adjuvant systemic therapy has been increasingly used since the late 1980s, and its dissemination and effectiveness have progressed since then (van de Velde et al., 2010).
Such developments have probably affected the impact of screening, also in service screening programmes (Berry et al., 2005). This section discusses studies of the effects of adjuvant systemic therapy and mammography screening in current health-care systems.

The effects of adjuvant treatment and mammography screening were calculated for the Netherlands using the Microsimulation Screening Analysis (MISCAN) model (de Gelder et al., 2015). [Models can extrapolate findings from screening and adjuvant treatment trials to actual populations, can allow for comparison of intervention strategies, and can separate effects on the natural history of disease, for example screening effects and adjuvant treatment effects (Berry et al., 2005; Mandelblatt et al., 2009) (see Section 5.1.2f).] In the MISCAN model, the progression was modelled as a semi-Markov process through the successive preclinical invasive stages T1a, T1b, T1c, and T2+. The mean duration of the preclinical detectable phase, the probability of a transition between the stages, and the mammography sensitivity were then estimated, using detailed data from screening registries. Data on adjuvant systemic therapy were derived from comprehensive cancer centres. Cure and survival rates after screen detection were based on RCTs (de Koning et al., 1995; Tabár et al., 2000; Nyström et al., 2002; Bjurstam et al., 2003). The risk of death from breast cancer after adjuvant treatment was modelled using the rate ratios from the meta-analysis of the Early Breast Cancer Trialists’ Collaborative Group (2005). In 2008, adjuvant treatment was estimated to have reduced the breast cancer mortality rate in the simulated population by 13.9%, compared with a situation without treatment. Biennial screening between age 50 years and age 74 years further reduced the mortality rate by 15.7%. Extending screening to age 48 years would lower the mortality rate by 1.0% compared with screening from age 50 years; 10 additional screening rounds between age 40 years and age 49 years would reduce this rate by 5.1%. Adjuvant systemic therapy and screening reduced breast cancer mortality by similar amounts.

A previous modelling study, which included six natural history models for the population in the USA, had estimated an approximately equal contribution of adjuvant therapy and screening to the observed mortality reduction in the USA (Berry et al., 2005), using very similar techniques to those described above.

These analyses have recently been updated, taking into account the receptor-specific heterogeneity of breast cancer (Munoz et al., 2014), by using six established population models with ER-specific input parameters on age-specific incidence, disease natural history, mammography characteristics, and treatment effects to quantify the impact of screening and adjuvant therapy on age-adjusted breast cancer mortality in the USA by ER status from 1975 to 2000. In 2000, actual screening and adjuvant treatment were estimated to have reduced breast cancer mortality by 34.8%, compared with the situation if no screening or adjuvant treatment had been present; a reduction by 15.9% was estimated to have been a result of screening, and 23.4% as a result of treatment. For ER-positive cases, adjuvant treatment made a higher relative contribution to breast cancer mortality reduction than screening, whereas for ER-negative cases the relative contributions were similar for screening and adjuvant treatment. Although ER-negative cases were less likely to be screen-detected than ER-positive cases (35.1% vs 51.2%), when they were screen-detected, the survival gain was greater for ER-negative cases than for ER-positive cases (5-year breast cancer survival, 35.6% vs 30.7%).
5.3 Adverse effects of mammography

5.3.1 False-positive rates

A screening test is not diagnostic but should identify asymptomatic women who are at risk of harbouring an undiagnosed cancer. The screening episode in organized screening should end with an unequivocal diagnostic report: there is or there is not cancer (Perry et al., 2006). A woman in whom an abnormality is detected by screening and whose investigations end with a negative result has a false-positive result. This result closes the screening episode.

In a recent survey of 20 population-based screening programmes in 17 European countries, the Euroscreen and EUNICE Working Group (Hofvind et al., 2012a) reported average recall rates varying from 9.3% at the initial screening episode (range, 2.2–15.6%) to 4.0% at subsequent screening episodes (range, 1.2–10.5%). The average rates of needle biopsy were 2.2% at the initial screening and 1.1% at subsequent screenings. The variation depends on differences between national protocols and a variety of local conditions. Over the whole diagnostic phase, the benign-to-malignant ratio ranged from 0.09 in the United Kingdom to 0.21 in Luxembourg, with an average of 0.11.

The difference in the performance of the assessment phase between opportunistic screening and service screening has been estimated by comparing screening in the USA and population-based programmes in Europe. Smith-Bindman et al. (2005) compared the performance of screening in the United Kingdom and the USA. The outcomes included (per 1000 women screened for 20 years) a detection rate of carcinoma in situ of 12.3 in the USA compared with 8.3 in the United Kingdom, a rate of non-invasive diagnostic tests for assessment of recalled women of 553 in the USA compared with 183 in the United Kingdom, and a biopsy rate of 142 in the USA compared with 85 in the United Kingdom, of which 54 and 25, respectively, were open surgical biopsies.

Hofvind et al. (2012b) compared the Norwegian mammography screening programme with screening practice in Vermont, USA (Vermont is a member of the Breast Cancer Surveillance Consortium, an initiative of the United States National Cancer Institute), showing that higher recall rates and lower specificity in the USA were not associated with higher sensitivity. These differences may be explained by professional practices, since screening centres in the USA usually have small volumes of mammography readings, and double reading is not a quality requirement in the USA as it is in Europe (Burnside et al., 2014).

The cumulative risk of a false-positive recall is one of the most important harms of screening. The false-positive rate is estimated from the recall rate by subtracting the cancer detection rate in the same screening episode. The cumulative risk of a false-positive result is defined as the cumulative risk of recall for further assessment at least once during the screening period (usually 10 biennial screening episodes in organized programmes) minus the cumulative risk of cancer detection over the same period. There is a similar definition for the cumulative risk of having an invasive procedure (needle biopsy or surgical biopsy) with a benign outcome.

A systematic review has been made of publications estimating the cumulative risk of a false-positive result in European population-based mammography screening programmes (Hofvind et al., 2012a). Four studies were included, based on data from the 1990s and conducted in Denmark, Italy, Norway, and Spain. Results updated with a further 9 years of experience in Norway have since been published (Román et al., 2013). The cumulative risk of any further assessment without cancer diagnosis varied from 8.1% to 20.4% in the most recent period (ending variously in 2001 to 2010), and the cumulative risk of
assessment with an invasive procedure without cancer diagnosis varied from 1.8% to 4.1%.

The cumulative risk of false-positives is higher in opportunistic mammography screening, which is the usual modality in the USA. Elmore et al. (1998) estimated that 41% of screened women had at least one false-positive result over 10 screening episodes. Hubbard et al. (2010) applied statistical models to more recent data from the Breast Cancer Surveillance Consortium for women aged 40–59 years at entry and followed up over their screening history. The risk of a false-positive over 10 screening mammograms varied between 58% and 77%.

Román et al. (2012) assessed factors affecting the false-positive rate after any assessment, and after assessment with an invasive procedure, in a retrospective cohort in Spain. The authors reported that the false-positive risk after assessment with an invasive procedure was less for digital mammography (RR, 0.83) than for non-digital mammography, and they estimated a total cumulative risk of 20.4%, ranging from 51.4% for the highest risk profile to 7.5% for the lowest risk profile. The risk after assessment with all procedures and with invasive procedures was estimated to be higher for younger women (OR, 1.30 for age 40–44 years; OR, 1.26 for age 40–54 years; reference category, age 65–69 years).

In the USA, Kerlikowske et al. (2013) assessed the cumulative risk by breast density and risk profile. The cumulative probability of a false-positive mammography result was higher among women with extremely dense breasts who underwent annual mammography and either were aged 40–49 years (65.5%) or used combined estrogen–progestogen hormone therapy (65.8%), and was lower among women aged 50–74 years who underwent biennial or 3-yearly mammography and had scattered fibroglandular densities (30.7% and 21.9%, respectively) or fatty breasts (17.4% and 12.1%, respectively).

Indicators of the cumulative risk of false-positives are included as possible harms of screening in the balance sheet of benefits and harms. The Euroscreen mammography screening balance sheet considered 1000 women who were aged 50 or 51 years at the start of their screening regimen. The cumulative risk of false-positives was estimated to be 200 over the 10 screening rounds from age 50 years to age 69 years; 170 women were recalled for further assessment without invasive procedures, and 30 women had further assessment with invasive procedures (Paci & EUROSCREEN Working Group, 2012).

5.3.2 Overdiagnosis

The definition of overdiagnosis and estimates of overdiagnosis in randomized trials of mammography screening have been presented in Section 4.2.3c. The quantification of overdiagnosis is important in observational studies because this harm was not a primary end-point of the RCTs and estimates are influenced by local screening practice and technological innovation. Other approaches, such as radiological doubling time, have been suggested as useful indicators for the study of overdiagnosis, but in this section overdiagnosis is considered as an epidemiological construct, based on a retrospective analysis of breast cancer diagnosis in the population.

Several approaches have been proposed for estimating overdiagnosis in observational studies.

The cumulative incidence method estimates overdiagnosis by following up a cohort of women, invited and not invited to screening or screened and not screened. The ideal study would require the follow-up of pairs of birth or enrolment age cohorts in which one cohort is invited to screening and the other is not invited (Møller et al., 2005; Biesheuvel et al., 2007). The attribution of an individual time zero to each invited woman allows for estimation of changes in incidence over the screening period in the population and monitoring of the compensatory drop phase after the end of screening (Fig. 5.2).
Fig. 5.2 Observed and modelled breast cancer incidence per 100,000 person–years in the presence and absence of screening in 1990–2006

Values after years indicate: percentage of the target population aged 49–69 years invited, fraction of prevalent screenings. (A) 1990: 9.2%, 74%; (B) 1992: 47.4%, 77%; (C) 1994: 74.3%, 49%; (D) 1996: 92.0%, 39%; (E) 1998: 80.8%, 20%; (F) 1999: 91.8%, 19%; (G) 2000: 94.4%, 18%; (H) 2002: 96.1%, 14%; (I) 2004: 95.8%, 14%; (J) 2006: 92.2%, 13%. Solid lines, modelled with screening; dashed lines, modelled without screening; triangles, observed.

The *incidence-rate method* compares the average annual incidence of breast cancer over a defined period of follow-up in a specified age group of women who were offered or accepted screening with an estimate of the average annual incidence of breast cancer during the same period in women who were not offered screening or were not screened. Overdiagnosis is taken to be any excess in incidence in the former over the latter once the screening lead time has been accounted for. Several methods have been suggested for the adjustment for lead time, with the aim of overcoming the frequent difficulty of too short a follow-up period for the lead time to have passed in all women under observation who had been invited to screening or were screened.

In a methodological study, Etzioni et al. (2013) contrasted an incidence excess approach with a lead-time approach. The *lead-time approach* uses the disease incidence under screening to make inferences about the lead time or the natural history of the disease. Using the *incidence excess approach*, the authors suggested that the estimate should consider the time needed for screening dissemination and the compensatory drop, as expressed by incidence rates at older ages. In the presence of a shorter follow-up time and/or unequal screening periods in the age cohorts of women, statistical adjustment for lead time is required. This can be based on estimates of lead time derived from clinical cancers (such as estimates derived from experience before the introduction of population screening programmes) or estimates from modelling studies.

Simulation, using statistical modelling, of lifetime individual histories with or without screening is often used to overcome the complexity of screening evaluation, in particular to account for lead time and to give understandable outcomes (see Section 5.1.2f). Complex models such as these need a set of assumptions about natural history of the disease and screening performance (Tan et al., 2006), which would ideally be clearly stated in reports based on the models’ use but generally are not. Importantly, too, a paucity of relevant empirical evidence means that assumptions about the proportion of preclinical cancers that are non-progressive and the range and distribution of lead time, which are critical to modelled estimates of overdiagnosis, are very uncertain.

Duffy & Parmar (2013), using estimates of the incidence rate in the United Kingdom and an exponential distribution of the lead time, simulated the time course of incidence rates during and after the screening period in the absence of overdiagnosis. With a 20-year period of screening (from age 50 years to age 69 years), a period of at least 10 years must elapse after the screening period (to when women are aged 79 years) for the *excess incidence rate* to be close to the rate observed in the absence of screening (to within 1% of excess with 30 years of follow-up from the start of screening). It is important to note that in the same simulation, 10 years of observation of a population of women screened from age 50–69 years at the start of screening will give an incidence excess of 50%. This model assumed an average lead time of 40 months. However, some estimates are much lower (see, for example, Feinleib & Zelen, 1969). Although there is disagreement over the average and distribution of lead time for breast cancer, the main conclusion is that an adequate correction for lead time is needed in the absence of a sufficient follow-up period to distinguish *excess of incidence* due to lead time from *overdiagnosis*.

An important factor determining the observational estimate of overdiagnosis is the estimate of the underlying incidence. In descriptive epidemiological studies, an estimate of incidence in the absence of screening is needed. In comparative studies, the reference population should be comparable to the invited population so far as is possible in terms of the background incidence rate, breast cancer risk factors, socioeconomic status, and use of health services other than for
mammography. If rates from the same or another historical (pre-screening) population are used, the time trend in the underlying incidence must be estimated, a projection made to the screened population, and sensitivity analyses of the estimates made that take account of variation in the trend due to unpredicted changes in population composition or the prevalence of risk factors. Self-selection bias should also be considered and adjusted for if attenders only are evaluated.

Adjustment for lead time and estimation of the underlying incidence of breast cancer in the absence of screening (control of confounding due to differences in breast cancer risk factors between screened and unscreened women) were considered as the main problems in estimating overdiagnosis in observational studies (Njor et al., 2013a), but these are not the only factors to be considered. Others include (Njor et al., 2013a): the nature and quality of the observational data used; what estimate was actually reported as a measure of overdiagnosis (ideally classified in the terms outlined by the Independent UK Panel on Breast Cancer Screening, 2012), which is sometimes not clearly described, and, for the Independent United Kingdom Panel’s measure A (the excess cancers expressed as a proportion of cancers diagnosed over the whole follow-up period in unscreened women), ranged from −0.7% to 76% for invasive cancer only and from 1% to 57% for invasive and in situ cancers together.

Observational studies of overdiagnosis for women aged 50–69 years are summarized in Table 5.11 and Table 5.12. Table 5.11 covers studies reviewed by the Euroscreen Working Group (Puliti et al., 2012), which included all 13 observational studies conducted in Europe that were published up to February 2011. Table 5.12 covers 17 studies conducted in Europe and published from February 2011 to November 2014, when the Handbook Working Group met, or conducted outside Europe and published up to November 2014.

Estimates of the overdiagnosis risk, principally the Independent United Kingdom Panel’s measure A (the excess cancers expressed as a proportion of cancers diagnosed over the whole follow-up period in unscreened women), ranged from −0.7% to 76% for invasive cancer only and from 1% to 57% for invasive and in situ cancers together.

The Euroscreen Working Group characterized overdiagnosis estimates as made with or without correction for lead time and underlying incidence trend. The reported estimates that were considered as adequately adjusted for both biases (from 6 of the 13 studies) ranged from 1% to 10% excess over the expected incidence for all breast cancers (measure A) (1% to 10% for invasive cancer only, from 4 studies, and 1% to 7% for invasive and in situ cancers, from 4 studies). The majority of the studies used temporal trends or geographical differences in dynamic populations to adjust for the underlying incidence. Only two studies used the cohort population approach, and a few studies used statistical modelling for the estimate. The Euroscreen Working Group derived a summary estimate of overdiagnosis of 6.5% of the incidence in the absence of screening. This is the estimate of the overdiagnosis in women screened between the ages of 50 years and 69 years and followed up for 10 years after the last screening, and included carcinoma in
### Table 5.11 Studies of the estimates of overdiagnosis in Europe

<table>
<thead>
<tr>
<th>Reference</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Country (area)</strong></td>
<td><strong>Type of population and study design</strong></td>
<td><strong>Age and interval of screening</strong></td>
<td><strong>Reference population</strong></td>
<td><strong>Adjustment for breast cancer risk</strong></td>
</tr>
<tr>
<td>Paci et al. (2004)</td>
<td>Italy (Florence) 1990–1999</td>
<td>Dynamic population Cohort</td>
<td>50–69 yr 2 yr 1990</td>
<td>Pre-screening incidence (1985–1990)</td>
</tr>
</tbody>
</table>

*Assumptions and methods vary across studies.*
<table>
<thead>
<tr>
<th>Reference</th>
<th>Country (area)</th>
<th>Calendar period of screening</th>
<th>Type of population and study design</th>
<th>Age and interval of screening</th>
<th>Reference population</th>
<th>Comparison</th>
<th>Adjustments for breast cancer risk</th>
<th>Adjustments for lead time</th>
<th>Mean follow-up after end of screening (range)</th>
<th>Estimate of overdiagnosis (only invasive)</th>
<th>Estimate of overdiagnosis (in situ and invasive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference</td>
<td>Country (area)</td>
<td>Population and study design</td>
<td>Age and interval of screening</td>
<td>Reference population</td>
<td>Adjustment for breast cancer risk</td>
<td>Adjustment for lead time</td>
<td>Mean follow-up after end of screening (range)</td>
<td>Estimate of overdiagnosis (in situ and invasive)</td>
<td>Estimate of overdiagnosis (only invasive)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------</td>
<td>-----------------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
<td>----------------------------------</td>
<td>-------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puliti et al. (2009)</td>
<td>Italy (Florence) 1990–2004</td>
<td>Birth cohort</td>
<td>50–69 yr 2 yr 1990</td>
<td>Pre-screening incidence (1986–1990)</td>
<td>Age and temporal trend</td>
<td>Compensatory drop</td>
<td>4.7 yr (1–14 yr)</td>
<td>0.99%</td>
<td>1.0%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martinez-Alonso et al. (2010)</td>
<td>Spain (Catalonia) 1990–2004</td>
<td>Dynamic population Statistical model</td>
<td>50–64 yr (extended to 65–69 yr) 2 yr 1990</td>
<td>Pre- and post-screening incidence (women aged 20–84 yr from 1980–2004)</td>
<td>Age, year of birth, fertility rate, and use of mammography</td>
<td>Statistical adjustment</td>
<td>Not applicable</td>
<td>0.4%–46.6%, depending on birth cohort</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Gelder et al. (2011b)</td>
<td>Netherlands 1989–2006</td>
<td>Dynamic population MISCAN model</td>
<td>49–69 yr (extended to 74 yr) 2 yr 1990</td>
<td>Predicted incidence without screening</td>
<td>Not needed</td>
<td>Compensatory drop</td>
<td>6.1 yr (1–16 yr)</td>
<td>NR</td>
<td>3.6%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Studies published up to February 2011 and included in the review by Euroscreen.
* Measures of overdiagnosis are equivalent to measure A of the [Independent UK Panel on Breast Cancer Screening (2012)](#).
* Period of screening that contributed to the estimate of overdiagnosis.
* First year of the screening programme or intervention to which the overdiagnosis estimate relates.
* A compensatory drop was observed by Zahl et al. (2004) (11% in Norway and 12% in Sweden) but was not taken into account in the estimation of overdiagnosis because it was not statistically significant.
* Recalculated as measure A by Puliti et al. (2012).
* AORH, Akershus, Oslo, Rogaland, Hordaland; CIS, carcinoma in situ; HRT, hormone replacement therapy; MISCAN, Microsimulation Screening Analysis; NR, not reported; yr, year or years.
* Modified from Puliti et al. (2012).
Table 5.12 Studies of estimates of overdiagnosis in Europe (published from February 2011 to November 2014) and in other countries (published up to November 2014)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Population</th>
<th>Intervention</th>
<th>Age and interval of screening</th>
<th>Reference population</th>
<th>Comparison</th>
<th>Adjustment for breast cancer risk</th>
<th>Measure of overdiagnosis</th>
<th>Estimate of overdiagnosis (only invasive)</th>
<th>Estimate of overdiagnosis (in situ and invasive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jørgensen &amp; Gøtzsche (2009)</td>
<td>Australia (New South Wales) 1996–2002</td>
<td>Dynamic population Ecological</td>
<td>50–69 yr 2 yr 1988</td>
<td>Pre-screening incidence (1972–1987)</td>
<td>Age and temporal trend</td>
<td>Compensatory drop: no drop was observed in women aged 70–79 yr</td>
<td>Measure A</td>
<td>38%</td>
<td>53%</td>
</tr>
<tr>
<td>Reference</td>
<td>Population</td>
<td>Intervention</td>
<td>Comparison</td>
<td>Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------</td>
<td>--------------</td>
<td>------------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Country (area)</td>
<td>Calendar period of screening</td>
<td>Reference population</td>
<td>Adjustment for breast cancer risk</td>
<td>Adjustment for lead time</td>
<td>Measure of overdiagnosis</td>
<td>Estimate of overdiagnosis (only invasive)</td>
<td>Estimate of overdiagnosis (in situ and invasive)</td>
<td></td>
</tr>
<tr>
<td>Junod et al. (2011)</td>
<td>France 1988–2005</td>
<td>Dynamic population Ecological</td>
<td>(i) For women aged 50–64 yr, incidence in the same age cohort born 15 yr earlier (1926–1930) (ii) For women aged 65–79 yr, incidence in the same age cohort born 15 yr earlier (1911–1915)</td>
<td>None</td>
<td>Measure A</td>
<td>(i) 76% (ii) 23%</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seigneurin et al. (2011)</td>
<td>France (Isère) 1991–2006</td>
<td>Statistical model of birth cohorts 1922–1956</td>
<td>Predicted pre-screening incidence (birth cohorts 1900–1950)</td>
<td>Age, temporal trend, and opportunistic screening</td>
<td>Simulation of sojourn times with various distributions of unknown parameters</td>
<td>(i) 3.3% (ii) 1.5%</td>
<td>Only in situ: (i) 31.9% (ii) 28.0% 31.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puliti et al. (2012)</td>
<td>Italy (Florence) 1991–2008</td>
<td>Dynamic population Cohort</td>
<td>Incidence in screening non-attenders</td>
<td>Age, marital status, and SES</td>
<td>Compensatory drop: 5–14 yr since last screen</td>
<td>Measure A</td>
<td>5%</td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Population</td>
<td>Intervention</td>
<td>Comparison</td>
<td>Outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>------------</td>
<td>--------------</td>
<td>------------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bleyer &amp; Welch (2012)</td>
<td>USA 1979–2008</td>
<td>Dynamic population Ecological</td>
<td>≥ 40 yr 1 yr 1971 Incidence before widespread screening (1976–1978)</td>
<td>Age, use of HRT, and temporal trend No explicit adjustment for lead time. Overdiagnosis estimated from difference between increase in incidence of early breast cancer and fall in incidence of advanced breast cancer when screening steady state reached Overdiagnosed cancers as a percentage of all cancers diagnosed in the population</td>
<td>20% 31%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reference</td>
<td>Country</td>
<td>Calendar period of screening</td>
<td>Type of population and study design</td>
<td>Age and interval of screening</td>
<td>Start year of screening</td>
<td>Adjustment for breast cancer risk</td>
<td>Adjustment for lead time</td>
<td>Measure of overdiagnosis</td>
<td>Estimate of overdiagnosis (only invasive)</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------</td>
<td>-----------------------------</td>
<td>-------------------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>-------------------------</td>
<td>------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>Lund et al. (2013)</td>
<td>Norway 2002–2010</td>
<td>52–69 yr 2 yr 2002</td>
<td>Incidence in unscreened women</td>
<td>Age, parity, use of HRT, family history, and BMI</td>
<td>Compensatory drop: included women up to age 79 yr in incidence</td>
<td>Measure A</td>
<td>7.5%</td>
<td>22.0%</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.12  (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Country (area) Calendar period of screening†</td>
<td>Type of population and study design</td>
<td>Age and interval of screening Start year of screening‡</td>
<td>Reference population Adjustment for breast cancer risk Adjustment for lead time Measure of overdiagnosis Estimate of overdiagnosis (only invasive) Estimate of overdiagnosis (in situ and invasive)</td>
</tr>
<tr>
<td>Njor et al. (2013b)</td>
<td>Denmark (Copenhagen and Funen) (i) Copenhagen: 1991–2005 (ii) Funen: 1993–2004</td>
<td>Dynamic population Birth cohorts: (i) 1921–1935 (ii) 1923–1934</td>
<td>56–69 yr 2 yr (i) 1991 (ii) 1993</td>
<td>Incidence in: (1) historical pre-screening birth cohorts from same regions (2) contemporary regions not invited to screening (3) national pre-screening historical birth cohort</td>
</tr>
<tr>
<td>Coldman &amp; Phillips (2013)</td>
<td>Canada (British Columbia) 2000–2009</td>
<td>Dynamic population Cohort</td>
<td>40–49 yr 1 yr ≥ 50 yr 2 yr 1988</td>
<td>Incidence in women who did not attend screening</td>
</tr>
</tbody>
</table>
Table 5.12  (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Population</th>
<th>Intervention</th>
<th>Comparison</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Country (area) Calendar period of screening</td>
<td>Type of population and study design</td>
<td>Age and interval of screening Start year of screening</td>
<td>Adjustment for breast cancer risk</td>
</tr>
<tr>
<td><strong>Heinävaara et al. (2014)</strong></td>
<td>Finland (Helsinki) 1986–1997</td>
<td>Dynamic population Ecological</td>
<td>50–59 yr 2 yr 1986</td>
<td>Incidence in: (i) last unscreened birth cohort (1930–1934) (ii) 5-yr birth cohorts from 1920–1924 to 1930–1934 (from statistical model)</td>
</tr>
</tbody>
</table>
### Table 5.12 (continued)

<table>
<thead>
<tr>
<th>Reference</th>
<th>Country (area)</th>
<th>Calendar period of screening</th>
<th>Type of population and study design</th>
<th>Age and interval of screening</th>
<th>Start year of screening</th>
<th>Reference population</th>
<th>Comparison</th>
<th>Adjustment for breast cancer risk</th>
<th>Adjustment for lead time</th>
<th>Measure of overdiagnosis</th>
<th>Estimate of overdiagnosis (only invasive)</th>
<th>Estimate of overdiagnosis (in situ and invasive)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beckmann et al. (2015)</td>
<td>Australia (South Australia) 1989–2010</td>
<td>Dynamic population Case–control study nested within a cohort</td>
<td>40–69 yr 1 yr (increased risk) or 2 yr 1989</td>
<td>Women who did not attend screening in 1989–2010</td>
<td>Age, temporal trend (1977–1988), SES, and area</td>
<td>Compensatory drop: included women up to age 85 yr in incidence and ≥ 10 yr since last screen</td>
<td>Measure A⁴</td>
<td>[8.3%]</td>
<td>[16.0%]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

⁴ Period of screening that contributed to the estimate of overdiagnosis.
⁵ First year of the screening programme or intervention to which the overdiagnosis estimate relates.
⁶ Measures are equivalent to measure A of the Independent UK Panel on Breast Cancer Screening (2012) unless otherwise indicated.
⁷ This estimate was not adjusted for lead time.
⁸ Results are those from the authors’ Approach 1, which the Working Group considered to be the preferred of the two approaches the authors took to adjustment for lead time.
⁹ The estimate of the percentage risk of overdiagnosis reported in Beckmann et al. (2015) is measure B, with women exposed to screening as the denominator. The Working Group recalculated this as measure A using data provided in Beckmann et al. (2015).

BMI, body mass index; HRT, hormone replacement therapy; NR, not reported; OD, overdiagnosis; SES, socioeconomic status; yr, year or years.
5.3.3 Overtreatment

Over the past 50 years, breast cancer care has moved from aggressive, mutilating surgery to breast-conserving treatment (Fisher et al., 2002; Veronesi et al., 2002). This change was the starting point for improvements in other treatment and assessment areas, such as, for example, the sentinel lymph node procedure, which has been well established in clinical practice since the early 2000s (Veronesi et al., 2003). Detection of early, indolent lesions, such as carcinoma in situ (Ernster et al., 2002), is a major area of concern. In a recent international survey, Lynge et al. (2014) documented the wide variability in the occurrence of in situ breast cancer across countries. In a comparison with European programmes, higher probabilities for the occurrence of carcinoma in situ were reported in the USA. This finding is associated with higher false-positive rates and biopsy rates in the diagnostic assessment phase (Smith-Bindman et al., 2005).

Carcinomas in situ have high survival rates after treatment, but studies have shown that only a proportion of them, depending mainly on the pathological grade, would have progressed to invasiveness over the lifetime of the woman in the absence of early diagnosis. Overdiagnosed breast cancer cases are all overtreated. Carcinoma in situ is considered a major area of overtreatment. However, overtreatment is a harm not limited to screen-detected cases. Clinicians follow shared guidelines, primarily based on the stage at presentation of the disease. Screen-detected cases, when treated in the same cancer unit, will receive treatment by tumour characteristics. Chemotherapy and hormone therapy for breast cancer are progressively being extended to very early and less-progressive cancers (Peto et al., 2012), with important implications when there is a growing proportion of early, high-survival-rate breast cancers.

An example of the relationship between overdiagnosis and overtreatment is the comparison of mastectomy rates in the screening and pre-screening epochs. In a Cochrane systematic review (Gøtzsche & Jørgensen, 2013), a 31% increase in mastectomy and lumpectomy rates (20% excess of mastectomies) was estimated in the intervention group compared with the control group. This estimate considered all breast cancer cases detected in the screening period (i.e. the excess of incidence observed in the screening arm).

Zorzi et al. (2006) evaluated the use of mastectomy in Italy in the period 1997–2001, during which a large number of screening programmes were implemented, using individual data classified by stage and modality of diagnosis in relation to screening. The probability of a mastectomy increased with age and primary tumour size, and screen-detected cases were half as likely to be treated with mastectomy as non-screen-detected cases. The increasing rates of early-stage cancers (< 30 mm) and the use of breast-conserving treatment paralleled a decline in the mastectomy rate and in the incidence of advanced-stage cancers (> 30 mm), showing an appropriate use of the surgical approach.

Suhrke et al. (2011), using population-based data in the epoch of change to a service screening programme, showed an increase in rates of breast surgery and also an increase in mastectomy...
Breast cancer screening

rates immediately after the start of the screening programme. They described a recent decline in mastectomy rates and suggested that the change affected all age groups and that it is likely to have resulted from changes in surgical policy.

5.3.4 Risk of breast cancer induced by radiation

Exposure of the breast to ionizing radiation may induce breast cancers (see Section 1.3.4). The low dose of X-ray photon radiation received during mammography is thus considered as a potential adverse effect of breast cancer screening. The number of cancers caused by screening with mammography must be estimated to evaluate the balance between benefits and risks. However, due to the small number of expected cases, it is not possible to estimate such a number from epidemiological data. Thus, numerous studies have used a quantitative risk assessment approach. This approach is based on a large number of hypotheses arising from current scientific knowledge and on hypotheses about screening modalities.

(a) Hypotheses for quantitative risk assessment

(i) Hypotheses about risk models

Hypotheses about risk models come from the selection of the most reliable studies on the relationship between radiation exposure and breast cancer risk (see Section 1.3.4). Hypotheses are made about the form of this relationship, the modifying effect of time and age at exposure, the latency time between exposure and risk, and transposition from high to low dose and low exposure rate.

The most recent models for such an exercise in the general population arise from the BEIR VII models of the United States National Academy of Sciences (National Research Council, 2006), with recommendations of the use of an excess absolute risk model for breast cancer risk (National Research Council, 2006; ICRP, 2007; Wrixon, 2008). This model assumes no threshold, even at a very low dose, and a decreasing effect with increasing age at exposure. Coefficients are estimated from atomic bomb survivors and women medically exposed to radiation (see Section 1.3.4). Because these studies are based on a higher dose and a higher dose rate than those typically involved in mammography screening, an effort was made by some authors to produce results taking into account transposition factors from high to low dose and dose rate (dose and dose rate effectiveness factor). Values of this factor in the context of mammography generally vary between 1 and 2 (National Research Council, 2006; Law et al., 2007; Heyes et al., 2009).

A hypothesis about the latency time for the induction of a breast cancer by radiation is also needed for risk assessment. A latency time of 10 years is generally used, with values varying from 5 years to 15 years.

(ii) Hypotheses about doses received during mammography

The estimation of doses received by the glandular tissue of the breast depends on breast thickness and density. Based on an extensive literature review, a historical reconstruction of doses received during mammography shows a strong decrease over time, with an estimated mean glandular dose to the breast of 2 mGy per view since 2000 (Thierry-Chef et al., 2012) (see Section 1, Fig. 1.16). Moreover, recent use of digital mammography (instead of screen-film mammography) has led to new estimates of doses received (Hendrick et al., 2010; Hauge et al., 2014).

(iii) Hypotheses about the target population and screening modalities

To fully develop the risk assessment, scenarios for the target population and screening modalities (age range, frequency, number of examinations at each screening, additional views, etc.) have been developed.
(b) Outcomes from risk assessment

Risk assessment studies provide estimated numbers of radiation-induced breast cancer cases and/or deaths, with a range of estimates according to variations in hypotheses. Estimation of prevented deaths based on assumptions about mortality reduction by screening modalities is performed in most studies, and calculation of benefit–risk is provided. Because the risk of radiation-induced cancer applies only to women who underwent mammography, hypotheses about mortality reduction should apply only to attendees; this is not always made explicit in publications. Thus, benefit–risk estimates provided by studies should be interpreted with caution.

(i) Risk assessment studies in the general population

Risk assessment studies performed in the early 2000s or earlier used risk models that are no longer recommended by international committees (Howe et al., 1981; Feig & Hendrick, 1997; Beemsterboer et al., 1998a; Mattsson et al., 2000; Law & Faulkner, 2001, 2002, 2006; León et al., 2001; Berrington de González & Reeves, 2005; Ramos et al., 2005). Since 2010, all studies have used the excess absolute risk model recommended by BEIR VII and contemporary estimates of mean glandular dose to the breast from either screen-film or digital mammography (Hendrick, 2010; O’Connor et al., 2010; de Gelder et al., 2011b; HPA, 2011; Yaffe & Mainprize, 2011; Hauge et al., 2014). These recent studies are now considered to be the most relevant and are summarized below (Table 5.13). In addition, one study used a biological model (Bijwaard et al., 2010, 2011).

(ii) Estimates for screening starting at about age 50 years

The Health Protection Agency estimated the number of cancer cases and cancer deaths after radiation exposure from a large number of sources, including screening mammography, in the United Kingdom population (HPA, 2011). The number of radiation-induced breast cancer cases after a single two-view screen every 3 years at age 47–73 years was estimated to be 28 per 100 000 women screened, and the number of breast cancer deaths under the same conditions was estimated to be 10 per 100 000 women screened. Assuming 500 prevented deaths from screenings, the authors estimated the net benefit (deaths prevented minus deaths induced) to be 490 [ratio of prevented to induced deaths of 50].

O’Connor et al. (2010) estimated the number of breast cancer cases induced by screen-film mammography, digital mammography, and other imaging techniques in a United States setting. They estimated that 21 cancer cases would be induced by digital mammography and 27 by screen-film mammography for annual screening per 100 000 women screened at age 50–80 years, and that there would be 6 or 7 induced deaths. Using different mortality reduction hypotheses, they estimated ratios of prevented to induced deaths of 116 and 135 for screen-film and digital mammography, respectively.

In Norway, Hauge et al. (2014) estimated the number of radiation-induced breast cancer cases after a single two-view digital mammography screening every 2 years from age 50 years to age 69 years to be 10 (range, 1.4–36) per 100 000 women screened, and the number of induced deaths per 100 000 women screened to be 1 (range, 0.1–3). Assuming a 40% mortality reduction among attendees, the authors estimated that 350 lives would be saved compared with 3 or fewer deaths induced [ratio of prevented to induced deaths of at least 117].

In the Netherlands, calculations were performed for a biennial digital mammography screening between the ages of 50 years and 74 years [12 screening sessions] (de Gelder et al., 2011b). The authors estimated 7.7 radiation-induced breast cancer cases (range, 5.9–29.6) and 1.6 radiation-induced breast cancer deaths.
Table 5.13 Risk assessment studies of breast cancer induced by mammography screeninga

<table>
<thead>
<tr>
<th>Reference Country</th>
<th>Mean glandular dose to the breast</th>
<th>Risk model</th>
<th>Target population, screening modalities</th>
<th>Lifetime calculation</th>
<th>Radiation-induced cases</th>
<th>Radiation-induced deaths</th>
<th>Benefit–risk: ratio of prevented to induced deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hendrick (2010)</td>
<td>3.7 mGy for 2-view DM 4.7 mGy for two-view SFM</td>
<td>EAR model from BEIR VII Modifying effect of age</td>
<td>Annual screening for 40–80 yr</td>
<td>NA</td>
<td>NA</td>
<td>20 (DM) and 25 (SFM) deaths</td>
<td>NA</td>
</tr>
<tr>
<td>O’Connor et al. (2010)</td>
<td>3.9 mGy for 2-view DM 4.9 mGy for 2-view SFM: inclusion of extra views</td>
<td>EAR model from BEIR VII Modifying effect of age Latency, 5 yr DDREF, 1.5</td>
<td>Annual screening for 40–80 yr and for 50–80 yr</td>
<td>Until 80 yr</td>
<td>Screening 40–80 yr: 56 (DM) and 71 (SFM) cases Screening 50–80 yr: 21 (DM) and 27 (SFM) cases Screening 40–49 yr: 35 (DM) and 44 (SFM) cases</td>
<td>Screening 40–80 yr: 15 (DM) and 19 (SFM) deaths Screening 50–80 yr: 6 (DM) and 7 (SFM) deaths Screening 40–49 yr: 9 (DM) and 11 (SFM) deaths</td>
<td>Assuming a mortality reduction of 15% from screening before age 60 yr and 32% after age 60 yr, ratio of prevented to induced deaths: Screening 40–80 yr: 44 (SFM) and 56 (DM) Screening 50–80 yr: 116 (SFM) and 135 (DM) Screening 40–49 yr: 3 (SFM and DM)</td>
</tr>
<tr>
<td>de Gelder et al. (2011b)</td>
<td>1.3 mGy per view (range, 1–5 mGy)</td>
<td>EAR model from BEIR VII Modifying effect of age No latency DDREF, 1.5</td>
<td>Screening for 40–74 yr or 50–74 yr Every 2 yr 2 views at first round 1 view at subsequent rounds</td>
<td>Until 100 yr</td>
<td>Screening 40–74 yr: 17.1 cases (range, 13.1–65.6) Screening 50–74 yr: 7.7 cases (range, 5.9–29.6)</td>
<td>Screening 40–74 yr: 3.7 deaths (range, 2.9–14.4) Screening 50–74 yr: 1.6 deaths (range, 1.3–6.3)</td>
<td>Assuming 26% mortality reduction, ratio of prevented to induced deaths: Screening 40–74 yr: 349 Screening 50–74 yr: 684 (range, 178–889)</td>
</tr>
<tr>
<td>HPA (2011)</td>
<td>4.5 mGy for 2-view screening</td>
<td>EAR model from Preston et al. (2007) (see Section 1.3.4) Modifying effect of age Latency, 10 yr</td>
<td>Screening for 40–73 yr Annually before 50 yr Every 3 yr after 50 yr</td>
<td>Until 85+ yr</td>
<td>Screening 40–47 yr: 61 cases Screening 47–73 yr: 28 cases</td>
<td>Screening 40–47 yr: 20 deaths Screening 47–73 yr: 10 deaths</td>
<td>Net benefit (deaths prevented minus deaths induced): 80 for age 40–47 yr; 490 for age 47–73 yr [ratio of prevented to induced deaths, 50]</td>
</tr>
<tr>
<td>Reference Country</td>
<td>Mean glandular dose to the breast</td>
<td>Risk model</td>
<td>Target population, screening modalities</td>
<td>Lifetime calculation</td>
<td>Radiation-induced cases</td>
<td>Radiation-induced deaths</td>
<td>Benefit–risk: ratio of prevented to induced deaths</td>
</tr>
<tr>
<td>-------------------</td>
<td>----------------------------------</td>
<td>------------</td>
<td>-----------------------------------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>--------------------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td><strong>Yaffe &amp; Mainprize (2011)</strong> Canada</td>
<td>3.7 mGy for 2-view DM</td>
<td>EAR model from BEIR VII Modifying effect of age Latency, 10 yr</td>
<td>Annual screening for 40–55 yr Every 2 yr for 55–74 yr</td>
<td>Until 109 yr</td>
<td>Screening 40–49 yr: 59 cases Screening 40–74 yr: 86 cases</td>
<td>Screening 40–49 yr: 7.6 deaths Screening 40–74 yr: 11 deaths</td>
<td>Assuming 24% mortality reduction, ratio of prevented to induced deaths: Screening 40–49 yr: 11.4 Screening 40–74 yr: 46</td>
</tr>
<tr>
<td><strong>Hauge et al. (2014)</strong> Norway</td>
<td>2.5 mGy for 2-view DM (range, 0.7–5.7 mGy)</td>
<td>EAR model from BEIR VII Modifying effect of age Latency, 5 or 10 yr DDREF, 1 or 2</td>
<td>Screening for 50–69 yr Every 2 yr</td>
<td>Until 85 or 105 yr</td>
<td>10 cases (range, 1.4–36)</td>
<td>1 death (range, 0.1–3.1)</td>
<td>Assuming 40% mortality reduction among attendees, 350 lives saved compared with 3 or fewer deaths induced [ratio of prevented to induced deaths, at least 117]</td>
</tr>
</tbody>
</table>

* Calculated values are per 100 000 women screened.

BEIR VII, Biologic Effects of Ionizing Radiation, Report VII (National Research Council, 2006); DDREF, dose and dose rate effectiveness factor; DM, digital mammography; EAR, excess absolute risk; NA, not available; SFM, screen-film mammography.
(range, 1.3–6.3) per 100,000 women screened, assuming a glandular dose of 1.3 mGy per view. Using a simulation model (MISCAN) to estimate deaths prevented due to screening, they estimated a ratio of prevented to induced deaths of 684. When a glandular dose of 5 mGy per view was assumed, the ratio decreased to 178 and the number of radiation-induced deaths increased to 6.3.

Bijwaard et al. (2010, 2011) performed a risk assessment using a mechanistic, biologically based model that assumes a two-stage mutation for carcinogenesis. With this approach, the authors estimated that for five mammography screenings of 2 mGy starting at age 50 years [biennial screening until age 60 years], 1.3 breast cancer cases would be induced per 100,000 women screened (Bijwaard et al., 2010), and 200 cases for 15 screenings of 4 mGy.

(iii) Estimates for screening starting at age 40 years

In the United Kingdom calculation (HPA, 2011), the number of radiation-induced breast cancer cases after annual two-view screening at ages 40–47 years was estimated to be 61 per 100,000 women screened. Using a hypothesis about survival, the authors estimated the number of radiation-induced breast cancer deaths after annual two-view screening at ages 40–47 years to be 20 per 100,000 women screened. Assuming 100 prevented deaths from screening, they estimated the net benefit (deaths prevented minus deaths induced) to be 80 [ratio of prevented to induced deaths of 5].

In the USA, Hendrick (2010) estimated the number of deaths induced by annual mammography per 100,000 women screened at age 40–80 years to be 20 for digital mammography and 25 for screen-film mammography. In the study of O’Connor et al. (2010), the authors estimated the number of breast cancers induced by annual mammography per 100,000 women screened at age 40–49 years to be 35 for digital mammography and 44 for screen-film mammography, and the number of radiation-induced breast cancer deaths to be 9 for digital mammography and 11 for screen-film mammography. According to a hypothesis about mortality reduction, they estimated a ratio of prevented to induced deaths of about 3 for both modalities.

In Canada, Yaffe & Mainprize (2011) estimated that mammography screening annually from age 40 years to age 55 years and biennially until age 74 years would induce 86 breast cancers cases (59 for the screening period 40–49 years) and 11 breast cancers deaths (7.6 for the screening period 40–49 years) per 100,000 women screened. Assuming a 24% reduction in mortality, they estimated a ratio of prevented to induced deaths of 46 for age 40–74 years (11.4 for age 40–49 years). The ratio of lives saved to lives lost is 78 for age 40–74 years (27 for age 40–49 years).

In the Netherlands, calculations were performed for biennial mammography screening between age 40 years and age 74 years; the authors estimated the number of breast cancer cases per 100,000 women screened to be 17.1 (range, 13.1–65.6) and the number of radiation-induced breast cancer deaths to be 3.7 (range, 2.9–14.4) (de Gelder et al., 2011a). Using a simulation model (MISCAN) to estimate deaths prevented due to screening, they estimated a ratio of prevented to induced deaths of 349. The study using a mechanistic model estimated 1.5 cases per 100,000 women screened for five mammography screenings of 2 mGy starting at age 40 years (Bijwaard et al., 2010).

(iv) Women at an increased risk

Among women at an increased risk of breast cancer, screening procedures are recommended earlier in life and at a higher frequency than in the general population (see Section 5.6). Due to the increased risk of radiation-induced breast cancer when exposure occurs at a younger age and because of the higher radiosensitivity of women
with a familial predisposition (see Section 1.3.6), separate risk assessment must be performed for women at an increased risk.

An excess relative risk model was used to estimate the lifetime risk of radiation-induced breast cancer mortality from five annual mammography screenings in young women harbouring a BRCA mutation (Berrington de González et al., 2009). They estimated the lifetime risk of radiation-induced breast cancer mortality per 10 000 women screened annually to be 26 (95% CI, 14–49) for screening at age 25–29 years, 20 (95% CI, 11–39) for screening at age 30–34 years, and 13 (95% CI, 7–23) for screening at age 35–39 years. [This calculation was based on model risk and coefficients estimated from the general population, and the higher sensitivity to radiation of these women was not taken into account.] A large European study among carriers of BRCA1/2 mutations suggested that exposure to diagnostic radiation before age 30 years for these women was associated with an increased risk of breast cancer at dose levels considerably lower than those at which increases had previously been found (Pijpe et al., 2012).

Benefit–risk estimates for women at an increased risk need to consider: the age-dependent higher risk of radiation in younger women and in women with specific gene mutations; their age-dependent overall measured breast cancer risk; and the contribution of mammography to early detection, which itself may depend on patient age, the type of genetic mutation (BRCA1 vs BRCA2), and the availability of magnetic resonance imaging (MRI).

5.3.5 Psychological consequences of mammography screening

Participation in breast cancer screening can have psychological or psychosocial consequences for women. Section 3.1.4 summarizes the psychological impacts of an invitation to screening, of a negative result, of a diagnosis of breast cancer, and of interval cancer, as well as the impact of a false-positive result on further participation. This section presents the studies reviewed for the evaluation of the psychological consequences of a false-positive result and of DCIS.

Several reviews have focused on the long-term psychological implications of a false-positive result (Rimer & Bluman, 1997; Steggles et al., 1998; Brodersen et al., 2004; Brett et al., 2005; Brewer et al., 2007; Hafslund & Nortvedt, 2009; Salz et al., 2010; Bond et al., 2013a, b). The two reviews by Bond et al. (2013a, b) evaluate the same set of studies, so one has been excluded. The review by Rimer & Bluman (1997) has also been excluded, due to its lack of relevance. In this section, the outcomes of the informative reviews (Table 5.14) and results from more recent individual studies are presented.

(a) Outcomes from reviews

Negative outcomes were reported from studies using validated measures during the period between receiving a recall letter and the recall appointment (Sutton et al., 1995; Chen et al., 1996; Lowe et al., 1999; Lampic et al., 2001; Sandin et al., 2002), at the recall appointment (Ellman et al., 1989; Cockburn et al., 1992; Swanson et al., 1996; Lowe et al., 1999; Ekeberg et al., 2001; Meystre-Agustoni et al., 2001), or immediately after receiving a recall letter (Cockburn et al., 1994; Lidbrink et al., 1995; Olsson et al., 1999; Lindfors et al., 2001).

The main psychological consequences of a false-positive result were psychological distress, somatization, depression, fear, anxiety, worry, an increase in women’s perceived likelihood of developing breast cancer, a decrease in the perceived benefits of mammography, and an increase in the frequency of breast self-examination (BSE) (Salz et al., 2010). [These outcomes may be contextualized as symptoms, but it is unclear how they would affect women in their everyday lives.]

Salz et al. (2010) performed a meta-analysis of the effect of false-positive mammograms on
Breast cancer screening

generic and specific psychosocial outcomes. From 17 studies presented in 21 articles, they found that across six generic outcomes, the only consistent effect was generalized anxiety (Ellman et al., 1989; Gram et al., 1990; Bull & Campbell, 1991; Lerman et al., 1991a, 1993; Cockburn et al., 1994; Ong et al., 1997; Scaf-Klomp et al., 1997; Brett et al., 1998; Pisano et al., 1998; Olsson et al., 1999; Aro et al., 2000; Lipkus et al., 2000; Brett & Austoker, 2001; Lampic et al., 2001, 2003; Meystre-Agustoni et al., 2001; Sandin et al., 2002; Barton et al., 2004; Jatoi et al., 2006; Tyndel et al., 2007).

(i) Short-term effects

All reviews concluded that there are short-term psychological consequences (up to 3 months) from having a recall. In one review (Brodersen et al., 2004), all 22 studies that investigated short-term consequences reported adverse short-term consequences. In a review based on 54 articles, Brett et al. (2005) concluded that the negative psychological impact was significantly higher for women who had a recall than for women who received a clear negative result after participation in mammography screening, although three studies reported no difference in the psychological impact of mammography screening between women who received a clear negative result and those who had a false-positive result (Bull & Campbell, 1991; Lightfoot et al., 1994; Aro et al., 2000). Other negative consequences reported in women who had a false-positive result were more intrusive thoughts, worry about breast cancer, greater requirements for social support, being more busy than usual to keep their thoughts away from the clinical visit, or difficulties sleeping (Bull & Campbell, 1991; Lightfoot et al., 1994; Scaf-Klomp et al., 1997; Gilbert et al., 1998; Aro et al., 2000). Two studies reported that 30% (Austoker & Ong, 1994) and 40% (Scaf-Klomp et al., 1997) of women felt very anxious when they received a recall letter. One study that looked at how having a false-positive result influences quality of life found a marked decrease in quality of life for recalled women (Lowe et al., 1999).

(ii) Long-term effects

Based on the available reviews, results about long-term consequences are more ambiguous and inconsistent (Brodersen et al., 2004; Brett et al., 2005; Brewer et al., 2007). Several studies did not find increases in long-term levels of anxiety among women who had a false-positive result (Gram et al., 1990; Cockburn et al., 1994; Lidbrink et al., 1995; Gilbert et al., 1998; Lowe et al., 1999; Ekeberg et al., 2001; Lampic et al., 2001; Sandin et al., 2002), and two studies were inconclusive (Scaf-Klomp et al., 1997; Aro et al., 2000). Other studies reported that the anxiety experienced was greater among women who had a false-positive result than among women who received a clear negative result, at 4–6 months after recall (Ellman et al., 1989; Brett et al., 1998; Olsson et al., 1999; Lampic et al., 2001), 6–12 months after recall (Lampic et al., 2001; Hislop et al., 2002), and 24 months after recall (Lipkus et al., 2000). One review found no long-term symptoms of depression among women who received a false-positive result (Brewer et al., 2007).

(iii) Breast cancer-specific measures

One review investigated the effects on healthcare use and symptoms (Brewer et al., 2007). The findings suggested that having a false-positive result increases anxiety related to breast cancer specific measures (Brewer et al., 2007). Three studies found that women who received a false-positive result reported conducting BSE statistically significantly more frequently (Bull & Campbell, 1991; Aro et al., 2000; Lampic et al., 2001). Women who had a false-positive result also reported higher levels of worry and increased concern about breast cancer (Lerman et al., 1991a, b; Scaf-Klomp et al., 1997; Brett et al., 1998; Aro et al., 2000; Lipkus et al., 2000; Sandin et al., 2002; Absetz et al., 2003). In their
meta-analysis, Salz et al. (2010) found statistically significant effects on all eight breast cancer-specific outcomes: distress about breast cancer, somatization or symptoms in the breast, fear of developing breast cancer, anxiety about breast cancer, worry about breast cancer, perceived likelihood of breast cancer, perceived benefits of mammography, and frequency of BSE. The largest effect was for anxiety about breast cancer ($r = 0.22$) and the smallest was for fear ($r = 0.08$); all eight pooled effect sizes were statistically significant.

**(iv) Screening factors**

Screening factors associated with greater adverse psychological effects were: previous false-positive results (Brett & Austoker, 2001; Haas et al., 2001; Lampic et al., 2001), pain at previous mammography screening (Ong & Austoker, 1997; Drossaert et al., 2002), dissatisfaction with information and communication during screening (Austoker & Ong, 1994; Brett et al., 1998; Brett & Austoker, 2001; Dolan et al., 2001), and waiting time between recall letter and assessment appointment (Gram et al., 1990; Thorne et al., 1999; Brett & Austoker, 2001; Lindfors et al., 2001).

Elements of the structure of the screening programme were also found to be important. The extent of further investigation seemed to determine the extent of negative psychological outcomes. Women who underwent a surgical biopsy before receiving a clear result experienced the greatest anxiety (Ellman et al., 1989; Lerman et al., 1991b; Ong & Austoker, 1997; Brett et al., 1998; Lampic et al., 2001), as did those asked to come back for further tests after 6 months or 1 year (Ong et al., 1997; Brett et al., 1998; Brett & Austoker, 2001). On-site evaluation was shown to reduce the stress of having a false-positive result (Lindfors et al., 2001). Biopsy-specific events appeared to be more distressing than follow-up mammography, and distress risk factors included younger age, less education, and no family history of breast cancer (Steffens et al., 2011).

Reported sociodemographic factors often associated with greater adverse psychological outcomes were younger age, less education, living in an urban area, having one child or no children, and manual occupation (Brett et al., 2005). Other studies found no impact of age (Brett et al., 1998; Brett & Austoker, 2001; Lampic et al., 2001) or employment (Olsson et al., 1999). One study with 910 participants in California, USA, found that Asian ethnicity, annual income greater than US$ 10 000, and weekly attendance of religious services were significantly associated with decreased depressive symptoms (Alderete et al., 2006).

**(b) Recent individual studies**

More recent studies, not included in the reviews, have used the Hospital Anxiety and Depression Scale, the Psychological Consequences Questionnaire, and the Consequences of Screening in Breast Cancer questionnaire to study psychological consequences of mammography screening (Table 5.14). Consistent with findings from a study conducted in 1996–1997 (Ekeberg et al., 2001), Schou Bredal et al. (2013) found that recall after mammography among women with a false-positive result was associated with transiently increased anxiety and a slight increase in depression. At 4 weeks after screening, the level of anxiety was the same and depression was lower compared with the general female Norwegian population (Schou Bredal et al., 2013).

In a study in Spain, participants were found to worry little until they underwent mammography, but levels of worry increased when the women were notified by telephone call of the need for further testing (Espasa et al., 2012). A substantial proportion of women requiring further assessment reported that they were at least somewhat worried about having breast cancer throughout the screening process, but
levels of anxiety and depression, measured by the Hospital Anxiety and Depression Scale, showed no statistically significant differences among women who had invasive complementary tests, non-invasive tests, and negative screening results (Espasa et al., 2012).

In a longitudinal study in Denmark, psychological effects of false-positive results were assessed with the Consequences of Screening in Breast Cancer questionnaire. At 6 months after the final diagnosis, women with a false-positive finding reported changes in existential values and inner calmness as great as those reported by women with a diagnosis of breast cancer; 3 years after the final diagnosis, women who had a false-positive result consistently reported greater negative psychosocial consequences in all 12 psychosocial outcomes compared with women who had a normal finding (Brodersen & Siersma, 2013). However, after 5 years, there was no statistically significant difference between the two groups in reported psychosocial aspects (Ostero et al., 2014).

When women who were first-time participants in mammography screening were compared with women with repeated screening experience, women in both groups reported experiencing high levels of anxiety before the diagnosis was known, and no differences were found in anxiety, depressive symptoms, or quality of life (Keyzer-Dekker et al., 2012).

In a study in 98 women, women reported a significant increase in anxiety after being notified of the need to return for follow-up testing, and significant positive associations were found between anxiety and behavioural testing, behavioural avoidance, cognitive approach, and cognitive avoidance coping in cross-sectional analyses (Heckman et al., 2004). Moreover, cognitive avoidance coping was a strong predictor of final levels of state anxiety in these women (Heckman et al., 2004).

These findings are consistent with qualitative studies in Scandinavia and North America. Norwegian women expressed mixed emotions over being recalled; information about recall rates and breast cancer risk was seen as alarming, and the short time between recall and examination was seen as reassuring but was also perceived as an indication of malignancy (Solbjør et al., 2011). Swedish women who were recalled described the recall process as “a roller coaster of emotions” (Bolejko et al., 2013). Qualitative studies from North America have described the psychological effects of the waiting process experienced by women, their unmet informational and psychosocial needs (Doré et al., 2013), anxieties generated by waiting and wondering, and fears of iatrogenic effects of follow-up tests such as...
biopsies and repeat mammograms (Padgett et al., 2001).

(c) Diagnosis of ductal carcinoma in situ

Psychological consequences of DCIS are included in this section because increasing participation in mammography screening means an increasing number of DCIS detections among women, but the effect of DCIS on psychological issues has been little explored. Women may not be aware of having DCIS, because surgeons might differ in how they inform women about this condition. Potentially, some women with DCIS are informed that they have breast cancer while others are informed that they do not have breast cancer. A study with semi-structured interviews of women previously diagnosed with and treated for DCIS identified six key themes: (i) invisibility of DCIS, (ii) uncertainty, (iii) perceptions of DCIS, (iv) acceptance of treatment, (v) social support, and (vi) moving on, which highlight the substantial challenges faced by women diagnosed with DCIS (Kennedy et al., 2008).

No articles focused on non-invasive breast cancer or DCIS before 1997 (Webb & Koch, 1997). A review of quality-of-life issues among patients with DCIS (Ganz, 2010) found that women with DCIS experience psychological consequences to a lesser extent than women with breast cancer, but few studies have compared these women with healthy women. Of greater concern, women with DCIS demonstrate severe misconceptions about their risk of invasive breast cancer (Ganz, 2010).

One study of 10 women with DCIS found that they would have liked to have received more information about DCIS when they were invited to routine screening (Prinjha et al., 2006). In another study, 45 women took part in an initial interview after a diagnosis of DCIS, and 27 took part in a follow-up interview 9–13 months later (Kennedy et al., 2012). Women’s early perceptions of DCIS merged with and sometimes conflicted with their beliefs about breast cancer, and their perceptions and experiences of the condition shifted over time.

A study in Australia also found misunderstanding and confusion among women diagnosed with DCIS and a desire for more information about their breast disease (De Morgan et al., 2011). Approximately half of the participants worried about their breast disease metastasizing, approximately half expressed high decisional conflict, 12% were anxious, and 2% were depressed. Logistic regression analysis demonstrated that worry about dying from the breast disease was significantly associated with not knowing that DCIS could not metastasize (De Morgan et al., 2011). In five focus group interviews involving 26 women diagnosed with DCIS, women were confused about whether or not they had cancer that could result in death, and this confusion was compounded by the use of the term “carcinoma” and by the recommendation of treatments such as mastectomy (De Morgan et al., 2002).

In a study of 487 women who were newly diagnosed with DCIS, financial status was inversely associated with anxiety and depression at the 9-month follow-up, and women with medium or low socioeconomic status were vulnerable to escalating anxiety and depression after a DCIS diagnosis (de Moor et al., 2010). A study in the USA of approximately 800 Latina and Euro-American women with DCIS found that younger age, not having a partner, and lower income were related to lower quality of life in various domains (Bloom et al., 2013).

5.4 Cost–effectiveness and balance of harms and benefits

Decisions about implementation of healthcare interventions are based primarily on benefits and a favourable harm–benefit ratio, but – to use limited resources efficiently – are also often based on cost–effectiveness analyses. A cost–effectiveness analysis compares different policies,
Breast cancer screening

including the current one, with no intervention (average cost–effectiveness) or compares a more-intensive programme with a less-intensive one (incremental cost–effectiveness). Often, the incremental cost–effectiveness ratio (ICER) is estimated for each policy, expressed as the ratio of the change in costs to the change in effects compared with a less-intensive alternative or the current policy. In a cost–effectiveness analysis, future costs and effects are taken into account and both are discounted at a chosen annual discount rate, to account for time preference. A new strategy is considered cost-effective if it results in an additional effect (compared with a baseline) at acceptable additional costs (or even savings). One should stress the fact that the change in effects is as important as, and in the practice of policy-making even more important than, the change in costs: how much will the population benefit from the resources invested? Effects are often defined as disease-specific deaths prevented and life years gained but are ideally adjusted for quality of life, resulting in quality-adjusted life years (QALYs) (Weinstein & Stason, 1977). For breast cancer screening, factors that could negatively affect quality of life are, among others, the screening examination, false-positive referrals, earlier and often more intensive treatment, over-diagnosis, and simply the earlier knowledge of cancer (Korfage et al., 2006). All such harms are included when adjusting the life years gained for negative quality-of-life effects. Positive side-effects, such as a reduced need for expensive palliative treatments because fewer women are dying of breast cancer, can and should also be incorporated into such cost–effectiveness analyses.

To determine whether an intervention produces reasonable amounts of benefits and limited harms for the resources invested, the cost–effectiveness ratios are usually compared with cost–effectiveness thresholds. A frequently used cost–effectiveness threshold is £30 000 per QALY gained (NICE, 2014). In the USA, interventions below the threshold of US$ 50 000 per QALY are generally considered cost-effective, interventions between US$ 50 000 per QALY and US$ 100 000 per QALY are considered moderately or borderline cost-effective, and those that exceed US$ 100 000 per QALY are generally not considered cost-effective (Grosse, 2008). It has recently been recommended that a threshold of US$ 200 000 per QALY should be used for the USA (Neumann et al., 2014). The relatively high threshold of US$ 200 000 per QALY relates to the fact that health-care costs in the USA are generally considerably higher than those in Europe. Looking more globally, the World Health Organization (WHO) has suggested a cost–effectiveness threshold of 3 times the national gross domestic product per capita (WHO, 2014). Practically, for low-income regions the maximal values for being cost-effective are about US$ 5000 (WHO, 2001). [A clear distinction has to be made for cost-efficacy estimates of trials, which often relate to the limited time frame of an RCT, in which not all benefits have accrued yet but where it is likely that cost and harms have already been prominent.]

Costs that should be considered in a cost–effectiveness analysis of breast cancer screening are costs associated with the organization of the programme (e.g. cost of invitations, screening costs), costs related to the diagnostic workup of both true-positives and false-positives, and additional treatment costs (e.g. due to more and earlier treatments). A few years after implementation, screening will lead to cost savings in treatment due to a decrease in the number of cases of advanced disease needing treatment (de Koning et al., 1992). The cost savings depend mostly on the cost for advanced disease and the magnitude of the effectiveness of the screening programme. In a full cost–effectiveness analysis, direct medical costs, direct non-medical costs (travel and time), and indirect costs (e.g. due to sick leave) must be considered.

Ideally, all possible screening policies that are relevant are compared in a cost–effectiveness
analysis. However, it is not feasible to compare all scenarios of interest in an RCT or an observational study. In addition, trials deliver (at best) costs per case detected. This is not an appropriate measure for cost–effectiveness because it lacks information about the effectiveness of screening (in terms of life years gained or breast cancer deaths averted). Furthermore, the aim of a cost–effectiveness analysis on breast cancer screening is to assess the effectiveness of a screening programme in an actual population rather than in a controlled setting. By the use of mathematical models, findings from RCTs and observational studies can be extrapolated to simulated populations (Berry et al., 2005). Models are especially suitable for a cost–effectiveness analysis because the key elements of screening, including the screening strategy (starting age, stopping age, and screening interval), the target population (e.g. at average or increased risk), and the time point of the analysis, can be altered and/or compared. Furthermore, long-term lifetime effects can be predicted, and life years gained or QALYs can be calculated (Groenewoud et al., 2007) (see Section 5.1.2f for further details).

5.4.1 Mammography screening programmes in developed countries

Under the assumption that mammography screening programmes are effective in reducing breast cancer mortality in women at average risk of breast cancer, numerous cost–effectiveness analyses have shown that organized mammography screening can be cost-effective (van Ineveld et al., 1993; Leivo et al., 1999; Stout et al., 2006; Groenewoud et al., 2007; Carles et al., 2011; Pataky et al., 2014).

Most population-based screening programmes screen women at biennial intervals (Giordano et al., 2012). Annual screening strategies may improve the detection of rapidly growing tumours. However, despite the greater effectiveness, screening strategies that consist of annual screening are often found to be less efficient and less cost-effective, due to a disproportionate increase in costs or due to diminishing returns; about 80% of the effect of annual screening is retained when screening is performed every 2 years (Mandelblatt et al., 2009; Stout et al., 2014). Schousboe et al. (2011) demonstrated that, in the United States setting, even if annual mammography is restricted to certain risk groups, based on age or breast density, the costs exceed US$ 100 000 per QALY gained. In contrast, Carles et al. (2011) reported several cost-effective annual screening strategies in Spain. However, ICERs increased markedly when comparing annual screening with biennial screening, as reported in other studies.

Organized mammography screening has been shown to be more cost-effective than opportunistic mammography screening (Bulliard et al., 2009; de Gelder et al., 2009). In Switzerland, the costs per life year gained of opportunistic screening were twice those of organized screening (de Gelder et al., 2009). This difference was caused predominantly by the higher costs of mammography for opportunistic screening and the more frequent use of additional imaging in combination with opportunistic screening.

Cost–effectiveness ratios obtained from studies of screening programmes in different countries are not easily comparable, due to differences in assumptions about effects and costs, time horizon, discount rate, and calculation methods (Brown & Fintor, 1993; de Koning, 2000). Furthermore, epidemiological factors (background risk of breast cancer), the performance of the screening test, and the organization of the national screening programme and the health-care system all influence cost–effectiveness. The cost–effectiveness of a screening programme also depends on its characteristics, including attendance rate, screening interval, and age group targeted for screening.
5.4.2 Screening in low- and middle-income countries

A detailed cost–effectiveness analysis on breast cancer screening in India has been reported, in which the natural history of breast cancer was calibrated against available data on breast cancer incidence, stage distribution, and mortality in India (Okonkwo et al., 2008). The model was used to estimate the costs of breast cancer screening in India, its effects on mortality, and its cost–effectiveness (i.e. costs of screening per life year gained or per life saved). Screening using CBE or mammography among different age groups and at various frequencies was analysed. Stage-dependent sensitivities of CBE in this study were based on data from the Canadian National Breast Screening Study (CNBSS) (Rijnsburger et al., 2004). Alternative (lower) estimates of stage-dependent sensitivities of CBE were based on data from 752,000 CBEs delivered to low-income women in the USA in 1995–1998 through the National Breast and Cervical Cancer Early Detection Program of the United States Centers for Disease Control and Prevention (Bobo et al., 2000).

Okonkwo et al. (2008) expressed costs in international dollars (Int.$), the currency used by WHO; an international dollar has the same purchasing power in a particular country as a United States dollar has in the USA. Under the assumption that such screening programmes are as effective as is seen in mammography trials, the estimated mortality reduction was the greatest for programmes targeting women between age 40 years and age 60 years. Using a 3% discount rate, a single CBE at age 50 years had an estimated cost–effectiveness ratio of Int.$ 793 per life year gained and resulted in a reduction in breast cancer mortality of 2%. The cost–effectiveness ratio increased to Int.$ 1135 per life year gained for every 5-yearly CBE (age 40–60 years) and to Int.$ 1341 for biennial CBE (age 40–60 years); the corresponding reductions in breast cancer mortality were 8.2% and 16.3%, respectively. CBE performed annually from age 40 years to age 60 years was predicted to be nearly as efficacious as biennial mammography screening for reducing breast cancer mortality, while incurring only half the net costs.

The main factors affecting cost–effectiveness were breast cancer incidence, stage distribution, and cost savings on palliative care averted (Okonkwo et al., 2008). The estimated cost–effectiveness of CBE screening for breast cancer in India compares favourably with that of mammography in developed countries. [The study relied on an assumption about the efficacy of CBE in reducing breast cancer mortality in India, which has not been verified in randomized trials comparing CBE with no screening but was based on the CNBSS 2 trial, assuming that the effect of stage shift from mammography trials can be extrapolated.]

More recently, several studies have investigated the expected cost–effectiveness of different strategies in Costa Rica and Mexico (Niëns et al., 2014), Ghana (Zelle et al., 2012), and Peru (Zelle et al., 2013). In Costa Rica, the current strategy of treating breast cancer at stages I to IV at a geographical coverage level of 80% seems to be the most cost-effective, with an ICER of US$ 4739 per disability-adjusted life year (DALY) averted. At a coverage level of 95%, biennial CBE screening could double life years gained and can still be considered very cost-effective (ICER, US$ 5964 per DALY averted). For Mexico, the results indicate that at a coverage level of 95%, a mass media awareness-raising programme could be the most cost-effective (ICER, US$ 5021 per DALY averted). If more resources are available in Mexico, biennial mammography screening for women aged 50–70 years (ICER, US$ 12,718 per DALY averted), adding trastuzumab (ICER, US$ 13,994 per DALY averted), or screening women aged 40–70 years biennially plus trastuzumab (ICER, US$ 17,115 per DALY averted) are less cost-effective options (Niëns et al., 2014).
Breast cancer in Ghana is characterized by low awareness, late-stage treatment, and poor survival. Biennial screening with CBE of women aged 40–69 years, in combination with treatment of all stages, seems the most cost-effective intervention (ICER, US$ 1299 per DALY averted). Mass media awareness-raising is the second-best option (ICER, US$ 1364 per DALY averted) (Zelle et al., 2013). The current breast cancer programme in Peru (US$ 8426 per DALY averted) could be improved by implementing 3-yearly or biennial screening strategies. These strategies seem the most cost-effective in Peru, particularly when mobile mammography is applied (from US$ 4125 per DALY averted) or when CBE screening and mammography screening are combined (from US$ 4239 per DALY averted).

The impact of the various screening interventions on stage distribution was estimated on the basis of a model using proportional detection rates (Duffy & Gabe, 2005). The authors applied a stage shift from developing countries to the Dutch screening programme and corrected this shift for locally relevant attendance rates and the epidemiology and demography. The age-specific sensitivity of tests and the sojourn times (CBE sojourn times are two thirds those of mammography) were based on the literature (Duffy & Gabe, 2005; NETB, 2014). The effectiveness of the awareness-raising interventions is based on a study in Malaysia (Devi et al., 2007), where a 2-fold reduction in advanced breast cancer was observed when a mass media campaign was applied. However, evidence on the effectiveness of awareness-raising, CBE, and mammography screening is absent in many countries. Also, these programmes require substantial organizational, budgetary, and human resources, and the accessibility of diagnostic, referral, treatment, and palliative care facilities for breast cancer should simultaneously be improved.

5.4.3 Harm–benefit ratio and generalizability

As already pointed out, the expected effects – both benefits and harms – and the cost of an intervention are context-specific. In public health, medicine, and any other field, inferences and extrapolations to other populations and individuals are needed. The average estimates for relative benefits, observed in IBM, nested case–control cohort, and case–control studies, in which biases have been minimized as much as possible, need to be extrapolated, as well as the estimates for overdiagnosis, false-positives, and radiation risk. To incorporate all of these and to estimate values as specifically as possible for different populations with different age structures, life expectancies, incidence, mortality, and treatment levels, statistical models are used.

The harm–benefit ratio has been calculated for different settings. The Independent United Kingdom Panel estimated that the United Kingdom screening programmes currently prevent 1300 deaths from breast cancer per year, equivalent to about 22 000 years of life being saved. Per 10 000 women invited to screening, it is estimated that 43 deaths from breast cancer are prevented and 129 cases of breast cancer represent overdiagnosis (Marmot et al., 2013). The Euroscreen Working Group estimated that for every 10 000 women screened biennially from age 50 or 51 years until age 68 or 69 years, about 80 deaths from breast cancer are prevented, versus about 40 cases overdiagnosed (Paci & EUROSCREEN Working Group, 2012). In the Netherlands, it has been estimated that each year 775 breast cancer deaths are prevented, versus 300 overdiagnosed cases (1 million invitations per year) (NETB, 2014).

5.4.4 Lower age limit for screening

Women younger than 50 years may benefit less from mammography screening, due to a lower breast cancer incidence, a lower
sensitivity of mammography due to denser breast tissue, a lower PPV, higher false-positive rates, and possibly more aggressive tumour growth (Carney et al., 2003; Buist et al., 2004). Therefore, the cost–effectiveness ratio is less favourable for younger women than for older women. For instance, a recent analysis showed that for Canada the most cost-effective strategies were biennial screening from age 50 years to age 69 years (ICER, US$ 28 921 per QALY), followed by biennial screening from age 40 years to age 69 years (ICER, US$ 86 029 per QALY) (Pataky et al., 2014).

In addition, the efficacy or effectiveness of screening, in terms of breast cancer mortality reduction, in women screened from age 40 years (Alexander et al., 1999; Smith et al., 2004; Moss et al., 2006; Hellquist et al., 2011) is less precisely estimated, due to small numbers of breast cancer deaths, than that in women screened from age 50 years, and may therefore be underestimated or overestimated in cost–effectiveness analyses. It could even be more cost-effective to screen women aged 50–69 years more frequently than to include women younger than 50 years (de Koning et al., 1991).

A study in which the Dutch MISCAN model was used to assess the cost–effectiveness of different policies for breast cancer screening in Catalonia, Spain (using Dutch data on costs) demonstrated that it is comparably cost-effective to extend screening from age 50 years to age 45 years and to extend screening from age 64 years to age 69 years (Beemsterboer et al., 1998b). The researchers emphasized that extending the upper age limit would result in a greater reduction in breast cancer mortality, whereas extending screening to younger women could lead to more life years gained. A more recently performed cost–effectiveness analysis, also focusing on screening in Catalonia, showed that biennial screening from age 45 years (to age 69 years or 74 years), annual screening from age 40 years (to age 69 years or 74 years), and annual screening from age 45 years (to age 69 years) (ranked in order of effectiveness) are all cost-effective strategies, with incremental costs per QALY gained of less than €30 000 (Carles et al., 2011).

A study based on data from the USA demonstrated that biennial mammography screening from age 40 years to age 49 years is cost-effective only for women with BI-RADS 3 or 4 breast density, women with both a previous breast biopsy and a family history of breast cancer, and women with BI-RADS 3 or 4 breast density and either a previous breast biopsy or a family history of breast cancer, assuming a cost–effectiveness threshold of US$ 100 000 per QALY gained (Schousboe et al., 2011). In contrast, another study, using five independent models of digital mammography screening in the USA, found that extending biennial screening from women aged 50–74 years to those aged 40–49 years would lead to incremental costs of US$ 55 100 per QALY gained, which was considered to be cost-effective (Stout et al., 2014). Annual mammography, which may improve detection of rapidly growing tumours that may be more common among younger women, was considered not cost-effective in both studies. As mentioned previously, age considerations may be different for developing countries.

5.4.5 Upper age limit for screening

Breast cancer incidence and breast cancer detection rates are higher in women aged 70 years and older, which may increase the effect of screening. However, compared with younger women, older women are more subject to numerous illnesses and conditions that negatively affect life expectancy, thereby limiting the beneficial effect of screening on life expectancy and potentially increasing costs of screening. Furthermore, attendance rates may be lower among older women, which would also negatively affect the cost–effectiveness ratio.
Women older than 74 years were not included in any breast cancer screening trial (see Section 4.2). Model simulations demonstrated that screening women aged 50–75 years and screening women with high bone mineral density up to age 79 years are both cost-effective strategies (Boer et al., 1995; Kerlikowske et al., 1999). Correspondingly, two systematic reviews showed that ceasing screening at age 75 years or 79 years instead of at age 65 years or 69 years is cost-effective, even for women who are not screened regularly before age 65 years (Barratt et al., 2002; Mandelblatt et al., 2003).

5.4.6 Digital mammography

In several countries, digital mammography has practically replaced film mammography (NHS, 2005; NETB, 2014). The sensitivity of digital mammography may be higher than that of film mammography for women younger than 50 years and for women with dense breasts (Pisano et al., 2008). However, the specificity of digital mammography may be slightly lower than that of film mammography (Skaane, 2009; Kerlikowske et al., 2011). Referral rates are likely to increase with digital mammography, depending on the baseline situation of referrals, but this is especially pertinent in the implementation phase. Because of the differences in test characteristics and in costs of mammography, cost-effectiveness ratios are likely to differ as well. A modelling study that used data from the DMIST trial found that, compared with film mammography, digital mammography is not cost-effective (US$ 331 000 per QALY gained), except when limited to women aged 40–49 years (Tosteson et al., 2008). However, digital mammography targeted to younger ages combined with film mammography from age 50 years is usually not a feasible strategy because film mammography has practically been replaced by digital mammography. Another study showed that digital mammography increases the number of false-positive findings by 220 per 1000 women compared with film mammography, leading to additional costs of US$ 350 000 per 1000 women, whereas the gain in benefits relative to film mammography is small (Stout et al., 2014).

5.4.7 Impact of individual risk factors

In most countries, organized mammography screening applies to all women in a targeted age group (usually 50–69 years or 50–74 years) with a relatively low (average) risk of breast cancer. Because breast cancer risk is associated with risk factors including age, reproductive history, a previous breast biopsy, and a family history of breast cancer (see Section 1.3), costs and benefits of screening may be affected by a woman’s individual risk of breast cancer. More personalized mammography screening, by selecting the starting and stopping ages and the screening interval based on a woman’s breast cancer risk profile, is therefore being considered in several research projects.

A cost-effectiveness study based on data from women in the USA showed that biennial mammography from age 40 years is cost-effective for women with high breast density (BI-RADS 3 or 4) and either a family history of breast cancer or a previous breast biopsy (< US$ 50 000 per QALY gained), and moderately cost-effective for women with high breast density only or both a previous breast biopsy and a family history of breast cancer (< US$ 100 000 per QALY gained) (Schousboe et al., 2011). Annual mammography was estimated to cost more than US$ 100 000 per QALY gained for any group at an increased risk, and was therefore not considered cost-effective. Another study based on population data from the USA, using five independent models, showed that annual digital mammography screening for women aged 40–74 years with high breast density (BI-RADS 3 or 4) resulted in 3-fold higher incremental costs per additional QALY gained relative to biennial screening for
all women aged 40–74 years (Stout et al., 2014). The incremental benefits of annually screening women aged 40–49 years with (extremely) dense breasts were small, predominantly accounting for the increase in ICERs.

Women with heterogeneously or extremely dense breasts and a negative screening mammogram may be considered for supplemental screening. The most readily available supplemental screening modality is ultrasonography, but little is known about its effectiveness when performed after negative screening mammography (see Section 5.5.1a). Sprague et al. (2015) used three independent simulation models to assess the lifetime benefits, harms, and cost–effectiveness from the payer perspective of supplemental ultrasonography screening for women with dense breasts compared with screening with digital mammography alone. They found that supplemental ultrasonography screening for women with dense breasts undergoing routine digital mammography screening would substantially increase costs while producing relatively small benefits in breast cancer deaths averted and QALYs gained. The cost–effectiveness ratio was US$ 325 000 per QALY gained (range, US$ 112 000–766 000). Restricting supplemental ultrasonography screening to women with extremely dense breasts would cost US$ 246 000 per QALY gained (range, US$ 74 000–535 000) relative to biennial mammography alone for women aged 50–74 years.

5.4.8 Quality of life

A Dutch analysis of cost–effectiveness and quality of life conducted in 1991 included estimates on 15 phases induced and/or prevented by the screening programme (de Koning et al., 1991). It appeared that 85% of the decrements in quality of life due to screening were due to the additional years in follow-up after diagnosis (of which about half were due to earlier detection and about half due to life years gained). False-positives comprised only a small component, as did the initial years of overdiagnosed cases. However, about 66% of the decrements were counterbalanced by gains; 70% of these gains imply reductions in palliative treatments for women with advanced disease. It was estimated that correcting the life years gained for quality of life would imply a 3% difference, that is, 3% fewer life years gained when adjusted for quality of life. The most unfavourable sensitivity analysis estimated a 19.7% decrease.

Vilaprinyo et al. (2014) estimated QALYs for the different breast cancer disease states. They used the health-related quality of life measures obtained from the EuroQol EQ-5D self-classifier in the study of Lidgren et al. (2007), which provided health-related quality of life measures for the first year after primary breast cancer (EQ-5D = 0.696), the second and following years after primary breast cancer or recurrence (EQ-5D = 0.779), and the metastatic breast cancer state (EQ-5D = 0.685). For false-positive mammograms, the authors assumed an average annualized loss of quality of life of 0.013. To obtain the value of 0.013, they assumed that 50% of women with a false-positive result would experience anxiety sufficient to increase the mood subscale of the EuroQol instrument from 0 to 1, lasting a total of 2 months. According to the United States EQ-5D tariffs, such a change for an entire year represents a decrease in the QALY value of 0.156. In the sensitivity analysis, the authors assessed the impact of changing the disutility by false-positives to 0 and to 0.026.

5.5 Other imaging techniques

This section reports evidence on the efficacy or effectiveness of imaging modalities other than screen-film mammography or standard digital mammography, where applied for population screening of asymptomatic women of about average (population) risk. Studies that included women at above average risk were considered, but
not those in which study subjects were restricted to classifications of increased risk. Studies of cohorts of women defined by dense breast tissue on mammography (but not restricted to women at an increased risk) were also reviewed.

The following imaging technologies were reviewed: breast ultrasonography, digital breast tomosynthesis, MRI (other than screening of women at increased risk), electrical impedance technology for breast imaging, scintimammography, and positron emission mammography. No RCTs examining the efficacy of these imaging technologies for population breast screening were available to the Working Group.

For two imaging technologies (ultrasonography in dense breasts and digital breast tomosynthesis in population screening), there was evidence from non-randomized studies of incremental (additional) cancer detection when applied as adjunct screening to mammography. The evidence for the preventive effects, adverse effects, and cost-effectiveness of these two technologies is presented in Sections 5.5.1, 5.5.2, and 5.5.3, respectively. Other imaging technologies, for which there was very little or no data on efficacy or effectiveness, or for which population screening studies have not been conducted, are briefly outlined in Section 5.5.4.

5.5.1 Preventive effects

(a) Breast ultrasonography

Ultrasonography has had a role in diagnosis of breast disease for approximately 30 years and has been used for the workup of screen-detected abnormalities and for image-guided needle biopsy (see Section 2.2.1 for technical details). Because dense breast tissue is a risk factor for breast cancer (McCormack & dos Santos Silva, 2006) and reduces the sensitivity of mammography, and hence is associated with a greater likelihood of an interval cancer in mammography screening (Ciatto et al., 2004a), evaluations of breast ultrasonography screening have often focused on populations defined by mammographic density (Buchberger et al., 2000; Houssami et al., 2009; Corsetti et al., 2011; Houssami & Ciatto, 2011; Venturini et al., 2013).

No RCTs examining the efficacy of screening by ultrasonography or of adjunct ultrasonography in women with dense breast tissue on mammography (i.e. mammography alone vs mammography plus ultrasonography) were identified by the Working Group. A recent Cochrane systematic review (Gartlehner et al., 2013) evaluated the literature to assess the effectiveness of ultrasonography screening as adjunct to mammography in women at average risk of breast cancer. None of the studies identified (no randomized, prospective, or controlled studies) reported sound evidence supporting ultrasonography as adjunct to mammography in population breast screening. An RCT on the efficacy of adjunct ultrasonography for breast cancer screening, called the Japan Strategic Anti-Cancer Randomized Trial, was noted (Ishida et al., 2014). This trial aimed to recruit 100 000 women aged 40–49 years and has recently closed to recruitment; its results have not yet been reported.

Several studies of breast ultrasonography screening, all non-randomized and without a comparison or control group, have examined the incremental cancer detection of breast ultrasonography in women with dense breast tissue and negative mammography. Table 5.15 presents the studies that have reported data for both true-positive detection and false-positives (or additional recall) attributed to ultrasonography screening. Studies that recruited women with dense breast tissue conditional to also being classified as at an increased risk were not considered (e.g. Berg et al., 2008). However, studies that defined subjects on the basis of dense breast tissue but also included some women or subgroups with additional risk factors were included and reviewed.

The majority of the studies were retrospective, and all were designed to assess incremental
Table 5.15 Studies of adjunct ultrasonography in screening asymptomatic women with mammography-negative dense breast tissue

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study characteristics; no. screened with US; age</th>
<th>Breast density</th>
<th>Preventive or screening effect</th>
<th>Adverse effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study characteristics; no. screened with US; age</td>
<td></td>
<td>Characteristics of US-only detected cancers (vs cancers detected by M, where reported): by tumour stage or pathological tumour size; axillary node status</td>
<td>No. of false-positives attributed to adjunct US (% of screens or subjects)</td>
</tr>
<tr>
<td></td>
<td>Country</td>
<td></td>
<td>Interval cancers</td>
<td>Surgical biopsy</td>
</tr>
<tr>
<td>Buchberger et al. (2000)</td>
<td>Non-randomized, retrospective, no comparison group; most had negative M/CBE; included some with PHBC or FHBC</td>
<td>2–4</td>
<td>32 (0.39%)</td>
<td>NR</td>
</tr>
<tr>
<td>Austria</td>
<td>n = 8103 asymptomatic women who had negative M and CBE (included some with PHBC)</td>
<td></td>
<td>Mean invasive cancer size, 9.1 mm (not significantly different from M-detected cancers)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35–78 yr (mean, 49 yr)</td>
<td></td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Kaplan (2001)</td>
<td>Non-randomized, retrospective, no comparison group; most had negative M/CBE</td>
<td>3, 4</td>
<td>6 (0.32%)</td>
<td>NR</td>
</tr>
<tr>
<td>USA</td>
<td>n = 1862</td>
<td></td>
<td>All 6 cancers early stage: 1 in situ, 5 stage I all node-negative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35–87 yr</td>
<td></td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td>Kolb et al. (2002)</td>
<td>Non-randomized, retrospective, no comparison group; most had negative M and CBE (included some with PHBC or FHBC)</td>
<td>2–4</td>
<td>33 cancers in 31 women (0.27%)</td>
<td>NR</td>
</tr>
<tr>
<td>USA</td>
<td>n = 12 193 screens (4897 women) who had negative M and CBE (included some with PHBC or FHBC)</td>
<td></td>
<td>89% in situ or stage I cancers; mean size: 9.9 mm (stage and size not different from M-detected)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean, 54.7 yr</td>
<td></td>
<td>89% node-negative</td>
<td></td>
</tr>
<tr>
<td>Corsetti et al. (2008, 2011)</td>
<td>Non-randomized, retrospective, no comparison group; most had negative M</td>
<td>3, 4</td>
<td>37 (0.40%)</td>
<td>NR</td>
</tr>
<tr>
<td>Italy</td>
<td>n = 9157 screens in self-referring women with negative M</td>
<td></td>
<td>Early-stage (in situ or small invasive) cancers: 64.8% vs 35.5%, P = 0.001 positive nodes: 13.5% vs 31.3%, P = 0.047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mean, 52 yr</td>
<td></td>
<td>8 interval cancers from 7172 negative screens at 1 yr: 1.1/1000</td>
<td></td>
</tr>
<tr>
<td>Reference Country</td>
<td>Study characteristics; no. screened with US(^a); age</td>
<td>Breast density(^b)</td>
<td>Preventive or screening effect</td>
<td>Adverse effect</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------------------------------------</td>
<td>---------------------</td>
<td>--------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td><strong>Kelly et al. (2010)</strong> USA</td>
<td>Non-randomized, retrospective (n = 6425) screens in 4419 asymptomatic women (included some with PHBC or FHBC) ≥ 35 yr</td>
<td>3, 4 with or without additional risk factor</td>
<td>Additional detection: no. of US-only detected cancers (% of screens or subjects)</td>
<td>Surgical biopsy</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>23 (0.52%)</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M detection: 3.6/1000</td>
<td>False-positives NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>US detection: 7.2/1000</td>
<td>recall 7.2% for US vs 4.2% for M ((P &lt; 0.01)); 9.6% for combined M + US</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Characteristics of US-only detected cancers (vs cancers detected by M, where reported): by tumour stage or pathological tumour size; axillary node status(^c)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>US detected more invasive cancers ≤ 10 mm (14 of 21) than mammography ((P &lt; 0.01))</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Interval cancers</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11 interval cancers at 1 yr: 1.7/1000</td>
<td></td>
</tr>
<tr>
<td><strong>Hooley et al. (2012)</strong> USA</td>
<td>Non-randomized, retrospective, no comparison group (n = 935) women with recent negative M who also had US (included some at intermediate or high risk) 29–89 yr (mean, 52 yr)</td>
<td>3, 4</td>
<td>Additional testing</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[3 (0.32%)] reported as 3.2; 95% CI, 0.8–10/1000 screens</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>All 3 cancers &lt; 10 mm (includes 1 DCIS) all node-negative</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td><strong>Weigert &amp; Steenbergen, (2012)</strong> USA</td>
<td>Non-randomized, retrospective chart review from radiology services, no comparison group (n = 8647) women with recent negative M who also had US screen age of cancer patients, 42–78 yr</td>
<td>3, 4 (&gt; 50% of breast dense)</td>
<td>Average size, 19 mm (for 17 invasive cancers) 1 node-positive</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>28 (0.32%) including 2 ADH and 1 LCIS; re-calculated as [25 (0.29%)]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 interval cancer at 6 mo</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>429 (4.96%) recommended to have biopsy</td>
<td></td>
</tr>
<tr>
<td><strong>Venturini et al. (2013)</strong> Italy</td>
<td>Non-randomized, prospective screening study tailored to breast density and (intermediate) risk: women with negative M and dense breasts (n = 835) women 40–49 yr</td>
<td>3, 4</td>
<td>Both cancers &lt; 15 mm 1 node-positive</td>
<td>False-positive invasive tests: 0.9% for US vs 0.1% for M</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 (0.24%)</td>
<td>Short-interval follow-up: 7.5% for US vs 0.3% for M</td>
</tr>
</tbody>
</table>
Table 5.15 (continued)

<table>
<thead>
<tr>
<th>Reference Country</th>
<th>Study characteristics; no. screened with US; age</th>
<th>Breast density&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Preventive or screening effect</th>
<th>Adverse effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brem et al. (2014) USA</td>
<td>Non-randomized, prospective screening study tailored to breast density (included some intermediate risk groups) n = 15 318 women ≥ 25 yr</td>
<td>3, 4</td>
<td>Additional detection: no. of US-only detected cancers (% of screens or subjects) 30 (0.19% of all screened women)</td>
<td>Characteristics of US-only detected cancers (vs cancers detected by M, where reported): by tumour stage or pathological tumour size; axillary node status&lt;sup&gt;c&lt;/sup&gt; Similar mean cancer size for M-detected (13 mm) and US-detected (12.9 mm) US-only detected cancers were more frequently invasive than M-detected cancers (P &lt; 0.05)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

<sup>a</sup> The study of Kelly et al. (2010) used automated whole-breast ultrasonography, and the study of Brem et al. (2014) used 3D automated breast ultrasonography. All other studies used handheld ultrasonography.

<sup>b</sup> Based on BI-RADS (Breast Imaging Reporting and Data System; D’Orsi et al., 2013) density categories: 1, almost entirely fatty (< 25% fibroglandular); 2, scattered fibroglandular densities (25–50% fibroglandular); 3, heterogeneously dense (51–75% fibroglandular); 4, extremely dense (> 75% fibroglandular).

<sup>c</sup> Based on women who underwent axillary node surgery or dissection.

ADH, atypical ductal hyperplasia; CBE, clinical breast examination; CNB, core needle biopsy; DCIS, ductal carcinoma in situ; FHBC, family history of breast cancer; PHBC, personal history of breast cancer; US, ultrasonography; yr, year or years.

cancer detection (as an indicator of potential effectiveness) within screened subjects; none of these studies were designed to assess screening benefit in terms of mortality reduction or using a surrogate for effectiveness of screening, such as a reduction in interval cancer rates. Incremental detection of breast cancer by ultrasonography was in the range of 0.19% to 0.52% of all screens. The highest estimate (Kelly et al., 2010) included women at an increased risk, including some women with a history of breast cancer, and reported a modest cancer detection rate for mammography. Therefore, the incremental detection of breast cancer by ultrasonography was substantial but heterogeneous, representing approximately 14% to 48% of the detected cancers (Corsetti et al., 2008; Venturini et al., 2013). [These data should be interpreted taking into account that several studies included, among women with dense breasts, subgroups of women at increased risk due to other risk factors (i.e. dense breasts plus other risk factors), and many studies included young women, and therefore the evidence may not be generalizable to population screening of women with dense breasts.] The two prospective studies reported the lowest incremental detection rates for ultrasonography, of 0.19% (Brem et al., 2014) and 0.24% (Venturini et al., 2013) of screens. Ultrasonography-only detected cancers were frequently early-stage cancers, generally at a comparable or earlier stage than cancers detected with mammography, although comparative data on cancer characteristics were not comprehensively reported.

Giuliano & Giuliano (2013) examined detection measures for automated breast ultrasonography screening in women with dense (density > 50%) breast tissue (test group) and used a different cohort of women with dense breasts from an earlier time frame as a control group for mammography screening. [This study is limited by the comparison of two cohorts with different underlying breast cancer prevalence (test group, 1.25%; control group, 0.60%).] For the test group (n = 3418; median age, 57 years) screened with mammography and ultrasonography, the screening sensitivity was 97.7%, the specificity was 99.7%, the cancer detection rate was 12.3 per 1000 screens, and the mean tumour size of detected cancers was 14.3 mm. For the control group (n = 4076; median age, 54 years) screened with digital mammography alone, the screening sensitivity was 76.0%, the specificity was 98.2%, the cancer detection rate was 4.6 per 1000 screens, and the mean tumour size of detected cancers was 21.3 mm. [This mean size is larger than expected for a screened population. The inferred 2.6-fold increase in the cancer detection rate, which represents one additional detection in approximately 0.70% of screens, was attributed to ultrasonography. This is well above estimates from all the other reviewed studies and is probably due to the comparison of cohorts with different underlying breast cancer risk. In addition, the relatively high specificity in the test group, based on the combined screening approach, is unusual and is inconsistent with all the other studies. Because of these limitations, this study was considered uninformative.]

One prospective screening study of ultrasonography in a multimodality setting (CBE, mammography, and ultrasonography) included 3028 Chinese women aged 25 years and older (Huang et al., 2012), not restricted to women with dense breasts. The sensitivity was higher for mammography (84.8%) than for ultrasonography (72.7%); however, ultrasonography detected 3 cancers not detected with mammography (all were in women with dense breasts). Ultrasonography yielded an incremental cancer detection rate of [0.99 per 1000] screens of all screening participants. Mammography-detected cancers were more frequently smaller than 20 mm and node-negative than those detected with ultrasonography or CBE.

Two non-randomized studies of adjunct ultrasonography for screening dense breasts reported data on interval cancers (Kelly et al.,
2010; Weigert & Steenbergen, 2012). [Given that these studies did not have a comparison estimate and had a relatively short follow-up period (12 months), it is difficult to interpret the estimated interval cancer rates.] Corsetti et al. (2008, 2011) reported indirect comparisons based on follow-up for first-year interval cancers in a cohort of self-referring women attending a breast service in Italy. The estimated first-year interval cancer rate was 1.1 per 1000 screens (from 7172 negative screens with follow-up) in women who underwent adjunct ultrasonography and had dense breasts, compared with 1.0 per 1000 screens (from 12,438 negative screens with follow-up) in women who received mammography only and did not have dense breasts.

(b) Digital breast tomosynthesis/three-dimensional mammography

Digital breast tomosynthesis is a derivative of digital mammography that produces quasi three-dimensional images, which reduces the effect of tissue superimposition and can therefore improve mammography interpretation (see Section 2.1.4 for details). A recent systematic review (Houssami & Skaane, 2013) examined the available evidence on the accuracy of digital breast tomosynthesis. The studies identified were relatively small ($n = 14$), comprised mostly test-set observer (reader) studies or clinical series that included symptomatic and screen-recalled cases, and were generally enriched with breast cancer cases. Taking into consideration the limitations of the studies, the evidence can be summed up as follows (Houssami & Skaane, 2013): (i) two-view digital breast tomosynthesis has accuracy that is equal to or better than that of standard two-view mammography; (ii) one-view digital breast tomosynthesis does not have better accuracy than two-view mammography; (iii) the addition of digital breast tomosynthesis to digital mammography increases interpretive accuracy; (iv) improved accuracy from using digital breast tomosynthesis (relative to, or added to, digital mammography) was the result of increased cancer detection or reduced false-positive recalls, or both; and (v) subjective interpretation of cancer conspicuity consistently found that cancers were equally or more conspicuous on digital breast tomosynthesis relative to digital mammography.

A review of the literature did not identify any RCTs examining the efficacy of digital breast tomosynthesis in population breast screening; however, digital breast tomosynthesis was the only other imaging technology investigated in population-based screening programmes in women at average (population) risk (Ciatto et al., 2013; Haas et al., 2013; Rose et al., 2013; Skaane et al., 2013a, b, 2014; Friedewald et al., 2014; Houssami et al., 2014a; Table 5.16). All these studies investigated digital mammography with tomosynthesis (also referred to as integrated two-dimensional/three-dimensional [2D/3D] mammography), using various methodologies (different design and reading/recall protocols). None were designed with the aim of assessing screening benefit in terms of mortality reduction or using a surrogate for effectiveness of screening, such as a reduction in interval cancer rates. Also, none of the studies reported estimates of overt-diagnosis. Two studies were prospective population-based trials embedded within organized screening programmes in Europe: the Screening with Tomosynthesis or Standard Mammography (STORM) trial in Italy (Ciatto et al., 2013) and the Oslo trial in Norway (Skaane et al., 2013a, b, 2014). Both studies used double reading according to European standards, but they used different recall protocols. Both studies performed digital mammography with tomosynthesis in all participants, and hence they reported paired data for screened women (within screening participant comparison).

The STORM trial (Ciatto et al., 2013; Houssami et al., 2014a) compared sequential screen-readings by the same readers for the same women: digital mammography alone and integrated 2D/3D mammography. The study reported
### Table 5.16 Studies evaluating tomosynthesis for population breast cancer screening: three-dimensional mammography as adjunct to digital mammography

<table>
<thead>
<tr>
<th>Reference</th>
<th>Study characteristics</th>
<th>Preventive or screening effect</th>
<th>Adverse effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cancer detection rates/1000</td>
<td>Absolute effect of 3D M on cancer detection compared with 2D alone</td>
</tr>
<tr>
<td></td>
<td>Design (no. of screens); screen-reading methods</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ciatto et al. (2013) [STORM trial] Italy</td>
<td>Prospective trial ($n = 7292$) in population-based programme, comparing 2D and integrated 2D/3D screening (paired data); sequential double reading, recall by either reader at either read</td>
<td>2D: 5.3 2D/3D: 8.1 $P &lt; 0.001$</td>
<td>Increase of 2.7/1000</td>
</tr>
<tr>
<td>Houssami et al. (2014a) [STORM follow-up study] Italy</td>
<td>Extended analysis of STORM trial ($n = 7292$), comparing various screening strategies, includes follow-up for year 1 interval cancers 2D double reading: 5.3 2D/3D single reading: 7.5 $P &lt; 0.001$ [other comparisons also reported]</td>
<td>Increase of 2.2/1000</td>
<td>See above</td>
</tr>
<tr>
<td>Skaane et al. (2013a, b) [Oslo trial] Norway</td>
<td>Prospective trial ($n = 12,631$) in population-based programme, comparing 2D and 2D/3D screening (paired data); randomized readings to 4 study arms with various screen-reading strategies; data shown are for analyses of single reading or double reading of tomosynthesis 2D: 6.1 2D/3D: 8.0 27% increase $P = 0.001$ Double reading: 2D: 7.1 2D/3D: 9.4 $P &lt; 0.001$</td>
<td>Increase of 1.9/1000</td>
<td>Cancers detected with 2D/3D only were mostly invasive and more frequently grade 2 or 3 (2 DCIS cases were detected with 2D/3D only)</td>
</tr>
</tbody>
</table>
Table 5.16 (continued)

<table>
<thead>
<tr>
<th>Reference [Study] Country</th>
<th>Study characteristics Design (no. of screens); screen-reading methods</th>
<th>Preventive or screening effect</th>
<th>Adverse effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cancer detection rates/1000 screens</td>
<td>Absolute effect of 3D M on cancer detection rate compared with 2D alone</td>
</tr>
<tr>
<td>Skaane et al. (2014) [Oslo trial] Norway</td>
<td>See above Analysis of 2D <em>syn</em>/3D</td>
<td>2D/3D: 7.8 2D <em>syn</em>/3D: 7.7 Not significantly different</td>
<td>Increase of 2.3/1000</td>
</tr>
<tr>
<td>Rose et al. (2013) USA</td>
<td>Retrospective: before vs after (13 856 vs 9499) introduction of 3D as adjunct to 2D screening; single reading from readers from several radiology services</td>
<td>2D: 4.0 2D/3D: 5.4 2D <em>syn</em>/3D: 7.7 For invasive cancer: 2D: 2.8 2D/3D: 4.3 2D <em>syn</em>/3D: 2.8 2D <em>syn</em>/3D: 5.4 P = 0.18</td>
<td>Increase of 1.4/1000 2D/3D: 5.4 Increase of 1.5/1000 2D/3D: 5.4</td>
</tr>
<tr>
<td>Haas et al. (2013) USA</td>
<td>Retrospective: services using 2D vs services using 2D/3D (7058 vs 6100) in same year; single reading from readers from breast or radiology services</td>
<td>2D: 5.2 2D/3D: 5.7 2D <em>syn</em>/3D: 7.7 P = 0.07</td>
<td>Increase of 0.5/1000 2D/3D: 5.7</td>
</tr>
<tr>
<td>Reference [Study] Country</td>
<td>Study characteristics Design (no. of screens); screen-reading methods</td>
<td>Preventive or screening effect</td>
<td>Adverse effect</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td><strong>Cancer detection rates/1000 screens</strong></td>
<td>Absolute effect of 3D M on cancer detection rate compared with 2D alone</td>
<td>Characteristics of cancers detected only with integrated 2D/3D M only</td>
<td>Interval cancers</td>
</tr>
<tr>
<td><strong>False-positive recalls</strong></td>
<td>Absolute effect of 3D M on FPR compared with 2D alone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friedewald et al. (2014) USA</td>
<td>Retrospective: before vs after (281 187 vs 173 663) introduction of 3D as adjunct to 2D M screening; single reading from readers from 13 radiology services</td>
<td>2D: 4.2 2D/3D: 5.4 ( P &lt; 0.001 )</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td>Increase of 1.2/1000</td>
<td>Cancers detected with 2D/3D only comprised invasive cancer; DCIS rates similar for 2D and 2D/3D; stage data NR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Increase of 1.2/1000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data for all recalls: 2D: 10.7% 2D/3D: 9.1% ( P &lt; 0.001 ) For all biopsies (includes cancer): 2D: 1.8% 2D/3D: 1.9% ( P = 0.004 )</td>
<td>Decrease of 1.6%</td>
<td></td>
</tr>
</tbody>
</table>

Note: 2D refers to digital mammography acquisition of 2-view mammographic images, whereas 2D_synth refers to 2D mammographic images synthesized (reconstructed) from the digital breast tomosynthesis acquisition.

Decrease in FPR is estimated for recall conditional to 3D-positivity (Ciatto et al., 2013; Houssami et al., 2014a), whereas false-positive scores from the Oslo study were based on pre-arbitration data (Skaane et al., 2013a, b).

2D, two-dimensional; 3D, three-dimensional; DCIS, ductal carcinoma in situ; FPR, false-positive recall; M, mammography; NR, not reported; STORM, Screening with Tomosynthesis or Standard Mammography; 2D_synth/3D, tomosynthesis with synthetically reconstructed 2D images.
Breast cancer screening

A significant incremental cancer detection rate of 2.7 per 1000 screens for integrated 2D/3D mammography versus digital mammography ($P < 0.001$). The Oslo trial (Skaane et al., 2013a, b) randomized readers to four screen-reading strategies that used digital mammography or integrated 2D/3D mammography, allowing assessment of reconstructed 2D mammography in one of the study arms (Skaane et al., 2014). The study showed a significant incremental cancer detection rate of 1.9 per 1000 screens for integrated 2D/3D mammography versus digital mammography in a reader-adjusted analysis ($P = 0.001$) (Skaane et al., 2013a) and of 2.3 per 1000 screens for double reading of integrated 2D/3D mammography versus digital mammography ($P < 0.001$) (Skaane et al., 2013b). A further analysis (Skaane et al., 2014) found that integrated 2D/3D mammography yielded a similar incremental cancer detection rate compared with digital mammography whether by dual acquisition of digital mammography with tomosynthesis (acquired 2D and 3D images) or by tomosynthesis acquisition with synthetic 2D mammography (3D acquisition only, and 2D images reconstructed from the 3D data).

A third prospective screening trial, also conducted within a population-based programme, was in progress in Malmö, Sweden, at the time of the Handbook Working Group Meeting, in November 2014. This trial differs from the other screening studies of this technology in that it compares screen-reading using digital mammography alone (two views) with screen-reading using tomosynthesis alone (one 3D mammography view); hence, it is the only population-based breast screening study reporting detection estimates for tomosynthesis alone. [Note added after the Meeting: The results of the trial have been published (Läng et al., 2015). The incremental cancer detection rate was 2.6 per 1000 screens using tomosynthesis alone versus digital mammography ($P < 0.0001$).]

Three retrospective studies have also examined digital mammography with tomosynthesis for population screening (Haas et al., 2013; Rose et al., 2013; Friedewald et al., 2014); all three studies were conducted in the USA and hence used single reading as practised in the USA. Two studies (Rose et al., 2013; Friedewald et al., 2014) used a before–after methodology, comparing detection measures before and after the introduction of integrated 2D/3D mammography, whereas one study (Haas et al., 2013) compared services using digital mammography with services using integrated 2D/3D mammography within the same time frame. The largest retrospective study (Friedewald et al., 2014) was a comparison of 281 187 versus 173 663 screens before and after the introduction of tomosynthesis as adjunct to digital mammography screening in 13 radiology services, and reported a significant incremental cancer detection rate of 1.2 per 1000 screens. Overall, the three studies showed a modest incremental detection rate with the use of adjunct tomosynthesis (range, 0.5–1.4 per 1000 screens) relative to the prospective trials; however, the direction of the estimated increased cancer detection is consistent across all studies.

Four out of five studies provided limited data on the characteristics of the cancers detected with integrated 2D/3D mammography compared with digital mammography. [Studies were generally not powered for such analyses.] Two studies indicated that the increased cancer detection achieved by digital mammography with tomosynthesis was mostly of invasive disease (Rose et al., 2013; Friedewald et al., 2014), whereas two studies showed incremental detection of both invasive and in situ disease (Ciatto et al., 2013; Skaane et al., 2013b).

Data on interval cancer rates for this technology are limited to the follow-up report from the STORM trial; the estimated interval cancer rate based on only 12 months of follow-up is 0.82 per 1000 (95% CI, 0.30–1.79) (Houssami et al., 2014a).
Several studies reported on the use of integrated 2D/3D mammography screening in reducing false-positive recalls (Table 5.16). The reduction in false-positive recalls is most marked in the retrospective studies reported from the USA (absolute decreases in false-positive results range from 1.6% to 3.6%), where the baseline false-positive recall rates for digital mammography alone are relatively high (range, 8.7–12.0%). The estimated reduction in false-positive recalls in the prospective studies, which were conducted in European population screening programmes and had relatively low recall rates, was modest (0.8% and 2%), and the latter was an estimate conditional to 3D mammography positivity. Furthermore, one of the studies (Skaane et al., 2013b) showed that for double reading, digital mammography with tomosynthesis reduced false-positive recalls compared with mammography alone, but increased overall recall (see Table 5.16). [It is likely that the potential for digital mammography with tomosynthesis to reduce false-positive recalls will depend on both the false-positive recall rates at digital mammography and the recall rules, which vary according to the screening programme.]

5.5.2 Adverse effects

(a) Breast ultrasonography

The adverse effects of breast ultrasonography screening have been examined in non-randomized retrospective and prospective studies in women with dense breast tissue (Buchberger et al., 2000; Kaplan, 2001; Kolb et al., 2002; Corsetti et al., 2008, 2011; Kelly et al., 2010; Hooley et al., 2012; Weigert & Steenbergen, 2012; Venturini et al., 2013; Brem et al., 2014). The main adverse effect is additional false-positive intervention. Ultrasonography caused additional testing (needle biopsy or imaging follow-up) in 1.2–6.3%, and also surgical biopsy (although some studies included non-surgical biopsy in this percentage) in 0.9–2.7% due to false-positives (Table 5.15).

The study of Kelly et al. (2010), which included some women at an increased risk, reported an overall recall rate [not distinctly false-positive recall] of 7.2% for ultrasonography (vs 4.2% for mammography; \( P < 0.01 \)), and the combined strategy had an overall recall rate of 9.6% in that study. Venturini et al. (2013) reported a false-positive biopsy rate for ultrasonography of 0.9% (vs 0.1% for mammography) in a cohort of young women (aged 40–49 years) with dense breast tissue and intermediate lifetime risk. Brem et al. (2014) reported an overall recall rate of 28.5% for adjunct ultrasonography with mammography (vs 15% for mammography alone; \( P < 0.001 \)).

Given that there is substantial increased detection of breast cancer using adjunct ultrasonography in women with mammography-negative dense breasts, it seems possible that overdiagnosis could occur in this context. However, overdiagnosis has not been reported in any of the studies reviewed (Buchberger et al., 2000; Kaplan, 2001; Kolb et al., 2002; Corsetti et al., 2008, 2011; Kelly et al., 2010; Hooley et al., 2012; Weigert & Steenbergen, 2012; Venturini et al., 2013; Brem et al., 2014). [It would be difficult to attempt to estimate overdiagnosis based on the available data, due to (but not limited to) the lack of a control or comparison cohort and the heterogeneity of the screened populations, including variable underlying risk profiles.]

(b) Digital breast tomosynthesis/three-dimensional mammography

All studies reviewed reported a reduction in false-positive recalls using integrated 2D/3D mammography (Table 5.16). Therefore, this does not seem to be an adverse effect of this technology. [The same may not apply for 3D screening alone.]

Given that there is increased detection of breast cancer using digital mammography with tomosynthesis, it seems possible that overdiagnosis could occur in this context. Several studies (Rose et al., 2013; Skaane et al., 2013a; Friedewald et al., 2014) have suggested that digital breast
Breast cancer screening

tomosynthesis mostly increases detection of invasive cancers. However, none of the studies have reported on overdiagnosis. [The currently available data do not allow inferences relating to overdiagnosis from the increased cancer detection attributed to tomosynthesis.]

The main potential adverse effect of digital mammography with tomosynthesis relates to the radiation dose to the breast if dual acquisition is used. Digital breast tomosynthesis is reported to deliver on average similar doses to digital mammography (Feng & Sechopoulos, 2012; Houssami & Skaane, 2013). Thus, using dual acquisition by digital mammography with tomosynthesis approximately doubles the radiation dose. In the two population screening studies, the mean glandular dose per view was 1.58 mGy for digital mammography and 1.95 mGy for digital breast tomosynthesis in the Oslo study (Skaane et al., 2013a) and 1.22 mGy for digital mammography and 2.99 mGy (1.22 + 1.77 mGy) for integrated 2D/3D mammography in the STORM study (Bernardi et al., 2014). Recent tomosynthesis technology allows reconstruction of the 2D images from the data obtained from the tomosynthesis acquisition (also referred to as synthetic 2D mammography), eliminating the need for dual acquisition. Reconstruction of the 2D images from the tomosynthesis acquisition decreases the radiation dose by 45% compared with the dual acquisition (Skaane et al., 2014) and performs similarly to digital mammography with tomosynthesis from dual acquisition (see Section 5.5.1 and Table 5.16).

5.5.3 Cost–effectiveness analysis

(a) Breast ultrasonography

There were no studies of breast ultrasonography for population breast screening that reported on cost per life year gained or QALY saved. Cost analyses were reported by four of the studies that investigated ultrasonography in women with dense breasts. Studies conducted in the USA (Hooley et al., 2012; Weigert & Steenbergen, 2012) reported relatively higher costs than those conducted in Europe (Corsetti et al., 2008; Venturini et al., 2013). Hooley et al. (2012) estimated the cost of adjunct ultrasonography, factoring in the costs of ultrasonography and related biopsy and short-interval imaging follow-up (using the Medicare reimbursement rate), to be $US 60 267 per detected breast cancer. Weigert & Steenbergen (2012), using the average reimbursement rate for ultrasonography and related biopsy, estimated the cost of adjunct ultrasonography screening to be $US 110 241 per detected breast cancer.

In the European setting, Corsetti et al. (2008) estimated the cost of adjunct ultrasonography, factoring in the costs of ultrasonography and related testing and any form of biopsy, to be in the range of €14 618–15 234 per detected breast cancer. Venturini et al. (2013) reported the cost of screening young women with dense breasts; mammography was estimated to cost €6377 per detected breast cancer, whereas adjunct ultrasonography in the same programme was estimated to cost €19 158 per detected breast cancer.

(b) Digital breast tomosynthesis/three-dimensional mammography

There were no studies available of the cost–effectiveness, or any cost analyses, of digital mammography with tomosynthesis in population breast screening. Digital breast tomosynthesis is more expensive than digital mammography and requires more imaging storage and display infrastructure, all of which increase the costs and the resources needed for screening implementation. Digital mammography with tomosynthesis also increases screen-reading time, resulting in an approximate doubling (Houssami & Skaane, 2013); based on the Oslo trial (Skaane et al., 2013a), the mean interpretation time was 91 seconds for integrated 2D/3D mammography versus 45 seconds for digital mammography ($P < 0.001$).
5.5.4 Other techniques

(a) Magnetic resonance imaging

Breast MRI has been shown to have superior screening sensitivity to mammography in women at an increased risk of developing breast cancer (see Section 5.6). Searches of the literature did not identify any studies of MRI for screening of women considered at average (population) risk. One recent study (Kuhl et al., 2014) of an abbreviated (fast) MRI protocol screened 443 women “referred to MRI screening on clinical grounds”; 82% of the women were considered to be at mildly or moderately increased risk, because of either dense breast tissue or a mild or moderate family history of breast cancer. The 146 women with a personal history of breast cancer were having imaging of the contralateral breast. In this selected subject group, reportedly “pre-screened” with digital mammography and ultrasonography [data not reported for either], MRI yielded an incremental cancer detection rate of 18 per 1000 screens. False-positive rates varied by the applied MRI protocol and were in the range of 5.6–29%. [The findings from this “proof-of-concept” reader study are early and do not represent population screening.]

(b) Electrical impedance imaging

The literature search did not identify any RCTs or population-based studies of electrical impedance scanning for breast screening. Studies of electrical impedance technologies for imaging of the breast have used various devices and instrumentation, operated at various frequencies and interpreted using variable methods (e.g. visual, computer algorithms, or other methods) (Malich et al., 2001; Martin et al., 2002; Wersebe et al., 2002; Diebold et al., 2005; Fuchsjaeger et al., 2005; Zheng et al., 2008, 2011; Wang et al., 2010; Lederman et al., 2011).

All these studies were relatively small clinical series or diagnostic studies of women who had suspicious or equivocal (mammography or other image-detected) findings and included both symptomatic and asymptomatic women; these studies were based on women who were undergoing biopsy (surgical or core needle biopsy), and hence the studies were highly enriched with breast cancer cases (prevalence in the range of 5–60%).

One relatively large study assessed electrical impedance imaging for “risk-stratification” and screening of asymptomatic young women (aged 30–45 years) (Stojadinovic et al., 2005, 2008). [One limitation of this study is that the study participants included women with mammographic findings or clinical abnormalities who were scheduled to undergo biopsy.] The study reported an extremely low sensitivity for screening of 26.4%, and specificity of 94.7%.

(c) Scintimammography (molecular breast imaging)

The literature search did not identify any studies evaluating the efficacy or effectiveness of this technology for breast screening of women at average (population) risk.

Scintimammography has been used and evaluated in various clinical applications for breast imaging, predominantly in small and/or highly selected clinical series and diagnostic studies highly enriched with breast cancer cases (19–100%), including, but not limited to: diagnostic workup of suspicious or indeterminate mammography-detected (or other image-detected) findings; breast assessment in women scheduled for biopsy on the basis of clinical or mammographic abnormalities; staging of a known cancerous breast lesion; monitoring response to treatment; and detecting breast cancer recurrence (Bekis et al., 2004; Rhodes et al., 2005; Adedapo & Choudhury, 2007; Duarte et al., 2007; Gommans et al., 2007; O’Connor et al., 2007; Spanu et al., 2007, 2008, 2009; Hruska et al., 2008; Kim et al., 2009; Sharma et al., 2009; Xu et al., 2011; Lee et al., 2012; Spanu et al., 2012; Weigert et al., 2012; BlueCross BlueShield, 2007; Stojadinovic et al., 2005, 2008).
A meta-analysis (Xu et al., 2011) of 45 extremely heterogeneous diagnostic accuracy studies of scintimammography reported meta-estimates of 83% for sensitivity and 85% for specificity; in the subgroup of subjects without a palpable mass, meta-estimates were 59% for sensitivity and 89% for specificity.

Three studies reported screening of defined asymptomatic populations, which included women at an increased risk. Brem et al. (2005) screened with scintimammography 94 women at an increased risk who had normal mammograms and CBE. They detected 2 additional invasive (9 mm) cancers (+2%); however, this was at the trade-off of 14 additional false-positives (+15%). Rhodes et al. (2011) screened 936 women (aged 25–89 years) with dense breasts and at an increased risk (personal history of breast cancer or lobular carcinoma in situ [LCIS] or atypical proliferations, or BRCA mutations) using dedicated dual-head gamma imaging (with the radiotracer $^{99m}$Tc-sestamibi). The detection yield was 3.2 per 1000 screens for mammography and 9.6 per 1000 screens for scintimammography (incremental cancer detection rate, 7.5 per 1000 screens). Most of the cancers detected on scintimammography only were node-negative invasive cancers (median size, 11 mm). [The sensitivity of mammography was extremely low (27%).] False-positive recall rates (9% for mammography, 8% for scintimammography) and specificity (91% for mammography, 93% for scintimammography) were similar for the two tests. Finally, Hruska et al. (2012) reported a study of molecular breast imaging with $^{99m}$Tc-sestamibi in 306 asymptomatic women (aged 37–88 years), including some women at an increased risk, such as those with a personal history of breast cancer, who were undergoing myocardial perfusion imaging. Scintimammography had an incremental cancer detection yield of 13 per 1000 screens (4 cancers) relative to mammography in the previous 12 months, and caused additional false-positives in approximately [6%] of subjects.

The radiation dose to the whole body from this technology (see Section 2.2.4 for details) is reported to be 15–30 times the radiation dose from digital mammography (BlueCross BlueShield Association, 2013).

**Positron emission mammography**

Literature searches did not identify any population breast screening studies of positron emission mammography. This technology has been evaluated in very specific and limited clinical applications of breast imaging, predominantly for staging of a lesion; for preoperative assessment of disease extent (generally in comparison with MRI); for “screening” of the contralateral breast in preoperative staging; for response monitoring, in very small series of women with a biopsy of suspicious findings; or in phantom studies (Raylman et al., 2000; Levine et al., 2003; Tafra et al., 2005; Berg et al., 2011, 2012a; Schilling et al., 2011; Schilling, 2012; Shkumat et al., 2011; Eo et al., 2012; Kalles et al., 2013). Positron emission mammography involves much higher doses of radiation (whole-body radiation) and a much longer acquisition time (for two views of both breasts) than mammography (see Section 2.2.3).

### 5.5.5 Psychosocial harm

Few studies have measured psychosocial harm from imaging techniques other than mammography. One study found that MRI screening was more distressing than X-ray mammography both shortly after and 6 weeks after the screening procedure (Hutton et al., 2011), whereas another study found no difference between MRI and mammography screening in psychological outcomes (Brédart et al., 2012). As with other screening processes, psychological harm may depend on the conduct of the technology, such as the number of false-positive and false-negative screens and the waiting time from examination to result (see also Sections 3.1.4 and 5.3.5).
5.6 Screening of women at an increased risk

In some women, the risk of developing breast cancer during their lifetime is increased compared with that of women in the general population, and usually with an earlier expected age of onset. This increased risk may be attributed to the presence of a genetic or familial predisposition to breast cancer, to a personal history of invasive breast cancer or DCIS, or to the presence of lobular neoplasia or atypical proliferations. It should be noted that a familial predisposition, if not assessed by a specialized genetic centre, should not be used as an indication for screening outside the scope of the population breast cancer screening programme.

In general, it is preferable that women at an increased risk be screened outside the scope of a population breast cancer screening programme, for two reasons. First, regular population screening programmes with mammography might be insufficient, due to the earlier age of onset of breast cancer in these women and due to the reduced sensitivity of mammography in these women. In addition, women with a BRCA1/2 mutation are more susceptible to radiation risk. Second, these women often require additional care, assessment, counselling, and information relevant to primary prevention and risk-reduction strategies (as might be provided, for example, through specialized genetics teams/units) that are generally well outside of the health-care brief of mammography screening programmes.

Evidence on the outcomes of screening for breast cancer in the several subgroups of women at an increased risk is summarized and discussed here.

5.6.1 High familial risk, with or without a BRCA1 or BRCA2 mutation

This section reports evidence on the effectiveness of screening with MRI alone, adjunct MRI, adjunct ultrasonography, or adjunct CBE as compared with mammography alone in women with a high familial risk, with or without a BRCA1 or BRCA2 mutation. Table 5.17 presents individual prospective studies, and Table 5.18 summarizes pooled and meta-analyses, and systematic reviews. The included studies are those that were performed prospectively, in which MRI and mammography were performed in the same screening round, and in which the review of the diagnostic test was performed blinded for the outcome of the other test. Studies that were performed retrospectively or unblinded, or in which MRI, ultrasonography, or mammography were not performed in parallel were excluded.

In addition, three reports reviewing the evidence of the effectiveness of adjunct MRI in the screening of women at an increased risk of breast cancer were identified (Table 5.18). One is a systematic review of the literature (Lord et al., 2007), one is a systematic review and meta-analysis at the level of published studies (Warner et al., 2008), and one is a pooled analysis of individual patient data (Phi et al., 2014).

(a) Adjunct magnetic resonance imaging

(i) Sensitivity and specificity in women with a BRCA1/2 mutation

Several studies focused on the added value of MRI compared with mammography and/or ultrasonography in the screening of women with a BRCA1 or BRCA2 mutation (Table 5.17 and Table 5.18). In the meta-analysis (Warner et al., 2008) and the pooled analysis (Phi et al., 2014), the estimates of the sensitivity of mammography were comparable, at about 40%, and increased with mammography combined with MRI similarly in both studies, to 94% (95% CI, 90–97%) in Warner et al. (2008) and 93.4% (95% CI,
Table 5.17 Prospective studies in women with a *BRCA1/2* mutation or a familial breast cancer risk screened with magnetic resonance imaging, mammography, ultrasonography, or clinical breast examination

<table>
<thead>
<tr>
<th>Reference, Country, study</th>
<th>Study period</th>
<th>Study design</th>
<th>Test results and related follow-up</th>
<th>Risk category</th>
<th>No. of women in study</th>
<th>No. of breast cancers</th>
<th>MRI Sens, Spec (%)</th>
<th>M Sens, Spec (%)</th>
<th>US Sens, Spec (%)</th>
<th>CBE Sens, Spec (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1/2 43</td>
<td>8</td>
<td>100</td>
<td>97.5</td>
<td>25</td>
<td>96.9</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FH 241</td>
<td>20</td>
<td>100</td>
<td>97.7</td>
<td>25</td>
<td>97.4</td>
<td>30</td>
</tr>
<tr>
<td>Leach et al. (2005) United Kingdom, MARIBS study</td>
<td>1997–2004</td>
<td>Multicentre Double reading Annual MRI and M</td>
<td>BI-RADS 0, 3, 4, 5: biopsy</td>
<td>Total 649</td>
<td>35</td>
<td>77</td>
<td>81</td>
<td>40</td>
<td>93</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1 82</td>
<td>13</td>
<td>92</td>
<td>79</td>
<td>23</td>
<td>92</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA2 43</td>
<td>12</td>
<td>58</td>
<td>82</td>
<td>50</td>
<td>94</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FH 524</td>
<td>10</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1/2 48</td>
<td>4</td>
<td>100</td>
<td>78</td>
<td>50</td>
<td>8.3</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1 445</td>
<td>19</td>
<td>84</td>
<td>84</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA2 46</td>
<td>2</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1 445</td>
<td>19</td>
<td>84</td>
<td>84</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA2 46</td>
<td>2</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1/2 80</td>
<td>3</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FH 110</td>
<td>3</td>
<td>100</td>
<td>66.7</td>
<td>33.4</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Reference*</td>
<td>Country, study</td>
<td>Study period</td>
<td>Study design</td>
<td>Test results and related follow-up</td>
<td>Risk category</td>
<td>No. of women in study</td>
<td>No. of breast cancers</td>
<td>MRI Sens, Spec (%)</td>
<td>M Sens, Spec (%)</td>
<td>US Sens, Spec (%)</td>
</tr>
<tr>
<td>------------</td>
<td>----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td>---------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1</td>
<td>80</td>
<td>6</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA2</td>
<td>13</td>
<td>2</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FH</td>
<td>234</td>
<td>20</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1</td>
<td>27</td>
<td>2</td>
<td>50</td>
<td>50</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA2</td>
<td>17</td>
<td>2</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FH</td>
<td>565</td>
<td>14</td>
<td>78.6</td>
<td>NR</td>
<td>35.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1/2</td>
<td>53</td>
<td>5</td>
<td>NR</td>
<td>NR</td>
<td>99.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FH</td>
<td>436</td>
<td>22</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Rijnsburger et al. (2010)</td>
<td>Netherlands, MRISC study</td>
<td>1999–2006</td>
<td>Multicentre Single reading Annual MRI and M Biannual CBE</td>
<td>BI-RADS 4, 5: biopsy BI-RADS 0, 3: biopsy or additional imaging After abnormal CBE: additional imaging</td>
<td>Total</td>
<td>2157</td>
<td>97</td>
<td>70.7</td>
<td>89.7</td>
<td>41.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA1</td>
<td>422</td>
<td>35</td>
<td>66.7</td>
<td>91</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>BRCA2</td>
<td>172</td>
<td>18</td>
<td>69.2</td>
<td>91</td>
<td>61.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FH</td>
<td>1563</td>
<td>44</td>
<td>73</td>
<td>89.2</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 5.17 (continued)
<table>
<thead>
<tr>
<th>Referencea</th>
<th>Country, study</th>
<th>Study period</th>
<th>Study design</th>
<th>Test results and related follow-upb</th>
<th>Risk category</th>
<th>No. of women in study</th>
<th>No. of breast cancers</th>
<th>MRI Sens, Spec (%)</th>
<th>M Sens, Spec (%)</th>
<th>US Sens, Spec (%)</th>
<th>CBE Sens, Spec (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Single reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Annual MRI and M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Biannual US and CBE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRCAl</td>
<td>75</td>
<td>6</td>
<td>83.3</td>
<td>NR</td>
<td>50</td>
<td>NR</td>
<td>33.3</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRCAl2</td>
<td>68</td>
<td>5</td>
<td>80</td>
<td>NR</td>
<td>60</td>
<td>NR</td>
<td>20</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH</td>
<td>41</td>
<td>1</td>
<td>100</td>
<td>NR</td>
<td>100</td>
<td>NR</td>
<td>0</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sardanelli et al. (2011)</td>
<td>Italy, HIBCRIT 1 study</td>
<td>2000–2007</td>
<td>Multicentre</td>
<td>BI-RADS 4, 5: biopsy BI-RADS 3: 4 mo follow-up</td>
<td>Total</td>
<td>501</td>
<td>52</td>
<td>91.3</td>
<td>96.7</td>
<td>50</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Single reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Annual MRI, M, US, and CBE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRCAl</td>
<td>184</td>
<td>21</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRCAl2</td>
<td>146</td>
<td>10</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>FH</td>
<td>171</td>
<td>21</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Passaperuma et al. (2012)</td>
<td>Toronto, Canada</td>
<td>1997–2009</td>
<td>Single centre</td>
<td>BI-RADS 0, 4, 5: biopsy BI-RADS 3: 6, 12, 24 mo follow-up</td>
<td>Total</td>
<td>496</td>
<td>57</td>
<td>86</td>
<td>90</td>
<td>19</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Single reading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Annual MRI, M, US and CBE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>US was stopped in 2005 due to lack of Sens and Spec</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRCAl</td>
<td>267</td>
<td>31</td>
<td>90</td>
<td>NR</td>
<td>19</td>
<td>NR</td>
<td>20</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BRCAl2</td>
<td>229</td>
<td>26</td>
<td>80</td>
<td>NR</td>
<td>20</td>
<td>NR</td>
<td>0</td>
<td>NR</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Data reported from the most recent publication.

b Based on BI-RADS (Breast Imaging Reporting and Data System; D’Orsi et al., 2013) density categories: 1, almost entirely fatty (<25% fibroglandular); 2, scattered fibroglandular densities (25–50% fibroglandular); 3, heterogeneously dense (51–75% fibroglandular); 4, extremely dense (>75% fibroglandular).

c Due to the design of the Lehman et al. (2005) and Lehman et al. (2007) studies, only sensitivity could be reported.

d Only data for the BRCA1/2 mutation carriers are reported, as no MRI was performed in the other risk groups.

e Only the results for digital mammography are reported, as they are close to those for screen-film mammography.

BI-RADS, American College of Radiology Breast Imaging Reporting and Data System; CBE, clinical breast examination; FH, family history suspicious for an increased risk of breast cancer; HIBCRIT, High Breast Cancer Risk Italian Trial; M, mammography; MARIBS, Magnetic Resonance Imaging for Breast Screening; mo, month or months; MRI, magnetic resonance imaging; MRISC, MRI Screening; NR, not reported in the most recent publication; Sens, sensitivity; Spec, specificity; US, ultrasonography.
<table>
<thead>
<tr>
<th>Study</th>
<th>Included studies</th>
<th>Study design</th>
<th>Main outcome parameters</th>
<th>Results on main outcome parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sens US and M</td>
<td>49–67%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sens MRI and M (with or without US)</td>
<td>93–100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Recall rate with MRI compared with that without MRI</td>
<td></td>
<td>Adjunct MRI may increase patient recall rates 3–5-fold due to increased false-positive findings</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sens M and MRI</td>
<td>94% (90–97%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spec M</td>
<td>94.7% (93.0–96.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spec M and MRI</td>
<td>77.2% (74.7–79.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sens MRI</td>
<td>85.3% (69.1–93.8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sens M and MRI</td>
<td>93.4% (80.2–98.0%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spec M</td>
<td>93.6% (88.8–96.5%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spec MRI</td>
<td>84.7% (79.0–89.1%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spec M and MRI</td>
<td>80.3% (72.5–86.2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>In women aged &gt; 50 yr:</td>
<td>Sens M</td>
<td>38.1% (22.4–56.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sens MRI</td>
<td>84.4% (61.8–94.8%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sens M and MRI</td>
<td>94.1% (77.7–98.7%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spec M</td>
<td>95.9% (92.1–97.9%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spec MRI</td>
<td>88.5% (83.5–92.2%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spec M and MRI</td>
<td>85.3% (78.5–90.2%)</td>
</tr>
</tbody>
</table>

CI, confidence interval; M, mammography; MRI, magnetic resonance imaging; Sens, sensitivity; Spec, specificity; US, ultrasonography.
Breast cancer screening

80.2–98.0%) in Phi et al. (2014). The specificity of adjunct MRI was also similar in the two analyses, to 77.2% (95% CI, 74.7–79.7%) in Warner et al. (2008) and 80.3% (95% CI, 72.5–86.2%) in Phi et al. (2014). Thus, adding MRI to mammography in the screening of women with a BRCA1/2 mutation leads to a statistically significant increase in sensitivity of the screening strategy, accompanied by a decrease in specificity that was also statistically significant (see Table 5.18).

In the pooled analysis using individual data in women with BRCA1/2 mutations, for the screening of women aged 50 years and older, the highest sensitivity was reported for adjunct MRI (94.1%; 95% CI, 77.7–98.7%) compared with mammography alone (38.1%; 95% CI, 22.4–56.7%) and compared with MRI alone (84.4%; 95% CI, 61.8–94.8%) (Phi et al., 2014); the specificity was lowest for adjunct MRI.

(ii) Sensitivity and specificity in women without a BRCA1/2 mutation

Only two informative studies assessed the sensitivity and specificity of mammography and MRI separately for women with a familial risk without a known BRCA1 or BRCA2 mutation (Kuhl et al., 2005; Rijnsburger et al., 2010). Two other studies were considered uninformative due to the small number of breast cancers in that category (Lehman et al., 2007; Trop et al., 2010; see Table 5.17). For mammography, the reported estimates for the sensitivity were 25–46% and for the specificity were 95–97%. For MRI, the reported estimates for the sensitivity were 73–100% and for the specificity were 89–98%. [All estimates reported by the earlier study (Kuhl et al., 2005) are outside the confidence intervals of the two published meta-analyses (Warner et al., 2008; Phi et al., 2014). Given the lower expected incidence of breast cancer among women without a BRCA1 or BRCA2 mutation, the PPV of screening with MRI will be much lower than that among women with a BRCA1 or BRCA2 mutation.]

(iii) Mortality reduction

There are no randomized trials assessing the efficacy of adjunct MRI in terms of mortality reduction in women at an increased risk with or without a BRCA gene mutation (Nelson et al., 2013). Several prospective observational studies with long-term follow-up reported on stage distribution and mortality reduction by annual MRI plus mammography screening compared with women without intensified screening.

Three studies analysed the stage distribution of cancers detected in follow-up rounds of intensified screening programmes (Schmutzler et al., 2006; Rijnsburger et al., 2010; Passaperuma et al., 2012). In two of the studies (Schmutzler et al., 2006; Rijnsburger et al., 2010), an increase of N0 stages was reported (N0 stages of 67% vs 52% and 83% vs 56%, respectively). In the third study (Passaperuma et al., 2012), a significant reduction of late stages from 6.6% to 1.9% with intensified screening was observed.

Prospective studies assessing the effectiveness of adjunct MRI in terms of mortality reduction are summarized in Table 5.19. In a four-country study (England, the Netherlands, Norway, and Scotland), the 5-year survival was assessed for 249 women (205 non-BRCA1/2 mutation carriers with a family history of breast cancer, 36 BRCA1 mutation carriers, and 8 BRCA2 mutation carriers) prospectively diagnosed with breast cancer during screening (Møller et al., 2002). All women were under breast cancer surveillance at a dedicated clinic, including annual mammography and CBE, and were diagnosed with breast cancer in this setting. The 5-year survival was 63% for women with a BRCA1 mutation compared with 91% in the women with a family history of breast cancer and without a known BRCA1/2 mutation.

In 2001, as part of a national initiative, women in Norway with a BRCA1 mutation were offered annual breast screening with MRI in addition to mammography. The observed 5-year
<table>
<thead>
<tr>
<th>Reference</th>
<th>Study period and location</th>
<th>Study population</th>
<th>Study design</th>
<th>Main outcome parameters</th>
<th>Percentage survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Møller et al. (2002)</td>
<td></td>
<td>249 women (205 non-BRCA1/2 mutation carriers with FHBC, 36 BRCA1 mutation carriers, and 8 BRCA2 mutation carriers) in 4 countries or regions (England, the Netherlands, Norway, and Scotland)</td>
<td>Women screened with M combined with CBE and diagnosed prospectively; comparison of 5-yr survival between BRCA1/2 mutation carriers and non-carriers with FHBC</td>
<td>5-yr survival: BRCA1 mutation carriers 63% Non-carriers with FHBC 91%</td>
<td>P = 0.04</td>
</tr>
<tr>
<td>Møller et al. (2013)</td>
<td></td>
<td>802 women with a BRCA1 mutation</td>
<td>Women screened with M + MRI for a mean of 4.2 yr and diagnosed prospectively; assessment of the impact of programme on 5-yr and 10-yr survival</td>
<td>5-yr survival 75% (95% CI, 56–86%) 10-yr survival 69% (95% CI, 48–83%)</td>
<td></td>
</tr>
<tr>
<td>Rijnsburger et al. (2010)</td>
<td>Netherlands, 1999–2006</td>
<td>2157 women with &gt; 15% cumulative risk of breast cancer: gene mutation carriers (n = 599) and FHBC with moderate or high risk (n = 1558)</td>
<td>Women screened with biannual CBE and annual M + MRI and diagnosed prospectively; assessment of overall survival at 6 yr</td>
<td>6-yr survival: BRCA1/2 mutation carriers (n = 42) Familial groups (n = 43) 92.7% (95% CI, 79.0–97.6%) 100%</td>
<td></td>
</tr>
<tr>
<td>Passaperuma et al. (2012)</td>
<td>United Kingdom, 1997–2009</td>
<td>496 women with a known BRCA1/2 mutation, of whom 380 had no previous cancer history, aged 25–65 yr</td>
<td>Women screened with annual M + MRI and diagnosed prospectively; assessment of survival (n = 54)</td>
<td>8-yr survival 1 out of 28 BRCA1 mutation carriers with invasive breast cancer died of breast cancer</td>
<td></td>
</tr>
<tr>
<td>Evans et al. (2014)</td>
<td>1990–2013</td>
<td>MRI + M cohort: two prospective cohorts of 959 (647 + 312) women with proven or likely BRCA1/2 or p53 mutations (25% mutation-negative) M-only cohort: prospective cohort of 1223 women with BRCA1/2 mutation or at equivalent risk of breast cancer, aged ≤ 55 yr (24% mutation-negative) Unscreened cohort: retrospective cohort of 557 women with BRCA1/2 mutation identified from the Manchester genetic database as having been diagnosed with breast cancer, aged ≤ 55 yr</td>
<td>MRI + M cohort: screened annually with MRI + M either simultaneously (cohort 1) or 6 mo apart (cohort 2) M-only cohort: screened with M only [annually] Unscreened cohort: identified retrospectively as diagnosed with breast cancer and not having undergone intensive surveillance (a subset aged 50–55 yr had received 3-yearly mammography) 10-yr survival analysis</td>
<td>10-yr survival among BRCA1/2 mutation carriers only: MRI + M 95.3% M 87.7% NS when compared with no screening NS when compared with MRI + M 73.7% MRI + M vs no screening: P = 0.002</td>
<td></td>
</tr>
</tbody>
</table>

CBE, clinical breast examination; CI, confidence interval; FHBC, family history of breast cancer; M, mammography; mo, month or months; MRI, magnetic resonance imaging; NS, not statistically significant; yr, year or years.
breast cancer-specific survival for breast cancer patients with a \(BRCA1\) mutation was 75% (95% CI, 56–86%) and the 10-year survival was 69% (95% CI, 48–83%) (Møller et al., 2013). These results are in contrast with those of two other recent studies (Rijnsburger et al., 2010; Passaperuma et al., 2012). In one study (Rijnsburger et al., 2010), the estimated overall survival at 6 years in \(BRCA1/2\) mutation carriers was 92.7% (95% CI, 79.0–97.6%). In the other study (Passaperuma et al., 2012), out of 28 previously unaffected women with a \(BRCA1\) mutation diagnosed with invasive breast cancer, only 1 died after relapse. [The Working Group noted that the study of Møller et al. (2013) included only women with a \(BRCA1\) mutation, whereas the other two studies also included women with \(BRCA2\) mutations, which could explain the difference in outcome.]

In a recent publication (Evans et al., 2014), a survival analysis was conducted between \(BRCA1/2\) mutation carriers screened with MRI plus mammography and unscreened \(BRCA1/2\) mutation carriers (Table 5.19). There were no differences in 10-year survival between the groups screened with MRI plus mammography and with mammography only, but survival was significantly higher in the group screened with MRI plus mammography (95.3%) compared with the unscreened cohort (73.7%; \(P = 0.002\)). After adjustment for age at diagnosis, this difference was still statistically significant (HR, 0.13; 95% CI, 0.032–0.53). [In this study, there were no deaths among the 21 \(BRCA2\) carriers who received adjunct MRI, indicating that there might be differences in growth time between \(BRCA1\) and \(BRCA2\) tumours.]

(iv) False-positive recall rates

The low specificity linked to screening with mammography plus MRI implies that after several screening rounds a significant percentage of screenees will have experienced either a recall or an image-guided (often MRI-guided) biopsy or will have undergone short-term follow-up (Hoogerbrugge et al., 2008). In one systematic review on the adverse effects of adjunct MRI in the screening of women at an increased risk of breast cancer (Lord et al., 2007), there was a 3–5-fold higher risk of patient recall for investigation of false-positive results compared with that of mammography alone.

(b) Ultrasonography

Overall, the sensitivity of ultrasonography for the screening of women at an increased risk of breast cancer is comparable to or lower than that of mammography, and it is always lower than that of MRI (Warner et al., 2004; Kuhl et al., 2005, 2010; Cortesi et al., 2006; Lehman et al., 2007; Riedl et al., 2007; Weinstein et al., 2009; Trop et al., 2010; Sardanelli et al., 2011; Berg et al., 2012b; Table 5.17).

(c) Clinical breast examination

As part of the screening programme offered to women at an increased risk of breast cancer with and without a \(BRCA1\) or \(BRCA2\) mutation, CBE is offered in some settings in addition to mammography and/or MRI. The evidence on the topic was recently reviewed (Roeke et al., 2014), including seven studies (Tilanus-Linthorst et al., 2000; Warner et al., 2001, 2004; Kuhl et al., 2010; Rijnsburger et al., 2010; Trop et al., 2010; Sardanelli et al., 2011). The percentage of breast tumours detected by CBE varies from 0 out of 120 (0%) (Warner et al., 2001, 2004; Kuhl et al., 2010; Trop et al., 2010; Sardanelli et al., 2011) to 1 out of 260 (0.04%) (Tilanus-Linthorst et al., 2000) and 3 out of 97 (3.1%) (Rijnsburger et al., 2010) screen-detected cancers. [These latter two studies reported lower screen detection by mammography and/or MRI compared with studies in which no additional cases were detected by CBE. Furthermore, it is not clear whether CBE was performed blinded for the other tests, or whether these cases were detected during the screening or between the screening rounds, as most studies had annual screening.
with MRI plus mammography (with or without ultrasonography) and biannual screening with CBE.]

### 5.6.2 Personal history of invasive breast cancer or DCIS

Women with a personal history of invasive breast cancer or DCIS are at an increased risk of developing breast cancer. This section reviews the evidence on the performance of screening with mammography and on whether adjunct ultrasonography or MRI improves screening performance in these women (Table 5.20).

Women with a personal history of breast cancer are at an increased risk of ipsilateral or contralateral breast recurrence, or of a second primary breast cancer. Several studies have shown that a follow-up surveillance programme, including annual mammography, may be considered beneficial to these patients (Ciatto et al., 2004b; Lash et al., 2007; Lu et al., 2009). Only studies that included a comparison group were considered by the Working Group.

One large multicentre cohort study affiliated with the Breast Cancer Surveillance Consortium assessed the accuracy and outcomes of mammography screening in women with a personal history of breast cancer compared with those without such a history (Houssami et al., 2011; Table 5.20). Mammography data of women with a personal history of early-stage breast cancer (58,870 mammograms in 19,078 women) were matched on age, breast density, and year of screening to women without a personal history of breast cancer (58,870 mammograms in 55,315 women). Mammography screening in women with a personal history of breast cancer had lower sensitivity and specificity and a higher interval cancer rate, but a similar proportion of detected early-stage disease, compared with that in women without such a history (Houssami et al., 2011).

In a large study on the detection of breast cancer with the addition of annual screening with ultrasonography or a single screening with MRI to mammography in women at an increased risk, about 50% of the women had a personal history of breast cancer, and at baseline, about 55% of the women had a visually estimated breast density at scan of more than 60% (Berg et al., 2012b; Table 5.20). In this study, 111 cancers were detected: 33 with mammography only, 32 with ultrasonography only, and 26 by the combination of mammography and ultrasonography. In a substudy, after three rounds of mammography and ultrasonography, 9 additional cancers were detected with MRI. Overall, adding ultrasonography to mammography gave a statistically significant increase in sensitivity of the screening (first round, 55.6% vs 94.4%; subsequent rounds, 52% vs 76%) as well as a statistically significant increase in the recall rate (first round, 11.5% vs 26.6%; subsequent rounds, 9.4% vs 16.8%) (Berg et al., 2012b). When women with a personal history of breast cancer were compared with those without such a history, there were no statistically significant differences in yield between the two groups. However, the increase in the recall rate due to adjunct ultrasonography was statistically significantly smaller in the group of women with a personal history of breast cancer compared with those without such a history.

In a substudy in which MRI was added to the combination of mammography and ultrasonography, the sensitivity increased from 43.8% to 68.8%, whereas the recall rate increased from 16.3% to 36.3% (Berg et al., 2012b; Table 5.20). [The low sensitivity of the combined mammography and ultrasonography screening compared with the whole study might indicate an overselection of women with dense breast tissue in this substudy. The change in the recall rate due to supplementary MRI was statistically significantly higher in the group of women with a personal history of breast cancer compared with those without such a history. In this study, at baseline,
Table 5.20 Studies of the effects of screening in women with at least one risk factor for breast cancer

<table>
<thead>
<tr>
<th>Study</th>
<th>Study design</th>
<th>Study population (N)</th>
<th>Main outcome parameters</th>
<th>Results for main outcome parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Screening of women with a personal history of invasive breast cancer or DCIS (PHBC)</strong></td>
<td>Multicentre Cohort study Annual M Breast Cancer Surveillance Consortium</td>
<td>58,870 screening M in 19,078 women with PHBC 58,870 screening M in 55,315 women without PHBC</td>
<td>Sens (%):</td>
<td>PHBC: 65.4 (95% CI, 61.5–69.0) Non-PHBC: 76.5 (95% CI, 71.7–80.7) Spec (%):</td>
</tr>
<tr>
<td>Houssami et al. (2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Berg et al. (2012b)</strong></td>
<td>Multicentre 2004–2006 Single reading Annual M and US Included women with PHBC and/or dense breasts</td>
<td>1,426 women with PHBC 1,236 women without PHBC</td>
<td>Cancer detection (N):</td>
<td></td>
</tr>
<tr>
<td>ACRIN 6666</td>
<td></td>
<td></td>
<td>All women: 111 M only: 33 US only: 32 M + US: 26 Screening with M + US: PHBC: 59 No PHBC: 52 NS</td>
<td>Increase in cancer detection when adding US to M: Similar in both PHBC and non-PHBC patients Recall rate (%):</td>
</tr>
</tbody>
</table>
### Table 5.20 (continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>Study design</th>
<th>Study population (N)</th>
<th>Main outcome parameters</th>
<th>Results for main outcome parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Berg et al. (2012b)</strong>&lt;sup&gt;a&lt;/sup&gt; &lt;br&gt;ACRIN 6666</td>
<td>Multicentre 2004–2008 &lt;br&gt;Single reading &lt;br&gt;Annual M + US, extended with a single MRI screening &lt;br&gt;Included women with PHBC and/or dense breasts</td>
<td>275 women with PHBC &lt;br&gt;336 women without PHBC</td>
<td>Cancer detection rate (/1000 screens): &lt;br&gt;PHBC 7.3 &lt;br&gt;No PHBC 26.7 &lt;br&gt;&lt;small&gt;&lt;em&gt;P = 0.063&lt;/em&gt;&lt;/small&gt;</td>
<td>9 out of 25 cancers detected with MRI, after M + US &lt;br&gt;Recall rate (%): &lt;br&gt;M + US 16.3 &lt;br&gt;M + US + MRI 36.3 &lt;br&gt;&lt;small&gt;&lt;em&gt;P &lt; 0.001&lt;/em&gt;&lt;/small&gt; Increase in recall rate when adding MRI to US + M: &lt;br&gt;PHBC 17.1 &lt;br&gt;No PHBC 27.3 &lt;br&gt;&lt;small&gt;&lt;em&gt;P = 0.002&lt;/em&gt;&lt;/small&gt;</td>
</tr>
<tr>
<td><strong>Houssami et al. (2014b)</strong></td>
<td>Multicentre 1996–2010 &lt;br&gt;Cohort study &lt;br&gt;Breast Cancer Surveillance Consortium</td>
<td>LCIS or ALH: 2505 screens &lt;br&gt;Reference population: 12 525 screens</td>
<td>Sens (%): &lt;br&gt;LCIS or ALH 76.1 (61.2–87.4) &lt;br&gt;Matched group 82.3 (70.5–90.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Houssami et al. (2014b)</strong></td>
<td>Multicentre 1996–2010 &lt;br&gt;Cohort study &lt;br&gt;Breast Cancer Surveillance Consortium</td>
<td>ADH or AH: 6225 screens &lt;br&gt;Reference population: 31 125 screens</td>
<td>Sens (%): &lt;br&gt;ADH or AH 81.0 (70.9–88.7) &lt;br&gt;Matched group 82.6 (76.0–88.1) &lt;br&gt;Spec (%): &lt;br&gt;ADH or AH 86.2 (85.3–87.0) &lt;br&gt;Matched group 90.2 (89.9–90.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Sung et al. (2011)</strong></td>
<td>Single centre 2003–2008 &lt;br&gt;Retrospective study of women with LCIS</td>
<td>840 MRI in 220 women; 670 were routine screens</td>
<td>Cancers diagnosed (N): &lt;br&gt;M alone 5 &lt;br&gt;MRI alone 12 &lt;br&gt;Sens M (%) 36 (13–65) &lt;br&gt;Sens MRI (%) 71 (42–91) &lt;br&gt;Spec M (%) 90 (85–94) &lt;br&gt;Spec MRI (%) 76 (70–82)</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Study design</td>
<td>Study population (N)</td>
<td>Main outcome parameters</td>
<td>Results for main outcome parameters</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
</tr>
<tr>
<td>Friedlander et al. (2011)</td>
<td>Single centre 1996–2009 Retrospective study of women with LCIS</td>
<td>307 MRI in 133 women; all were routine screens</td>
<td>% (N) of women with biopsy recall % (N) of women with malignant findings</td>
<td>20.3% (27/133) 4% (5/133)</td>
</tr>
<tr>
<td>Port et al. (2007)</td>
<td>Single centre 1999–2005 Retrospective study of women with LCIS or AH</td>
<td>182 women screened with annual M 196 women screened with annual M and adjunct MRI</td>
<td>% (N) of women with screen-detected and interval cancer % (N) of women with biopsy recall</td>
<td>In both groups there were 2.5% (5) screen-detected cancers and 1% (2) interval cancers</td>
</tr>
<tr>
<td>King et al. (2013)</td>
<td>Single centre 1999–2009 Prospective study of women with LCIS</td>
<td>4321 women screened with annual M 455 women screened with annual M and adjunct MRI</td>
<td>Cancer detection rate (%): M only M + MRI Characteristics of tumours</td>
<td>13% 13% MRI was not associated with earlier stage, smaller size, or node-negativity</td>
</tr>
</tbody>
</table>

* The study by Berg et al. (2012b) included an MRI substudy. These results are presented here separately.

ADH, atypical ductal hyperplasia; AH, atypical hyperplasia of the breast; ALH, atypical lobular hyperplasia; CI, confidence interval; DCIS, ductal carcinoma in situ; FH, family history suspicious for an increased risk of breast cancer; LCIS, lobular carcinoma in situ; M, mammography; MRI, magnetic resonance imaging; NS, not significant; PHBC, personal history of breast cancer; Sens, sensitivity; Spec, specificity; US, ultrasonography.
about 55% of the women had a visually estimated breast density at scan of more than 60%.

5.6.3 Lobular neoplasia or atypical proliferations

Women with lobular neoplasia or atypical proliferations are estimated to be at an increased risk of developing breast cancer (Collins et al., 2007; Tice et al., 2013). One large study affiliated with the Breast Cancer Surveillance Consortium assessed the accuracy and outcomes of screening women with LCIS, atypical lobular hyperplasia, atypical ductal hyperplasia, or atypical hyperplasia compared with those without such lesions (Houssami et al., 2014b; Table 5.20). The cancer rates in the cohorts of women with LCIS or with atypical lobular hyperplasia were 2–3 times that in the reference cohort, and the cancer rate in the cohort of women with atypical ductal hyperplasia was 3–4 times that in the reference cohort. There were no statistically significant differences in sensitivity between the four cohorts. However, mammography screening of women with LCIS, atypical lobular hyperplasia, atypical ductal hyperplasia, or atypical hyperplasia resulted in lower specificities and higher interval cancers rates compared with their referent population. [The higher interval cancer rates partly reflect the higher underlying breast cancer risk.]

A few studies have examined the sensitivity of MRI in screening women with LCIS (Friedlander et al., 2011; Sung et al., 2011; King et al., 2013) and those with LCIS or atypical hyperplasia (Port et al., 2007). In the two studies that did not have a comparison group, high sensitivities were reported for MRI screening in women with LCIS (Friedlander et al., 2011; Sung et al., 2011). [The Working Group noted that in the study of Sung et al. (2011), only 80% of the screens were routine screens; the remaining 20% had non-specified indications, and the indications for the routine screens were not specified. Similarly, the study of Friedlander et al. (2011) reported only results from routine breast MRI screens, but the indications for the routine screens were not specified. The estimated sensitivities are thus likely to be biased in both studies.]

In the other two studies (Port et al., 2007; King et al., 2013), women with high-risk lesions (LCIS and/or atypical hyperplasia) screened annually with mammography plus MRI were compared with women with high-risk lesions screened with annual mammography only. [In both studies, women with high-risk lesions selected to undergo adjunct MRI screening were younger and had stronger family histories of breast cancer compared with those screened by mammography only.] In both studies, adjunct MRI screening generated more follow-up biopsies compared with mammography alone.

5.7 Clinical breast examination

5.7.1 Preventive effects of clinical breast examination

Randomized trials of CBE versus no screening have shown a significant shift from late-stage (T3/T4) to early-stage (T1/T2) breast cancers in the intervention arm (Pisani et al., 2006; Mittra et al., 2010; Sankaranarayanan et al., 2011; see Section 4.3). Compliance with screening is one of the factors that determine effectiveness. In all three trials of CBE, the compliance with screening was high (> 85%), indicating acceptance of the procedure and ease of administering CBE. Access to care after recall and diagnosis is of paramount importance in the success of any screening trial, as is evident in the two randomized trials in India of CBE versus no screening (Mitra et al., 2010; Sankaranarayanan et al., 2011). This was the major reason that the study in the Philippines was discontinued (Pisani et al., 2006). The active intervention was stopped after the first screening round due to poor compliance (35% of screen-positive women) of participants
with clinical follow-up for confirmation of diagnosis and treatment.

5.7.2 Adverse effects

In the Mumbai study, the recall rate after CBE was 0.71%. Out of 153,130 screens by CBE, 1539 women were recalled for diagnostic investigations and 81 were confirmed to have invasive cancers (Mittra et al., 2010).

Some harm of CBE may be attributed to pain or discomfort. Baines et al. (1990) carried out a survey of women who participated in the CNBSS to document women’s attitudes to screening by CBE and mammography. Of those who underwent CBE, 8.4% reported moderate discomfort and 2.1% extreme discomfort, whereas of those who underwent mammography, 36.2% reported moderate discomfort and 8.7% extreme discomfort.

5.7.3 Cost–effectiveness analysis

Determining the cost–effectiveness of CBE alone is difficult because no trial has reported independent efficacy of CBE versus no screening. There have been many reports of cost–effectiveness analyses (Okonkwo et al., 2008; Ahern & Shen, 2009) on screening with reference to CBE. [The Working Group noted that most reports made assumptions about mortality reductions to simulate or estimate cost–effectiveness that were not realistic. It may be appropriate to look at cost analysis instead.] The cost of delivering breast cancer screening by CBE is less than one third that of mammography (Sarvazyan et al., 2008).

5.8 Breast self-examination

5.8.1 Preventive effects of teaching breast self-examination

Randomized trials and multiple observational studies have generally shown little or no reduction in mortality from breast cancer in women who practised BSE (see Section 4.4). If BSE is to have an effect on breast cancer mortality, it will have to be practised competently, and more frequently than in the Shanghai trial (see Section 4.4). Table 5.21 shows results of 11 surveys on BSE practice, based on self-reports, conducted primarily in countries with limited resources. Proficiency of BSE practice was not assessed in any of the studies. [It is unlikely that the proportion of women who reported practising BSE in any of the studies was sufficiently high to result in a meaningful reduction in breast cancer mortality rates in the populations surveyed.]

Results of two studies of BSE practice before and after BSE instruction have been reported. Approximately 1000 women aged 30–50 years in Madhya Pradesh, India, attended BSE instruction sessions in which a film was shown, reinforced by a lecture with flip charts showing proper technique, and including a question-and-answer period (Gupta et al., 2009). None of the women were practising BSE before the instruction. Two months after the instruction, 53% reported practising BSE regularly. [It is uncertain what regular practice means in just 2 months of alleged practice.] In Lower Saxony, Germany, women invited to instruction sessions received a lecture on BSE techniques followed by individual BSE training by a gynaecologist (Funke et al., 2008). The self-reported prevalence of monthly BSE practice was 21% before the instruction and 62% 1 year after the instruction. Proficiency of BSE practice was not assessed in either of these studies. [It is therefore unclear whether a sufficient number of women in either study practised BSE with sufficient competence and frequency to result in a reduction in mortality from breast cancer.] In three studies, BSE practice after BSE instruction was compared with BSE practice in a control group that did not receive instruction. In a study in rural women in the Republic of Korea (Lee et al., 2003), women were given BSE instruction after appraisal of their individual risk on the basis of a questionnaire. Three months after the
instruction, 30.5% of the women reported practising BSE regularly, compared with 10.2% in a control group. In a study of Latinas in the USA (Jandorf et al., 2008), women were randomized to a group receiving information on BSE and CBE or to a control group. Telephone interviews 2 months after the instruction revealed that 45% of the women in the instruction group practised BSE compared with 27% in the control group. [Proficiency was not assessed in either of these studies.] In a BSE instruction programme in Ribe County, Denmark, up to 20 women at a time attended an intensive BSE training session lasting up to 2 hours that included videos as well as individual instruction on breast models and on the women’s own breasts (Sørensen et al., 2005). An unreported number of years later (< 5 years), a questionnaire was mailed to the women who had participated and to a sample of women in the county who had not participated; 485 (77%) and 313 (53%) responded, respectively. Women were asked about frequency of BSE practice and

### Table 5.21 Percentage of women who reported practising breast self-examination in surveys conducted in selected countries

<table>
<thead>
<tr>
<th>Country Reference</th>
<th>Age of participants (years)</th>
<th>Definition of sample</th>
<th>Definition of BSE practice</th>
<th>Number of women</th>
<th>Percentage practising BSE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Africa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethiopia</td>
<td>16–37</td>
<td>Health extension workers</td>
<td>Regularly</td>
<td>390</td>
<td>14.4%</td>
</tr>
<tr>
<td>Azage et al. (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nigeria</td>
<td>20–65</td>
<td>Market workers</td>
<td>Regularly</td>
<td>238</td>
<td>0.4%</td>
</tr>
<tr>
<td>Obaji et al. (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>East and South Asia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td>Mean, 40.5 (SD, 15.5)</td>
<td>Rural women</td>
<td>Classified as good</td>
<td>86</td>
<td>7.0%</td>
</tr>
<tr>
<td>Rosmawati (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td>Not given</td>
<td>Teachers</td>
<td>Regular</td>
<td>425</td>
<td>19.0%</td>
</tr>
<tr>
<td>Parsa et al. (2011)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>Mean, 32.4 (SD, 10.9)</td>
<td>Outpatients</td>
<td>Regularly</td>
<td>373</td>
<td>25.9%</td>
</tr>
<tr>
<td>Sobani et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>20–64</td>
<td>Rural women</td>
<td>Monthly in past year</td>
<td>705</td>
<td>49.3%</td>
</tr>
<tr>
<td>Satitvipawee et al. (2009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>West Asia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iraq</td>
<td>18–62</td>
<td>Women affiliated with universities</td>
<td>Ever practised</td>
<td>858</td>
<td>53.9%</td>
</tr>
<tr>
<td>Alwan et al. (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Islamic Republic of Iran</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khalili &amp; Shahnazi (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pakistan</td>
<td>Mean, 29 (SD, 5.6)</td>
<td>Health-care workers</td>
<td>Monthly</td>
<td>246</td>
<td>17.0%</td>
</tr>
<tr>
<td>Güleser et al. (2009)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West Bank and Gaza Strip</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Azaiza et al. (2010)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Europe</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poland</td>
<td>22–45</td>
<td>Nursing students, hospital workers, and gynaecological outpatients</td>
<td>Regularly</td>
<td>492</td>
<td>33.7%</td>
</tr>
<tr>
<td>Lepecka-Klusek et al. (2007)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

BSE, breast self-examination; SD, standard deviation.
whether they practised the various components of the BSE technique that was taught (positioning, use of mirror, and palpation pattern). On the basis of their answers, women were classified as performing BSE correctly, nearly correctly, or partly correctly. A higher percentage of women in the intervention group than in the control group practised BSE monthly (30.7% vs 21.1%) and practised it correctly or nearly correctly (27.6% vs 10.2%).

[The level of BSE practice in women taught BSE in all five of the evaluations of BSE instruction summarized in this section was lower than that in the trial in Shanghai, which showed no reduction in breast cancer mortality from BSE instruction. It is therefore reasonable to conclude that the level of BSE activity that was probably achieved in these studies was insufficient to have a meaningful impact on breast cancer mortality rates in the populations in which they were conducted. All of these studies except one were conducted in developed countries in which women, like the women in the Shanghai trial, had reasonable access to care, and in which women would be expected to seek medical attention for breast symptoms suggestive of breast cancer early in the course of the disease. The study in India may be an exception. In that country, many women with breast cancer typically present with advanced disease. It is unknown whether breast cancer mortality would be reduced if women in that country could be motivated to practise BSE on a regular basis, as was reported in the study by Gupta et al. (2009), and to do so competently.]

5.8.2 Adverse effects

In both randomized trials of BSE, more women in the instruction group than in the control group found breast lumps that required further evaluation and that were subsequently confirmed as not being breast cancer (Section 4.4). In the trial in St Petersburg (Semiglazov et al., 2003), nearly twice as many women were referred for further evaluation in the instruction group than in the control group; in the Shanghai trial, 80% more women in the intervention group than in the control group were found to have a histologically confirmed benign lesion (Thomas et al., 2002). Such false-positives on screening can produce considerable anxiety, and the further evaluation of suspicious findings is not a trivial expense. Given that there is no proven benefit of BSE in reducing mortality from breast cancer, the risk–benefit ratio is very high.

5.8.3 Cost–effectiveness analysis

Given that there is no good evidence that BSE, as it has been reported to be practised in studies to date, contributes to a reduction in mortality from breast cancer, there can be no estimate of the cost per life year gained by practising BSE. Based on data from the study in Ribe County, Denmark, Sørensen & Hertz (2003) estimated the cost per avoided cancer with spread to lymph nodes to be €15 410 and the cost of avoiding a cancerous tumour larger than 20 mm to be €16 318. [In their model, they assumed that there was considerable shift to a lower stage as a result of BSE practice, but as discussed in Section 4.4, the evidence for this is questionable and inconsistent, and the results of their estimates are highly dependent on the assumptions that they made as to the magnitude of the stage shift. They used only the cost of the BSE programme in their model. Their estimates did not take into account the costs of diagnostic confirmation or of changes in treatment if there is a stage shift at the time of diagnosis by BSE practice. If there truly is a stage shift, then this could result in less aggressive and less costly treatment, which would be a benefit even in the absence of a reduction in mortality. However, given the uncertainties as to any beneficial effects of BSE, no meaningful cost–effectiveness estimates are possible.]
References


Broeders MJ, Veerbeek AL, Straatman H, Peer PG, Jong PC, Beex LV et al. (2002). Repeated mammographic screening reduces breast cancer mortality along the


Breast cancer screening


Breast cancer screening


Kalager M, Adam HO, Bretthauer M, Tamimi RM (2012). Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian...


Lampic C, Thurfjell E, Bergh J, Sjödén PO (2001). Short- and long-term anxiety and depression in women recalled...


Breast cancer screening


Semiglazov VF, Manikhas AG, Moiseenko VM, Protosenko SA, Kharikova RS, Selezniev IK et al. (2003). Results of a prospective randomized investigation [Russia (St Petersburg)/WHO] to evaluate the significance of self-examination for the early detection of breast cancer [in Russian]. Vopr Onkol, 49(4):434–41. PMID:14569932


Spanu A, Chessa F, Meloni GB, Sanna D, Cottu P, Manca A et al. (2008). The role of planar scintimammography with high-resolution dedicated breast camera in the...


