4. EFFICACY OF BREAST CANCER SCREENING

4.1 Methodological and analytical issues

To evaluate the efficacy of screening, it is important to consider the definitions of efficacy and effectiveness for an intervention, to define outcome measures, and to consider potential biases.

4.1.1 Efficacy versus effectiveness

The term "efficacy" should be distinguished from the term "effectiveness". Efficacy is "the extent to which a specific intervention, procedure, regimen, or service produces a beneficial result under **ideal conditions**" (Porta, 2014), whereas effectiveness is "a measure of the extent to which a specific intervention, procedure, regimen, or service, when deployed in the field in the **usual circumstances**, does what it is intended to do for a specified population" (Porta, 2014). In practice, true efficacy [under ideal conditions] can rarely be estimated. Randomized controlled trials (RCTs), which are conducted to initially assess whether screening works, assess efficacy by estimating a primary outcome, such as reduction in breast cancer mortality in the study arm compared with the control arm. However, the measure of efficacy is limited by the implementation of the intervention and other practical issues – for instance, less than 100% compliance in the study arm and unintended screening in the control arm. Hence, an intention-to-treat analysis of RCTs, i.e. an analysis in which the data are analysed according to the original randomized design, may actually have a limited ability to address efficacy, due to non-ideal circumstances (Gulati et al., 2012).

This section focuses primarily on the assessment of efficacy; methodological issues in the assessment of effectiveness are addressed in Section 5.1.

4.1.2 Primary outcome measures

The primary outcome measure is reduction in breast cancer mortality, although increasing life expectancy or reduction of metastatic disease can also be considered efficacy measures. Given the natural history of the disease, a minimum requirement in addressing efficacy is a sufficiently long follow-up (Hanley, 2011). Some authors have suggested that the use of breast cancer mortality as the end-point of a trial may have led to unreliable estimates of the relative risk reduction, due to possible uncertainties surrounding the determination of breast cancer death (leading to misclassification of deaths), and that the use of all-cause mortality as the end-point of a trial would resolve this bias (Black et al., 2002; Gøtzsche & Jørgensen, 2013). However, others have argued that all-cause mortality is not an appropriate end-point for screening trials for a specific disease (Tabár et al., 2002; Marmot et al.,

2013; Weiss, 2014). Although using all-cause mortality avoids the need to determine cause of death precisely, breast cancer deaths reflect a small fraction of all-cause mortality, and trials of the size needed to have sufficient statistical power to detect the expected small effects of screening on all-cause mortality would be logistically and financially impracticable. A Swedish review, which incorporated all Swedish RCTs of breast cancer screening, showed a 2% non-significant reduction in all-cause mortality (Nyström et al., 2002a), which is in line with the expected 0.94% (Nyström et al., 2002b).

4.1.3 Biases

Several sources of bias have important effects on the estimation of screening efficacy.

The first important bias is lead-time bias. The general concept of screening is that by early detection of disease and subsequent treatment, prognosis is improved and the probability of death from the disease is reduced. The time between screen detection and the point at which a tumour would have presented and been clinically diagnosed (in the absence of screening) is referred to as "lead time" (Cole & Morrison, 1980). The survival time, the time from breast cancer diagnosis to death, of screen-detected cases is increased because of this lead time, even for individuals who do not benefit from screening. Lead-time bias may therefore appear to act in favour of screening, if efficacy is evaluated by survival analyses.

The second important bias is length bias (Cole & Morrison, 1980) (sometimes referred to as length-time bias). The probability of a tumour being detected at screening is (partially) dependent on the growth rate of the tumour, because slow-growing tumours have a longer preclinical detectable phase (sojourn time) and are therefore more likely to be detected than fast-growing tumours. Tumours detected at screening thus reflect a biased sample of preclinical lesions, including slower-growing tumours, which are generally thought to be associated with a better prognosis and therefore longer survival. This again leads to bias apparently in favour of screening. The most extreme form of length bias is referred to as overdiagnosis. Some ductal carcinoma in situ (DCIS) may never progress to invasive cancer or present clinically (in the absence of screening) (Yen et al., 2003), and some invasive cancers may be sufficiently indolent that they would never have presented clinically during the woman's lifetime if they had not been detected by screening (see Section 4.2.3c).

The last important bias in evaluating screening is selection bias. Women attend screening voluntarily, and participants might therefore generally be more health-conscious and have a lower baseline risk of breast cancer than non-participants, although in practice this assumption may not hold true (Paap et al., 2011). The decision to attend screening may also be influenced by certain demographic and social factors (see Section 3.1) that affect disease prognosis, for example familial risk. In RCTs with mortality as the end-point, such a selection may hamper the generalizability of the results.

Evaluations of efficacy and effectiveness must control for the above-mentioned biases if they are to provide credible estimates. To eliminate lead-time and length bias, differences in breast cancer mortality rates (between the trial arms or different populations) should be the end-point of a study rather than survival, because survival time in cancer patients is extended due to lead time and is more favourable due to length-biased sampling. Selection bias can partially be quantified by comparing non-participants with historical or recent data on mortality or risk factors and can, perhaps, be controlled for by adjusting for risk factors or their surrogates (e.g. socioeconomic status; Allgood et al., 2008) or by the application of an empirically derived adjustment factor (Paap et al., 2014). In addition, it has been argued that any bias due to selection

for screening is likely to be small in organized programmes with invitation schemes based on population registries and with high attendance rates (van Schoor et al., 2011a, b).

4.1.4 Use of randomized controlled trials

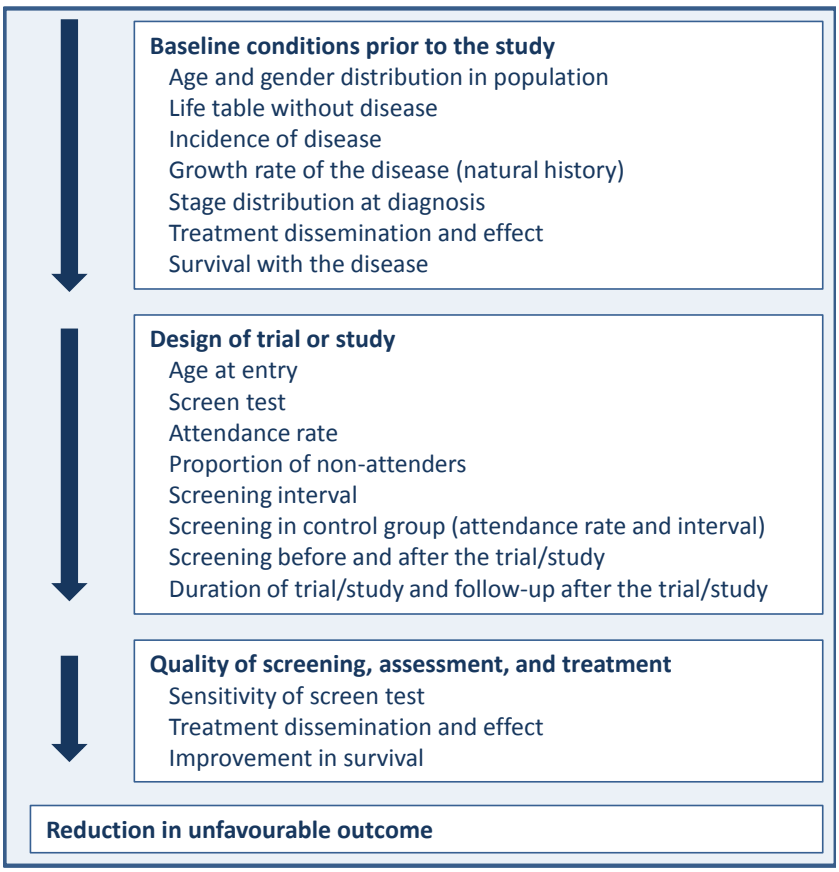
Reduction in breast cancer mortality in women offered screening relative to women not offered screening is the appropriate measure of benefit of an RCT. Lead-time and length bias are then eliminated in the analyses. Women are followed up from the time of randomization instead of from the time of diagnosis, which avoids lead time, and all deaths from breast cancer that occur during the follow-up period are included in the analysis. The RCTs of breast cancer screening are evaluated in accordance with the intention-to-screen principle, taking into account in the intervention group both women who accept the invitation to screening and women who decline the invitation. The resulting point estimate of reduction in breast cancer mortality therefore does not evaluate the screened groups of individuals only.

In RCTs, participants are randomly assigned to either the intervention group or the control group to prevent confounding at baseline, accounting for both observable characteristics and unknown confounders. However, even well performed randomization schemes may not prevent potential imbalances completely. To take into account possible differences in risk factors for death from breast cancer between the intervention group and the control group, an assessment should preferably be made of the distribution of risk factors in both groups at trial entry, which would permit adjustment in the analysis (although most known risk factors for breast cancer seem to have limited predictive value). If individual randomization is not feasible, for example when the same clinician would be required to use a simple screening test in one individual and not use it in another, randomization by cluster is an alternative. Both types of randomization have been used in the RCTs of breast cancer screening. Recruitment and randomization are less complex with cluster randomization, but an equal distribution of risk factors between the intervention group and the control group is less likely to be achieved than with individual randomization. Furthermore, subjects with a previous diagnosis of breast cancer at the time of randomization are, ideally, excluded from the trial. Whereas a previous diagnosis can be determined more easily in RCTs with individual randomization, this may be more difficult to achieve beforehand with cluster randomization. An important advantage of cluster randomization is that contamination of (screening in) the control group may be reduced.

As mentioned above, the screening effect in RCTs is dependent on, among other things, the compliance in the intervention group and the limitation of contamination of the control group. Low compliance reduces the estimate of effect and must therefore be reported. Screening of controls by services outside of the trial will also dilute the effect of screening on breast cancer mortality. Possible contamination of the control group is often difficult to measure, especially because mammography is also used for clinical diagnosis of breast cancer and this use may not be easily distinguished from use for opportunistic screening. Methods to adjust for contamination and poor compliance have been proposed (Cuzick et al., 1997; Baker et al., 2002). Furthermore, unless the breast cancer mortality analysis is limited to those diagnosed with breast cancer during the screening phase of the trial period, with longer follow-up, screening of the control group can influence the observed difference between the intervention group and the control group.

The difference in outcome between the groups of subjects randomized is further determined by a large number of varying factors. The age groups at entry, screening interval, attendance

Fig. 4.1 Trajectory of a screening outcome



From de Koning (2009).

of trial screening, and opportunistic screening in both women randomized to screening and control women all influence the ultimate extent of the effects. Such "analyses per protocol" were not routinely conducted in the available trials. Through modelling, it has been shown that these relatively simple differences alone could make one trial exhibit a 25% greater effect than another (de Koning et al., 1995).

However, for estimating the magnitude of (true) efficacy, it is equally important to consider how much earlier the diagnosis was made as a result of screening and the effect this has. Therefore, the more important questions relate to the quality of the screening, how many women were referred for further examination, and how many tumours were detected and at which stage. The baseline conditions before the study, or in this case those of the control group, are also significant. If women in one region on average receive health care at an earlier stage, this can mean that the difference between "early" and "late" (read: intervention group compared with control group) is smaller in one region than in another, even if the quality of screening and therapy may be the same. In standard meta-analyses, all of these differences are ignored, and modelling has been proposed to estimate the impact of such effects and to lead to better estimates of "efficacy under ideal circumstances". Fig. 4.1 exemplifies the most important different steps that ultimately lead to the (reported) reduction in the unfavourable outcome of disease – for

example, breast cancer mortality – that should be considered when estimating the true efficacy.

4.1.5 *Use of observational studies in assessing efficacy*

Estimates of screening efficacy from contemporary observational studies may be considered more relevant than those from the RCTs, most of which were initiated in the 1970s or early 1980s. Recent studies can take into account improvements in mammography techniques and in treatment that have occurred over the past 30 years. However, observational studies are prone to the biases discussed above, and adequate control for these biases by design or analysis is difficult. The presence of other potential biases differs between studies and is dependent on the study design, the duration and completeness of follow-up, and, in a case–control study, the definition of exposure to screening. In practice, these observational studies have been used primarily to assess the effectiveness of screening programmes (see Section 5.1.2).

4.2 Mammography

The basic characteristics of the randomized trials of the efficacy of screen-film mammography screening are shown in Table 4.1. All of these trials were considered by the previous IARC Working Group on breast cancer screening (IARC, 2002). All ages given in this section, unless otherwise stated, refer to age at entry into the trial.

4.2.1 *Description of randomized trials*

(a) *Health Insurance Plan trial*

In December 1963, the Health Insurance Plan of Greater New York, USA, had 85 000 female members aged 40–64 years (Shapiro et al., 1966). In 23 of the plan's 31 medical groups, women were individually randomized to annual film mammography screening and clinical breast examination (CBE) for 4 years or to a control arm receiving the usual care within the plan but no screening. Randomization was pair-matched by age, size of the insured family, and employment group through which the family had joined the plan. Of those randomized to screening, 67% attended the first screening round. Although data on risk factors were not collected from all participants, there were no differences between a 10% sample of the examined group, a 20% sample of non-attenders, and a 20% sample of controls with respect to age, socioeconomic status, and history of pregnancies (Shapiro et al., 1988).

Gøtzsche & Olsen (2000) suggested that the exclusions after randomization and the review of causes of death may have led to lack of comparability between the screened and unscreened groups. Miller (2001) advised that the decisions made on the deaths reviewed were entirely masked. [Miller was a member of the death review committee.] A difference in the numbers of women with breast cancer initially excluded from the two arms of the trial arose because previously diagnosed breast cancers were identified in women in the screened group when they attended screening, but this was not possible for the controls. However, the 18-year follow-up enabled identification of deaths from breast cancer in the two groups; determination of the date of diagnosis was then made from hospital records. Women who had died from breast cancers diagnosed before randomization were then excluded.

[The Working Group concluded that the Health Insurance Plan trial was valid and could be included in its overall evaluation of screening by mammography. The technology used produced images of comparable quality to those from screen-film mammography (see Sections 2.1.1 and 2.1.2 for details on the history of screening techniques).]

Table 4.1 Basic characteristics of randomized trials of the efficacy of screen-film mammography screening

Trial, country	Randomization	No. of women	Accrual period for screening		Age at entry (years)	Intervention	No. of mammography views	Screening interval (months)	No. of rounds	Attendance rate at first round (%)	Determination of end-point
			Invited group	Control group							
Health Insurance Plan trial, USA	Individual	60 696	December 1963–June 1966		40–64	M + CBE	2	12	4	67	Independent death review
Malmö I trial, Sweden	Individual	42 283	October 1976–August 1978	October 1992–February 1993	45–70	M	2	18–24	6–8	74	Independent death review Official statistics
Malmö II trial, Sweden	Individual	17 786	September 1978–November 1990	September 1991–April 1994	43–49	M	2	18–24	1–7	75–80	Official statistics
Two-County trial: Kopparberg County, Sweden	Cluster	57 897	July 1977–February 1980	September 1982–December 1986	40–74	M	1	24 (40–49) 33 (≥ 50)	2–4	89	Death review
Two-County trial: Östergötland County, Sweden	Cluster	76 617	May 1978–March 1981	April 1986–February 1988	40–74	M	1	24 (40–49) 33 (≥ 50)	2–4	89	Death review Official statistics
Edinburgh trial, United Kingdom	Cluster	54 643	1978–1985		45–64	M + CBE	2	24	2–4	61	Death certificates
CNBSS 1 trial, Canada	Individual	50 430	January 1980–March 1985		40–49	M + CBE	2	12	4 or 5	100	Independent death review Official statistics
CNBSS 2 trial, Canada	Individual	39 405	January 1980–March 1985		50–59	M + CBE	2	12	4 or 5	100	Independent death review Official statistics
Stockholm trial, Sweden	Cluster	60 117	March 1981–May 1983	October 1985–May 1986	40–64	M	1	28	2	81	Official statistics

Table 4.1 (continued)

Trial, country	Randomization	No. of women	Accrual period for screening		Age at entry (years)	Intervention	No. of mammography views	Screening interval (months)	No. of rounds	Attendance rate at first round (%)	Determination of end-point
			Invited group	Control group							
Gothenburg trial, Sweden	Individual Cluster	51 611	December 1982–April 1984	November 1987–June 1991	39–59	M	2	18	4 or 5	85	Official statistics
United Kingdom Age trial	Individual	160 921	1991–1997	On reaching age 50–52 years	39–41	M	2, first screen 1, subsequently	12	4–7	68	Official statistics

CBE, clinical breast examination; CNBSS, Canadian National Breast Screening Study; M, mammography.

(b) Malmö trials

In the first of two trials in Malmö, Sweden (Malmö I), starting in October 1976 all women born in 1908–1932 were identified from the population register and randomized by a computer program within each birth-year cohort. The resulting lists were divided; the 21 088 women in the first half were invited, and the 21 195 women in the second half served as controls (Andersson et al., 1988). Women were invited to screen-film mammography alone, with two views (craniocaudal and mediolateral oblique) in the first two rounds, and with either both views or only the oblique view, depending on the parenchymal pattern, in the subsequent rounds, every 18–24 months. A single mediolateral oblique view was taken for women whose breasts were mainly fatty on mammography, and both views were taken for women with dense breasts. The attendance rate was higher for the first round (74%) than for subsequent rounds (70%), and was higher among younger women than among older women.

After August 1978, the investigators aimed to continue to recruit women who reached the age of 45 years and to randomize them to either receive or not receive an invitation to mammography. In the second trial (Malmö II), 17 786 women born in 1933–1945 were recruited, with 9574 in the invited group and 8212 in the control group. The randomization and screening procedures were the same as in the first trial, and recruitment continued until 1990 (Andersson & Janzon, 1997).

(c) Two-County trial (Kopparberg and Östergötland)

In 1975, the Swedish National Board of Health and Welfare invited five county councils to start a mammography screening trial. Two counties, Kopparberg (now Dalarna) County and Östergötland County, accepted the invitation. Women in this trial were randomized by cluster within geographical areas (municipalities, parishes, tax districts). The municipalities in Östergötland County were grouped pairwise with respect to the size of the population and geographical characteristics. The more-populated municipalities of Linköping, Norrköping, and Motala were split into six, eight, and two clusters, respectively, of similar size, creating three, four, and one pairs, respectively, to increase the number of clusters. The clusters were randomly allocated to an invitation group or to a control group. A total of 76 617 women aged 40–74 years were randomized to mammography or the usual care (Nyström et al., 2002a). In Kopparberg County, the invited group was planned to be twice as large as the control group. Thus, triplets of geographical areas were identified by dividing each block into three units of roughly equal size, of which two were randomly allocated by local politicians to receive screening and one to the control group. A total of 57 897 women aged 40–74 years were included (Tabár et al., 1985). In total, 77 080 women were randomized to regular invitation to screening (active study population [ASP]) and 55 985 to no invitation (passive study population [PSP]) in 45 geographical clusters (Duffy et al., 2003a). In the ASP, women aged 40–49 years were invited to screening by single-view mammography every 24 months, and those aged 50 years and older were invited on average every 33 months. The overall compliance with the invitations for women aged 40–74 years was 89% for the first screen and 83% for the second screen. Women aged 40–49 years had the highest compliance, 93% for the first screen and 89% for the second screen, and women aged 70–74 years had the lowest compliance, 79% for the first screen and 67% for the second screen (Tabár et al., 1985). Women aged 70–74 years at randomization were not invited to a third screen. The compliance for the third screen was 88% for women aged 40–49 years, 86% for those aged 50–59 years, and 78% for those aged 60–69 years (Tabár et al., 1992).

When this trial was conducted, adjuvant chemotherapy and hormone therapy were not available in Sweden, and therefore they were not used for the treatment of breast cancer cases in the trial (Holmberg et al., 1986, Tabár et al., 1999). Furthermore, because the controls (PSP) were not contacted until a decision was made to screen them at the end of screening of the ASP, no data on breast cancer risk factors were collected to permit confirmation that balance in the distribution of risk factors was achieved by the cluster randomization.

In response to suggestions that there were various potential problems with the randomization in the Two-County trial (Olsen & Gøtzsche, 2001), Nyström et al. (2002a) reported that the breast cancer incidence and mortality rates in the clusters of the screened and control groups in Östergötland County before the trial (1968–1977) were similar. They suggested that there is no reason to believe that the cluster randomization in this component of the trial was biased, as any bias would have manifested in breast cancer incidence and mortality rates. Duffy et al. (2003a) reanalysed the available data, taking into account the cluster randomization. Although there was no significant difference in prior breast cancer mortality between the ASP and PSP clusters, the authors reported an analysis adjusting for prior mortality within clusters. This yielded a significant 27% reduction in mortality in the ASP, a minor dilution of the unadjusted estimate (30%). [This suggested that there was no substantial bias in terms of prior risk of breast cancer mortality as a result of the cluster randomization.]

Issues have been raised about the numbers of cases included in the analyses of the Two-County trial (Zahl et al., 2006). Dean (2007) advised that the analysis of Zahl et al. (2006) was inaccurate with respect to trial dates and did not take into account the staggered entry of districts into the trial (Fagerberg & Tabár, 1988).

Verification of the cause of death is crucial in any trial. Holmberg et al. (2009) characterized and quantified differences in the number of breast cancer cases and deaths identified in the Two-County trial by the local end-point committee compared with the Swedish overview committee. Of the 2615 outcomes included by the local end-point committee or the overview committee, there were 478 (18%) disagreements, of which 82% were due to differences in application of inclusion/exclusion criteria and 18% were due to disagreement with respect to cause of death or vital status at ascertainment. For Östergötland County, the overview committee-based determination of cause of death resulted in a reduction of the estimate of the effect of screening compared with the local end-point committee, but for Kopparberg County the difference was modest.

The Two-County trial was closed after completion of the first round of screening in the PSP; participants in both groups continued with service screening. All cases of breast cancer in both groups diagnosed up to and including the end of the first screen of the PSP were followed up for death from breast cancer (Holmberg et al., 2009).

(d) Edinburgh trial

In Edinburgh, United Kingdom, in 1978–1981, 87 general practitioners' practices, covering 44 268 women aged 45–64 years, were randomized for a breast cancer screening trial (Alexander et al., 1999). The 22 926 women in the practices in the intervention group were invited to participate in a screening programme, which included CBE every year and two-view mammography every 2 years. The 21 342 women in the practices in the control group received only the usual care. Subsequently, additional eligible women who joined these practices and existing patients who reached the age of 45 years were recruited into two further cohorts: 4867 women in 1982–1983 and 5499 women in 1984–1985 (Alexander et al., 1999).

Alexander et al. (1989) reported that the cluster randomization in the Edinburgh trial

resulted in differences by socioeconomic category and also in rates of mortality from all causes between the two comparison groups.

[The Working Group noted concerns about the potential for bias resulting from the cluster randomization procedure. Although the authors adjusted for socioeconomic status in their analyses, it is not clear that this entirely removed the bias. Nevertheless, the Working Group concluded that this trial could be included in the evaluation.]

(e) Canadian National Breast Screening Study trials

The Canadian National Breast Screening Study (CNBSS) was originally designed as a single trial in women aged 40–59 years (Miller et al., 1981), and was managed as such, but after the first mortality reports (Miller et al., 1992a, b), it was regarded as two trials: CNBSS 1, in women aged 40–49 years, and CNBSS 2, in women aged 50–59 years. Women were eligible for the trials if they had not had breast cancer, had not had a mammogram in the previous 12 months, were not currently pregnant, and completed a questionnaire providing full identification and data on risk factors for breast cancer (Miller et al., 1981). Before randomization, all participants gave written informed consent after having been told that they had a 50% chance of having a mammogram. They then received CBE and instruction in breast self-examination (BSE), and the findings were recorded. While the participant remained in the examining room, the examiner went to receive the results of randomization from the centre coordinator, and then told the participant whether she would receive mammography screening. Subsequently, women randomized to screening in both trials were offered annual CBE and mammography (Miller et al., 1992a, b). Control women aged 40–49 years in the CNBSS 1 trial received a questionnaire every year. Control women aged 50–59 years in the CNBSS 2 trial were offered annual CBE.

Women were invited to volunteer to participate in the trials by several methods (Baines et al., 1989) and were recruited in 1980–1985. A total of 50 430 women aged 40–49 years were enrolled in the CNBSS 1 trial, and 39 405 women aged 50–59 years were enrolled in the CNBSS 2 trial. The distribution of breast cancer risk factors in the two groups in both trials was identical, confirming that balance was achieved by randomization (Miller et al., 1992a, b). The treatment administered to breast cancer cases in women aged 40–49 years in the CNBSS 1 trial was evaluated to be compatible with standards then applied in North America for adjuvant chemotherapy and hormone therapy (Kerr, 1991).

For women in the mammography group of the CNBSS 1 trial, full compliance with screening (mammography plus CBE) after the first screen (when compliance was 100% with CBE) varied from 89.4% (for the second screen) to 85.6% (for the fifth screen). In addition, a small proportion (1.7–2.9%) of the women accepted CBE but refused to undergo mammography. More than 90% of the participants in the control group (ranging from 93.3% to 94.9% in the various years) returned their annual questionnaire (Miller et al., 1992a). For women in the mammography group of the CNBSS 2 trial, full compliance with screening after the first screen varied from 90.4% (for the second screen) to 86.7% (for the fifth screen). In addition, a small proportion (1.8–3.2%) of the women accepted CBE but refused to undergo mammography. In the control group, compliance with annual CBE screening varied from 89.1% (for the second screen) to 85.4% (for the fifth screen); questionnaires only were obtained for 2.8–7.0% of the women (Miller et al., 1992b).

Boyd et al. (1993) criticized the process of randomization in the trials, but a systematic external review of the randomization records showed no evidence of subversion of randomization (Bailar & MacMahon, 1997). The mammography equipment used in these trials has also

been criticized (Kopans, 1990, 1993, 2014; Moskowitz, 1992; Kopans & Feig, 1993; Tabár, 2014), and these criticisms have been addressed by the investigators (Miller et al., 1990, 2014a, b).

(f) Stockholm trial

A trial was performed in the south-eastern part of Greater Stockholm, Sweden, in which about 60 000 women aged 40–64 years in March 1981 were randomized by day of birth to invitation to mammography screening or to a control group (Frisell et al., 1986). Women born on days 1–10 and 21–31 of the month were invited to screening, and women born on days 11–20 constituted the control group. Attendance was 81% for the first round. In the review of Swedish trials by Nyström et al. (2002a), women born on day 31 were not included, and the totals analysed were 39 139 in the intervention group and 20 978 in the control group.

(g) Gothenburg trial

From December 1982 to April 1984, all women born in 1923–1944 and living in the city of Gothenburg, Sweden, were randomized to mammography screening or to a control group; of the 51 611 women, 25 941 were aged 39–49 years. Randomization was by cluster on the basis of date of birth for the cohorts born in 1929–1935 and by individual birth date for those born in 1936–1944 (Bjurstam et al., 1997). To enable rescreening of women every 18 months, with a limited capacity for mammography, the ratio of women randomized to the invited group and the control group was 1:1.2 in the age group 39–49 years and 1:1.6 in the age group 50–59 years. Attendance of invited women was 85% for the first round and 77% on average for subsequent rounds.

(h) United Kingdom Age trial

In 1991, a national, multicentre RCT was set up by the United Kingdom Coordinating Committee on Cancer Research (Moss, 1999). Women aged 39–41 years were randomized 1:2 to annual mammography screening for 7 years or to no screening, followed up without screening until they reached the age of 50 years, and then invited to participate in the United Kingdom National Health Service Breast Screening Programme of 3-yearly mammography. This is the only randomized screening trial that completely avoids "age creep" (the delayed benefits of screening for women randomized in their forties but diagnosed with breast cancer after their fiftieth birthday) (de Koning et al., 1995; Smith, 2000). The aim was to recruit 195 000 women, with 65 000 forming a study group and the remaining 130 000 a control group. However, recruitment was slower than anticipated, and a total of 160 921 women were randomized (Johns et al., 2010b). Attendance of women invited to routine screening was 68% for the first round and 69% for subsequent rounds. A total of 43 709 women in the intervention arm (81%) attended at least one routine screen, and 23 262 (43%) attended at least seven screens; 31 392 women attended 75% or more of all routine screens to which they were invited. To estimate the level of unscheduled screening in the control arm, Kingston et al. (2010) analysed data obtained from questionnaires sent to a random sample of 3706 women at five centres in the control arm of this trial, with a response rate of 58.8%. Overall, 24.9% of women surveyed reported having had a mammogram, but only about one third of the mammograms (8.4%) were for non-symptomatic reasons.

4.2.2 Beneficial effects

In this section, the data available from the randomized trials on breast cancer mortality, incidence of advanced breast cancer, and less-extensive therapy are summarized.

(a) Reduced breast cancer mortality

Of the 12 trials considered by the previous IARC Working Group on breast cancer screening (IARC, 2002), 11 had results on breast cancer mortality. The results from the United Kingdom Age trial were subsequently reported after 10 years of follow-up, and those for the CNBSS trials and the Two-County trial were subsequently updated.

For the Health Insurance Plan trial, the relative risk of death from breast cancer 18 years after recruitment was estimated by the previous IARC Working Group on breast cancer screening (IARC, 2002) from the data of Shapiro et al. (1988) to be 0.78 (95% confidence interval [CI], 0.61–1.00) overall.

In the Malmö I trial (women aged 45–70 years at randomization) with a follow-up of 19.2 years, the relative risk of death from breast cancer was 0.81 (95% CI, 0.66–1.00). In the Malmö II trial (women aged 43–49 years at randomization) after 9.1 years of follow-up, the corresponding relative risk was 0.65 (95% CI, 0.39–1.08) (Nyström et al., 2002a).

For the Two-County trial, after 29 years of follow-up, the relative risk of death from breast cancer among breast cancer cases diagnosed in the screening phase of both components of the trial (women aged 40–74 years at randomization) was 0.69 (95% CI, 0.56–0.84) according to data from the local end-point committee and 0.73 (95% CI, 0.59–0.89) according to consensus data from the overview committee appointed by the Swedish Cancer Society (Tabár et al., 2011).

For the Edinburgh trial, a report based on 14 years of follow-up and 577 518 person–years in the initial cohort (women aged 45–64 years at recruitment) showed a rate ratio for breast cancer mortality of 0.87 (95% CI, 0.70–1.06). After adjustment for socioeconomic status, the rate ratio was 0.79 (95% CI, 0.60–1.02) (Alexander et al., 1999).

For the CNBSS trials, after 20–24 years of follow-up, the breast cancer mortality hazard ratio based on the breast cancer cases ascertained in the 5-year screening period for both trials combined was 1.05 (95% CI, 0.85–1.30). The breast cancer mortality hazard ratio remained similar if the cancer accrual period was extended to 6 years (1.06; 95% CI, 0.87–1.29) or 7 years (1.07; 95% CI, 0.89–1.29) (Miller et al., 2014a).

In the Stockholm trial (women aged 40–64 years at assignment), the relative risk of death from breast cancer was 0.90 (95% CI, 0.63–1.28) after a median follow-up of 14.9 years. Although the possibility of double counting of controls in earlier analyses has been raised, in the most recent analysis reassurance was provided that there was no double counting (Nyström et al., 2002a).

In the Gothenburg trial (women aged 39–59 years at assignment), the overall relative risk of death from breast cancer was 0.79 (95% CI, 0.58–1.08) after a median follow-up of 14 years (Bjurstam et al., 2003).

In the United Kingdom Age trial (women aged 39–41 years at assignment), the ratio of breast cancer deaths in the study group relative to the control group was 0.83 (95% CI, 0.66–1.04) after a mean follow-up of 10.7 years (Moss et al., 2006).

(b) Age-specific effects of screening

The results from randomized trials that have published results related to mammography screening for women aged 40–49 years at entry are presented in Table 4.2. Relative risks of death from breast cancer ranged from 0.64 to 1.52, with a median of 0.76.

Limited data are available for the Health Insurance Plan trial, although Shapiro et al. (1988) noted that the benefit appeared to be restricted to women diagnosed with breast cancer after the age of 50 years, and took many years to appear.

Table 4.2 Age-specific results of randomized trials of the efficacy of mammography screening, with and without clinical breast examination – women aged 40–49 years

Trial, country References	Age (years) at enrolment/ screening	Mean duration of follow-up (years)	No. of women	Breast cancer mortality per 100 000 person–years (no. of breast cancer deaths) in screened/control group	RR	95% CI
Health Insurance Plan trial, USA Shapiro et al. (1988), IARC (2002)	40–49/40–54	18	NR	(49)/(65)	0.77	0.52–1.13
Malmö I and II trials, Sweden Andersson & Janzon (1997)	45–49/45–69	15.5 (Malmö I) 10 (Malmö II)	25 770	34 (57)/54 (78)	0.64	0.45–0.89
Malmö II trial, Sweden Nyström et al. (2002a)	43–49/43–57	9.1 (Malmö II)	17 793	26 (29)/38 (33)	0.65	0.39–1.08
Two-County trial: Östergötland County, Sweden Nyström et al. (2002a)	40–49/40–54	17.4	20 744	18 (31)/17 (30)	1.05	0.64–1.71
Two-County trial: Kopparberg County, Sweden Tabár et al. (2000)	40–49/40–54	20	NR	NR	0.76	0.42–1.40
Edinburgh trial, United Kingdom Alexander et al. (1999)	45–49/45–56	14	21 746	34 (47)/42 (53)	0.75	0.48–1.18
CNBSS 1 trial, Canada Miller et al. (2014a)	40–49/40–54	22	50 430	NR	1.09	0.80–1.49
Stockholm trial, Sweden Nyström et al. (2002a)	40–49/40–54	14.9	22 324	17 (34)/11 (13)	1.52	0.80–2.88
Gothenburg trial, Sweden Bjurstam et al. (2003)	39–49/39–55	14	25 941	(25)/(46)	0.65	0.40–1.05
United Kingdom Age trial Moss et al. (2006)	39–41/39–48	10.7	160 921	18 (105)/22 (251)	0.83	0.66–1.04

CI, confidence interval; CNBSS, Canadian National Breast Screening Study; NR, not reported; RR, relative risk.
From IARC (2002).

For the Malmö trials, Andersson & Janzon (1997) combined the data from the Malmö I and Malmö II trials, with a relative risk of death from breast cancer of 0.64 (95% CI, 0.45–0.89). This is the only relative risk presented in Table 4.2 where the upper 95% confidence limit is less than 1.0. In the Malmö I trial, the cumulative mortality curves did not begin to separate until after 5 years of follow-up, but in the Malmö II trial, separation began after the first year. For the Malmö II trial, Nyström et al. (2002a) presented age-adjusted data for women aged 43–49 years.

For the Two-County trial, updated data by age have not been reported for women aged 40–49 years or for women aged 50 years and older, but have been reported by separate segments of the age ranges in different publications. Table 4.2 presents the results from the Swedish overview analysis, where the findings only from Östergötland County were reported

Table 4.3 Age-specific results of randomized trials of the efficacy of mammography screening, with and without clinical breast examination – women aged 50 years and older

Trial, country References	Age (years) at enrolment/ screening	Mean duration of follow-up (years)	No. of women	Breast cancer mortality per 100 000 person–years (no. of breast cancer deaths) in screened/ control group	RR	95% CI
Health Insurance Plan trial, USA Shapiro et al. (1988), IARC (2002)	50–64/50–69	18	NR	(77)/(98)	0.79	0.58–1.08
Malmö I trial, Sweden Andersson et al. (1988)	55–69/55–79	8.8	26 210	(35)/(44)	0.79	0.51–1.24
Two-County trial: Östergötland County, Sweden Nyström et al. (2002a)	50–59/50–64 60–69/60–74	17.4	23 506	27 (53)/29 (54)	0.94 0.72	0.66–1.35 0.52–1.00
Two-County trial: Kopparberg County, Sweden Tabár et al. (2000)	50–59/50–64 60–69/60–74 70–74/70–78	20	22 435	39 (64)/54 (83)	0.46 0.58 0.76	0.30–0.71 0.39–0.87 0.44–1.33
Edinburgh trial, United Kingdom Alexander et al. (1999)	50–54/50–61 55–59/55–66 60–64/60–71	14	11 046 11 858 9 993	56 (44)/52 (35) 55 (43)/76 (55) 67 (42)/76 (44)	0.99 0.65 0.80	0.62–1.58 0.43–0.99 0.51–1.25
CNBSS 2 trial, Canada Miller et al. (2014a)	50–59/50–64	22	39 405		1.02	0.77–1.36
Stockholm trial, Sweden, Nyström et al. (2002a)	50–59/50–64 55–64/55–69	14.9	24 367 26 347	12 (25)/20 (24) 17 (39)/23 (28)	0.56 0.75	0.32–0.97 0.46–1.21
Gothenburg trial, Sweden Bjurstam et al. (2003)	50–59/50–61	14	25 670	(38)/(66)	0.91	0.61–1.36

CI, confidence interval; CNBSS, Canadian National Breast Screening Study; NR, not reported; RR, relative risk.

(Nyström et al., 2002a), with a relative risk of 1.05 (95% CI, 0.64–1.71). In the report by Tabár et al. (2000), the relative risk of death from breast cancer in Kopparberg County was 0.76 (95% CI, 0.42–1.40) for women aged 40–49 years.

Relative risks of less than 1.0 were reported from the Edinburgh trial and the Gothenburg trial for women aged 45–49 years and 39–49 years at entry, respectively; relative risks of more than 1.0 were reported from the CNBSS 1 trial and the Stockholm trial for women aged 40–49 years at entry.

Although analyses are based on women aged 40–49 years at entry into the trials, screening after age 49 years could have influenced the estimated relative risks of breast cancer mortality, so-called "age creep" (de Koning et al., 1995; Smith, 2000). Only the United Kingdom Age trial (Moss, 1999) was designed to overcome this. As stated above, in this trial of women aged 39–41 years at assignment, the ratio of breast cancer deaths in the study group relative to the control group was 0.83 (95% CI, 0.66–1.04) after a mean follow-up of 10.7 years (Moss et al., 2006).

Table 4.3 summarizes the available data on the efficacy of mammography screening for women aged 50 years and older at entry. For the Malmö I trial, data were available only for

women aged 55–69 years at entry. For many trials, data are available only for 10-year age groups. Partly because of this age separation, many of the relative risks presented show upper 95% confidence limits of more than 1.0. However, the upper 95% confidence limit was less than 1.0 for women aged 50–59 years and for those aged 60–69 years in Kopparberg County, for women aged 55–59 years in the Edinburgh trial, and for women aged 50–59 years in the Stockholm trial.

In a model-based analysis, Rijnsburger et al. (2004) evaluated whether the lack of benefit from mammography in the CNBSS 2 trial could have been due to a beneficial effect of the CBE performed in both arms for women aged 50–59 years. Using data derived from the CNBSS 2 trial, the Netherlands breast screening programme, and the Two-County trial, it was estimated that a mortality reduction of more than 20% could have been derived from the CBE if compared with a no-screening arm.

The only trial to enrol women aged 70–74 years was the Kopparberg component of the Two-County trial (Tabár et al., 1992). The participation rate of this group was poor, and only two screens were offered. At 15 years after randomization, the relative risk of death from breast cancer in the screened group compared with the control group was 0.79 (95% CI, 0.51–1.22) (Tabár et al., 1995). At 20 years after randomization, the relative risk of death from breast cancer in Kopparberg County was 0.76 (95% CI, 0.44–1.33) (Tabár et al., 2000).

(c) *Meta-analyses of results of randomized trials of mammography screening*

The previous IARC Working Group on breast cancer screening (IARC, 2002) reported the results of its own meta-analysis of the trials, including those using mammography alone compared with no screening as well as all valid trials in women aged 40–49 years. The results are summarized in Table 4.4, together with the results of subsequent meta-analyses. [None of these meta-analyses included the updated results of the Two-County trial or of the CNBSS trials.]

(d) *Reduced incidence of advanced breast cancer*

Most investigators consider that advanced breast cancer should be defined as extensive local involvement or metastatic disease, although the exact definition by stage will vary according to the level of detail recorded. In the randomized screening trials, this level of detail was rarely captured. The available data as reported by the authors of the various trials are presented in Table 4.5.

For the Health Insurance Plan trial, Shapiro (1977) reported that of 299 breast cancers in the study arm detected within 5 years of entry, 102 (34%) were node-positive (for 27, the nodal status was unknown) compared with 121 of 285 (42%) in the control arm (34 of unknown status).

For the Malmö I trial, Andersson et al. (1988) reported that, after an excess of stage II–IV breast cancers ascertained during the first screen, the numbers of breast cancers at these stages gradually became greater in the control group, resulting at 10 years in a cumulative rate per 100 000 person–years of 980 in the study group and 1210 in the control group [relative risk (RR), 0.81]. Most of the excess in the control group was from stage II cancers. There were 26 stage III and 22 stage IV breast cancers ascertained in the study group, and 27 stage III and 32 stage IV breast cancers in the control group (Andersson et al., 1988). No similar data have been reported for the Malmö II trial.

For the Two-County trial, Tabár et al. (1992) estimated the cumulative incidence of breast cancers of stage II or higher during the first 10 years of follow-up. There was an excess incidence in the ASP at year 1, which disappeared by year 3. Subsequently, the rate increased much more slowly in the ASP than in the PSP. At 10 years, the rate per 1000 person–years was just more than 10 in the PSP and less than 8 in the

Table 4.4 Meta-analyses of randomized controlled trials of the efficacy of mammography screening

Reference	Screen[a]	Age at entry (years)	No. of trials[b]	Population (thousands)		Breast cancer deaths		RR	95% CI
				Screened	Control	Screened	Control		
IARC (2002)[c]	M alone	40–49	6	58.6	49.1	166	173	0.81	0.65–1.01
	All	40–49	8					0.88	0.74–1.04
	M alone	50–69	6	188.5	147.8	496	549	0.75	0.67–0.85
Nelson et al. (2009)	All	39–49	8	152.3	195.9	448	625	0.85	0.75–0.96
		50–59	6					0.86	0.75–0.99
		60–69	2					0.68	0.54–0.87
		70–74	1					1.12	0.73–1.72
Canadian Task Force on Preventive Health Care (2011)	All	40–49	8	152.3	195.9	448	625	0.85	0.75–0.96
		50–69	7	135.1	115.2	639	743	0.79	0.68–0.90
		70–74	2	10.3	7.3	49	50	0.68	0.45–1.01
Magnus et al. (2011)[d]	All	39–49	7	144.6	191.6	427	615	0.83	0.72–0.97
Gøtzsche & Jørgensen (2013)	All	39–49	8	142.9	186.6	385	567	0.84	0.73–0.96
		≥ 50	7	146.3	122.6	599	701	0.77	0.69–0.86
Marmot et al. (2013)	All	40–74	9					0.80	0.73–0.89

[a] "All" indicates trials with mammography with or without CBE screening.
[b] The Two-County trial is regarded as two trials: Kopparberg County and Östergötland County.
[c] Excluded the Edinburgh trial.
[d] Included the Edinburgh trial but excluded Kopparberg County and Östergötland County.
CBE, clinical breast examination; CI, confidence interval; M, mammography; RR, relative risk.

ASP [rates approximated from Fig. 4 of Tabár et al. (1992)]. Tabár et al. (1995) reported that the cumulative incidence rate of lymph node-positive breast cancers together with those with distant metastases for women aged 40–49 years at 14 years of follow-up was 28.0 per 100 000 in the ASP and 32.8 per 100 000 in the PSP; the corresponding rates per 100 000 for women aged 50–74 years were 45.1 in the ASP and 64.4 in the PSP.

For the Edinburgh trial, Alexander et al. (1994) reported that of 489 breast cancers ascertained in the study group, 189 (39%) were of stage II (21 mm or larger), III, or IV (10 were of unknown stage), compared with 221 of 400 (55%) in the control group (7 of unknown stage).

For the CNBSS trials, no data have been reported on the incidence of advanced breast cancers, but data were reported on the nodal status of the majority of the breast cancers detected during the screening period, and for an average of 8.5 years of follow-up from enrolment (Miller et al., 1992a, b), and subsequently on tumour size (Miller et al., 2000, 2002). For the CNBSS 1 trial, the total of node-positive breast cancers in the mammography arm was 81 of 245 (33%) with known nodal status (for 33, the nodal status was unknown). The corresponding numbers were 59 of 203 (29%) for the control arm (45 of unknown nodal status) (Miller et al., 1992a). For the CNBSS 2 trial, the corresponding numbers were 83 of 281 (30%) in the mammography arm (47 of unknown nodal status) and 64 of 200 (32%) in the control arm (38 of unknown nodal status) (Miller et al., 1992b).

For the Stockholm trial, data were reported on breast cancers of stage II or higher. There was a cumulative incidence of 4.27 per 1000 in the intervention arm compared with 4.86 per 1000

Table 4.5 Incidence of advanced breast cancer in randomized trials of breast cancer screening

Trial, country[a]	Definition of advanced breast cancer	No. of patients with advanced breast cancer		Cumulative incidence of advanced breast cancer (‰)		RR	95% CI
		Intervention	Control	Intervention	Control		
Health Insurance Plan trial, USA	Stage II or higher	160	188	5.29	6.21	0.85	0.69–1.05
Malmö I trial, Sweden	Stage II or higher	190	231	9.01	10.90	0.83	0.68–1.00
Two-County trial, Sweden	Stage II or higher	524	555	6.80	9.91	0.69	0.61–0.78
CNBSS 1 trial, Canada	Size ≥ 20 mm	111	115	4.40	4.56	0.97	0.74–1.25
CNBSS 2 trial, Canada	Size ≥ 20 mm	114	136	5.78	6.91	0.84	0.65–1.07
Stockholm trial, Sweden	Stage II or higher	172	97	4.27	4.86	0.88	0.68–1.12
Gothenburg trial, Sweden	One or more nodes involved	85	144	3.93	4.81	0.80	0.61–1.05
United Kingdom Age trial	Size ≥ 20 mm	171	386	3.17	3.61	0.88	0.73–1.05

[a] Follow-up periods may differ between trials.
CI, confidence interval; CNBSS, Canadian National Breast Screening Study; RR, relative risk.
Adapted from Autier et al. (2009).

in the control arm, for a relative risk of 0.88 (95% CI, 0.68–1.12) (Table 4.5).

In the Gothenburg trial, the incidence of lymph node-positive breast cancers in the study group was 0.65 per 1000, compared with 0.81 per 1000 in the control group, for a relative risk of 0.80 (95% CI, 0.61–1.05). For women aged 50–59 years, the relative risk was 1.02 (95% CI, 0.70–1.48) (Bjurstam et al., 2003).

In the United Kingdom Age trial, which defined advanced breast cancers as those of 20 mm or larger, the cumulative incidence rate per 1000 was 3.17 in the intervention arm and 3.61 in the control arm, for a relative risk of 0.88 (95% CI, 0.73–1.05) (Moss et al., 2005a) (Table 4.5).

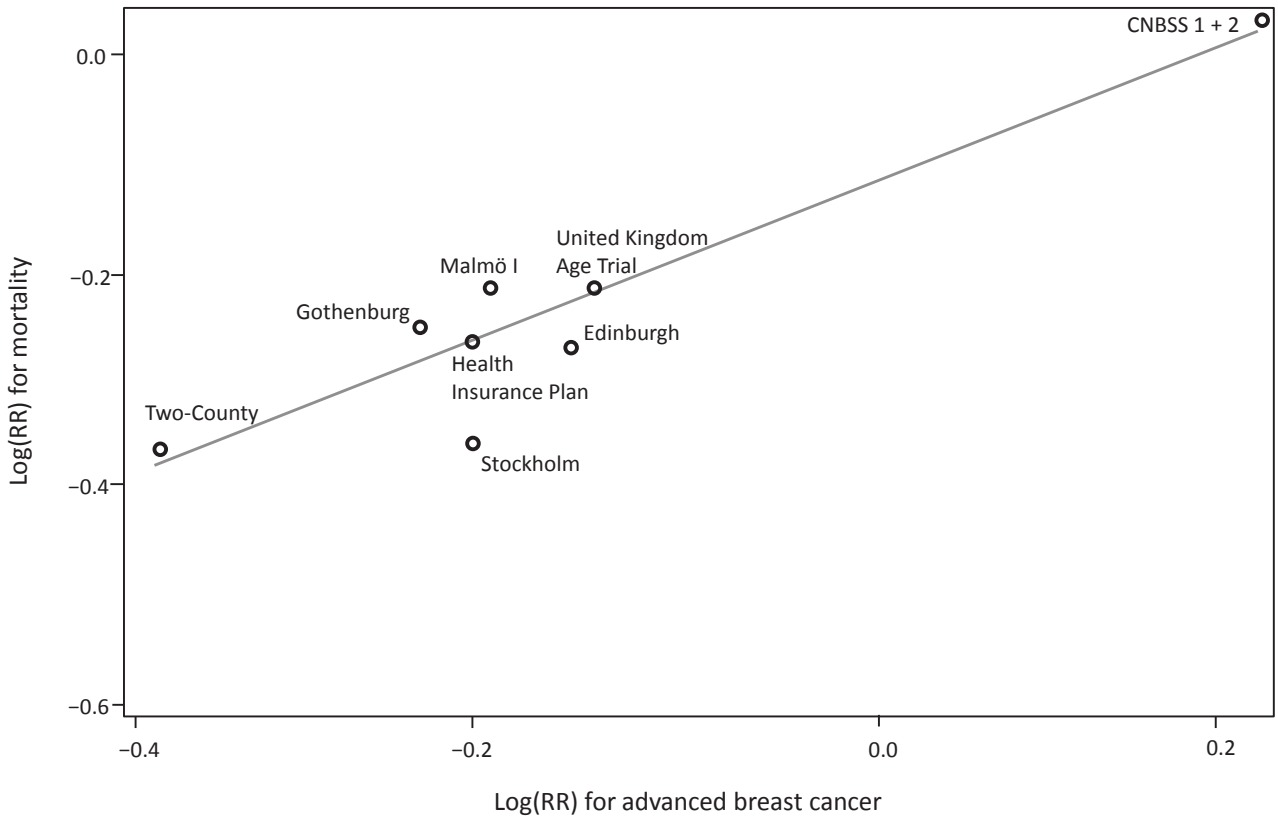
Based on the available data from randomized controlled trials, an association has been observed between the risk of advanced breast cancer and breast cancer mortality (Autier et al., 2009; Tabár et al., 2015a, b; Fig. 4.2).

(e) More-conservative surgery

The extent of use of breast-conserving surgery was reported for the Malmö I trial, although data were missing from some control subjects with stage 0 disease (Andersson et al., 1988). Overall, of 575 women with breast cancer ascertained in the study group, 137 (24%) received breast-conserving surgery, compared with 80 (18%) of 436 in the control group.

Gøtzsche & Jørgensen (2013), in a Cochrane review, reported that the risk ratio for mastectomies in the screened versus unscreened groups based on 5 trials was 1.20 (95% CI, 1.11–1.30) and for lumpectomies and mastectomies combined was 1.35 (95% CI, 1.26–1.44), thus suggesting that screening in the trials did not result in more-conservative surgery. [The Working Group noted that the sources of the data from which these estimates were made are unclear.]

Fig. 4.2 Plot of data from randomized controlled trials, showing the association between the logarithm of relative risk (RR) of advanced breast cancer and of disease-specific mortality, with meta-regression line



From Tabár et al. (2015a).
CNBSS, Canadian National Breast Screening Study.

4.2.3 Performance indicators

For consistency with Section 5.1 on indicators for monitoring effectiveness of screening, the data available on false-positive mammograms and interval cancers are summarized here, as process indicators of performance in these trials.

(a) False-positive mammograms

For the Malmö trials, Andersson & Janzon (1997) reported that in women younger than 50 years, further examination of false-positives was required in 1260 per 100 000 person–years; the rate of surgery for benign disease was 56 per 100 000 person–years, and the rate of treatment of clinically insignificant cancer was 10 per 100 000 person–years. No data on false-positives were reported for older women.

For the Two-County trial, the rate of recall for assessment for those not found to have breast cancer was 44 per 1000 at the first screen and 22 per 1000 at subsequent screens; the rate of biopsy for benign conditions was 6 per 1000 at the first screen and 1 per 1000 at subsequent screens (Tabár et al., 1992).

In the CNBSS trials, with screening by both mammography and CBE, it is not possible to fully distinguish the contribution of mammography to false-positive detections. As a result of the referrals by the study surgeon, the overall

rates of surgical intervention after the first screen were 64 per 1000 in the mammography group and 37 per 1000 in the control group in the CNBSS 1 trial (Miller et al., 1992a) and 58 per 1000 in the mammography group and 25 per 1000 in the control group in the CNBSS 2 trial (Miller et al., 1992b). After subsequent screens, the rates were approximately one half of those after the first screen. These procedures resulted after the first screen in overall rates of biopsy detection of benign lesions of 33.6 per 1000 in the mammography group and 11.5 per 1000 in the control group in the CNBSS 1 trial (Miller et al., 1992a) and 34.8 per 1000 in the mammography group and 8.7 per 1000 in the control group in the CNBSS 2 trial (Miller et al., 1992b). After subsequent screens, the rates of biopsy with detection of benign breast lesions were approximately one third of those after the first screen (Miller et al., 1992a, b).

For the Stockholm trial, Frisell & Lidbrink (1997) reported that the recall rate was 0.8% for all subjects and 1.0% for those in the age group 40–49 years. With only two screening rounds, the rate of false-positives was 242 per 100 000 person–years in women older than 50 years compared with 355 per 100 000 in those younger than 50 years. The rate of benign surgical biopsies in the second round was 21 per 100 000 in women older than 50 years and 49 per 100 000 in those younger than 50 years. In women aged 40–49 years, 1 out of 2.5 surgical biopsies was benign, compared with 1 out of 7 in those older than 50 years.

For the Gothenburg trial, Bjurstam et al. (2003) reported that 5.9% of the participants in the study group were recalled for supplemental mammography at the first screen, and 2.6% at subsequent screens. The percentages of women who had clinical examination and fine-needle aspiration cytology who were not found to have cancer were 1.5% at the first screen and 0.7% at subsequent screens; the corresponding percentages for surgery were 0.3% and 0.1%, respectively.

For the United Kingdom Age trial, Johns et al. (2010a) reported that 14.6% of women in the intervention arm and 18.1% of women attending at least one routine screen experienced one or more false-positive screens during the trial.

(b) Interval cancers

In the Malmö I trial, 100 (17%) breast cancers were detected in the 2-year interval before the next screen was due, out of 581 breast cancers ascertained in the study group (Andersson et al., 1988). Corresponding data have not been reported from the Malmö II trial.

For the Two-County trial, Tabár et al. (1992) reported the incidence of interval cancers as a percentage of the incidence in the control group by age. Over all intervals between screens, the percentage for women aged 40–49 years was 45% in the first year and 62% in the second; the percentages over the 3-year intervals were 17%, 34%, and 63%, respectively, for women aged 50–59 years, 17%, 27%, and 46%, respectively, for those aged 60–69 years, and 8%, 44%, and 48%, respectively, for those aged 70–74 years.

In the CNBSS 1 trial, the rate of interval cancers after the first screen was 0.75 per 1000 in the mammography group and 1.11 per 1000 in the control group. For the second, third, fourth, and fifth screens in the mammography group, the rates were 0.71, 0.36, 0.46, and 0.64 per 1000, respectively (Miller et al., 1992a). In the CNBSS 2 trial, data on interval cancer rates were available for both the mammography group and the control group after all five screens. The rates per 1000 in the mammography group and the control group, respectively, were 0.76 and 0.81 after the first screen, 0.57 and 0.92 after the second screen, 0.46 and 1.52 after the third screen, 0.52 and 0.95 after the fourth screen, and 0.51 and 1.64 after the fifth screen (Miller et al., 1992b).

For the Stockholm trial, Frisell et al. (1986) reported that 60 interval cancers (6 in situ) occurred in the 24 months between the two screens (1.8 per 1000 examinations), and 38 of

the cases occurred in the second year. A review of the original mammograms found no indication of an abnormality in 31 cases (2 in situ); 45% of them were in women aged 40–49 years, and only 8 occurred in the first year.

For the Gothenburg trial, Bjurstam et al. (2003) reported that 52 [24%] invasive interval cancers occurred of the total of 220 invasive cancers ascertained in attenders. There were an additional 2 in situ interval cancers of 36 ascertained in attenders. The proportion of invasive interval cancers decreased with increasing age, from 36% at ages 39–44 years to 31% at 45–49 years, 16% at 50–54 years, and 15% at 50–59 years [percentages calculated by the Working Group]. The two in situ interval cancers were ascertained in women younger than 50 years.

In the United Kingdom Age trial, there were 125 (26%) interval cancers and 229 (48%) screen-detected cancers of the total of 482 breast cancers ascertained (Moss et al., 2005b). However, of the total, 9 breast cancers were diagnosed between randomization and invitation, and 61 breast cancers occurred in never-attenders, 44 in lapsed attenders, and 14 in women lost to screening. If these are excluded from the denominator, the percentages become 35% and 65%, respectively.

(c) Overdiagnosis of breast cancer

(i) Definition

Overdiagnosis of breast cancer is detection by screening of a breast cancer (DCIS or invasive carcinoma) that would never have presented clinically during the woman's lifetime if it had not been detected by screening. Overdiagnosis is invariably associated with the use of any method that is able to effectively bring forward the date of diagnosis. The probability that a tumour represents an overdiagnosis versus a timely diagnosis is determined by two components: the speed of growth, which determines the time the tumour would have required to present clinically, and the remaining lifespan of a patient, which depends on the patient's age at diagnosis and other competing causes of death. Overdiagnosis is an important harm caused by screening because of the otherwise unnecessary investigation, treatment, and psychosocial consequences that a diagnosis of cancer entails. Overdiagnosed cases cannot be identified individually, but, based on the above-mentioned components, the majority of overdiagnoses represent slower-growing, lower-grade cancers, both in situ and invasive.

(ii) Counting overdiagnosed cancers

Conceptually, overdiagnosed cancers can be counted as the difference between the numbers of breast cancer cases, including in situ and invasive, accumulated in screened and unscreened cohorts from the beginning of screening in the screened cohort until the end of the compensatory drop in incidence that occurs after screening has ended (i.e. when the lead time of all breast cancer cases diagnosed as a result of screening has elapsed) (Puliti et al., 2011). In principle, randomized screening trials, in which there is a clearly defined end to trial screening and a period of follow-up for new incident cases in both screened and unscreened women beyond the end of the compensatory drop, provide the best estimates of overdiagnosis under the assumption that there is no further screening outside of the trial, or at least that the accrual of diagnosed breast cancers outside of the trial is approximately the same in the two arms (Moss, 2005; Biesheuvel et al., 2007; Independent UK Panel on Breast Cancer Screening, 2012; Marmot et al., 2013). However, this requirement is rarely, if ever, met, or known with any certainty to have been met, by any trial. The time interval that should be allowed for the compensatory drop is uncertain. Information on the timing of the compensatory drop is available from established screening programmes (de Gelder et al., 2011). For a randomized screening trial in which two cohorts of women are

recruited, screened or not screened for a period, and followed up for a period, Duffy & Parmar (2013) depicted in Fig. 1 of their article that excess cancers due to lead time accumulate for 10 years, this excess remains constant for as long as screening lasts, and then the excess dissipates over 10 years. Therefore, given the assumptions of Duffy & Parmar (2013) as to median lead time and its distribution, 10 years after screening has ended seems a suitable point at which to attribute any remaining excess to overdiagnosis.

(iii) Estimating the proportion of incident cancers that are overdiagnosed

The Independent United Kingdom Panel on Breast Cancer Screening (Marmot et al., 2013), following earlier work by de Gelder et al. (2011), defined four measures of the overdiagnosis rate based on data from randomized screening trials. In each, the numerator was a count of overdiagnosed cancers. The four denominators were: (A) breast cancers diagnosed over the whole follow-up period in unscreened women (where the follow-up period extends from the beginning of screening in the screened women until the end of follow-up in both screened and unscreened women); (B) breast cancers diagnosed over the whole follow-up period in women invited to screening; (C) breast cancers diagnosed during the screening period in women invited to screening; and (D) breast cancers detected by screening in women invited to screening. The United Kingdom Panel preferred denominators (B), as representing the population perspective, and (C), as representing the perspective of a woman invited to screening.

(iv) Estimates of overdiagnosis rates from the trials

For the Health Insurance Plan trial, cumulative in situ and invasive breast cancer incidence rates at 10 years after the beginning of the trial (~6 years after the end of the trial) were reported as 2.11 per 1000 in women offered screening and 2.09 per 1000 in control women (Table 1 in Shapiro, 1997), from which an overdiagnosis rate of 1% can be estimated, as a proportion of breast cancers diagnosed in unscreened women over the whole follow-up period. The excess number of incident invasive breast cancers at 10 years was 0 (Table 5.1 in Shapiro et al., 1988). However, the year-by-year data on invasive breast cancer do not show a decrease in incident breast cancers in screened women from years 1–4 (screening) to years 5–10 (after screening); the average annual numbers were 62 and 61, respectively. Instead, there was an increase in incident cases in the control group; the corresponding annual average numbers were 55 and 66, respectively. [Therefore, there may have been a period of "catch-up" screening in the control group after trial screening ended, which would bias the estimate of overdiagnosis from the Health Insurance Plan trial downwards.]

In updating results from the Malmö I trial, Zackrisson et al. (2006) reported incidence data separately for women aged 45–54 years and those aged 55–69 years at entry. However, conclusions on overdiagnosis could be drawn only for women aged 55–69 years, whose controls were never screened, in contrast to women aged 45–54 years, whose controls were offered screening after the end of the screening period. In women aged 55–69 years at entry, the relative risk of in situ and invasive breast cancer was 1.10 (95% CI, 0.99–1.22) and the relative risk of invasive breast cancer was only 1.07 (95% CI, 0.96–1.18). Thus, 15 years after the trial ended the rate of overdiagnosis of breast cancer was 10% in women randomized to screening at age 55–69 years compared with an unscreened control group. Njor et al. (2013) questioned the validity of this estimate on several grounds. They argued that older screened women would not have been followed up long enough for the whole of the compensatory drop to have occurred, with resulting upward bias in the overdiagnosis estimate. In addition, since mammography screening was available outside of the

screening trial for the whole period, women in the screening arm would have continued to participate in screening after the end of the trial, which would also have biased the overdiagnosis estimate upwards. They presented data [percentages calculated by the Working Group from data in Table 1 in Njor et al. (2013)] showing that 20% of cancers diagnosed in all screened women (34% of cancers in the youngest women) in the 10 years after trial screening ended were asymptomatic, i.e. probably screen-detected. [The Working Group considered both the overdiagnosis estimates of Zackrisson et al. (2006) and the updated estimates of Njor et al. (2013) difficult to interpret.]

At the end of the Two-County trial, in 1985, cumulative in situ and invasive breast cancer incidence rates were 18.50 per 1000 in women offered screening and 18.61 per 1000 in control women, and the excess breast cancer incidence in screened women relative to that in control women was −0.06% (Duffy et al., 2003b). [The numbers of breast cancers contributing to these rates are stated elsewhere to have been those at the end of 1992 (Tabár et al., 1995).] In 2012, cumulative breast cancer incidence numbers every 5 years from the start of the trial until 29 years later were published for the Dalarna (formerly Kopparberg) County component of the trial (Yen et al., 2012). Screening of the control group began after an average of three screens of women in the screened group, 6–8 years after the start. The relative cumulative risk of breast cancer in the screened group was 1.34 (95% CI, 1.13–1.59) at 5 years after the start, 1.03 (95% CI, 0.91–1.16) at 10 years, 1.04 (95% CI, 0.94–1.15) at 15 years, 1.06 (95% CI, 0.97–1.16) at 20 years, 1.02 (95% CI, 0.94–1.11) at 25 years, and 1.00 (95% CI, 0.92–1.08) at 29 years. The authors concluded that "there was no overdiagnosis associated with the additional 3 screens of the [screened group] in the first 8 years of observation." [Because screening of the control group began after the end of scheduled screening in the screened group and continued in that group also, it is not possible to make an estimate of the extent of overdiagnosis caused by the screens in this trial.]

Incidence data from the Edinburgh trial have been reported to 10 years, 3 years beyond the end of the intervention period (Alexander et al., 1994, 1999). Organized service screening began in Scotland in 1988; women in the screening arm of the trial received their first invitation to service screening about 3 years after their last trial screen (year 7). Although it is not stated, it is assumed that women in the control arm could have begun service screening in 1988 if they were then aged 50–64 years, the target age group for service screening. Cumulative in situ and invasive breast cancer incidence rates to 10 years were 22.4 per 10 000 in women randomized to screening and 20.0 per 10 000 in control women (Alexander et al., 1994), from which an overdiagnosis rate of 12% can be estimated, as a proportion of cancers diagnosed in unscreened women over the whole follow-up period. There were 57% fewer incident breast cancers in screened women than in control women during the 3 years of post-screening follow-up, consistent with a substantial compensatory drop (Alexander et al., 1994). [It is doubtful whether 3 years after the end of screening in the trial would have been sufficient for the compensatory drop to have been completed.]

For the CNBSS trials, initiated in 1980, the period of screening was the first 5 years after randomization, and the follow-up period was 20–25 years after randomization (Miller et al., 2014a). Screening was provided in the intervention groups for four or five annual screening rounds. The subsequent history of screening in the intervention and control groups after the end of trial screening was not reported. In the first 5 years, the cumulative incidence of invasive breast cancer in the group offered mammography relative to that in the control group was 1.27 (95% CI, 1.13–1.42), with an excess of cancers in the screened group of 142. After 10 years of follow-up,

it was 1.09 (95% CI, 1.01–1.18), and after 25 years it was 1.04 (95% CI, 0.99–1.08) [relative risks and confidence intervals estimated from data in Table 1 in Miller et al. (2014a)]. The excess of breast cancer in the group offered mammography became constant at 106 cancers 15 years after enrolment (i.e. 10 years after the end of screening). This excess was 22% of all screen-detected invasive cancers in the trial (484). Miller (2014) reported that if in situ cancers are included in these estimates, the proportion of screen-detected cancers that were overdiagnosed increases to 35%. [There is a potential contribution of CBE to overdiagnosis in the CNBSS 2 trial, which has not been assessed. Women in both arms of the trial could have joined service mammography screening between 1988 and 1998, when organized screening services were rolled out across Canada, and opportunistic screening could also have occurred. Therefore, the excess cancers in the intervention arm may not be attributable exclusively to the screen-detection in the trials. Correspondingly, accrual of cases in the control arm may also have been inflated by screening. The resulting potential for bias makes the overdiagnosis estimate from the CNBSS trials difficult to interpret.]

The Stockholm trial offered two rounds of mammography screening at an interval of about 2.5 years to 40 318 women, beginning in 1981 and ending in 1985. In 1986, one round of screening was offered to the 19 343 control women, and recording of incident breast cancers in both groups ceased at the end of 1986. At the end of 1985, 371 cancers, both in situ and invasive, had been diagnosed in women randomized to screening, and 257 in control women (adjusted to the size of the population randomized to screening; Frisell et al., 1991), a 44% excess of breast cancer in screened women relative to that in control women. At the end of 1986, 428 cancers had accumulated in women offered screening, and 217 in control women (Frisell et al., 1997) (439 when adjusted as described above; Frisell et al., 1991). [Lack of follow-up for incident breast cancers after the end of the trial period prevents any estimate of overdiagnosis from the Stockholm trial.]

At the end of the Gothenburg trial, both groups were invited to service screening. Incidence of breast cancer (DCIS and invasive) was ascertained until the end of 1996, about 8 years after the end of the trial, and also at the end of the screening phase, which included the first service screening round for control women aged 50–69 years. There was a clear excess of breast cancers 4 years after the start of the trial in women randomized to screening (Fig. 2 in Bjurstam et al., 2003), but there was no excess at the end of the screening phase (excess over control group, −6.0%) or at the end of follow-up, 8 years after the end of the trial (−6.6%, invasive cancer only) [estimates based on data in Table 1 and text in Bjurstam et al. (2003)]. [No explanation has been offered by the authors for this paradoxically lower incidence of breast cancer in the control group than in the screened group.]

(d) Frequency of mammography screening

Only one trial provided informative data about the effects of varying screening frequency. The effect of annual versus 3-yearly mammography screening in increasing the likelihood of an improved outcome was tested in one trial (Breast Screening Frequency Trial Group, 2002). The measured outcomes included tumour size, nodal status, and histological grade of invasive tumours. These data were incorporated into two models to predict breast cancer mortality. Although the tumours diagnosed in women in the study arm were significantly smaller than those in women in the control arm, there was no difference in terms of nodal status or histological grade. The relative risks of predicted deaths from breast cancer for annual versus 3-yearly screening were 0.95 (95% CI, 0.83–1.07) and 0.89 (95% CI, 0.77–1.03) in the two models.

In most of the randomized screening trials, a 1–2-year screening interval was used. In the Two-County trial, a 24-month interval was used for women aged 40–49 years and a 33-month interval for those aged 50–74 years. [Given the different designs of these trials, it is not possible to derive estimates of the comparative efficacy of screening by different intervals by comparing their results.]

(e) Digital mammography

No trials of digital mammography with breast cancer mortality as the end-point have so far been reported. Trials that had breast cancer detection as the end-point are discussed in Section 2.1.3.

4.3 Clinical breast examination

4.3.1 Randomized clinical trials

Comparisons of the efficacy of CBE versus no screening come from three randomized studies (Pisani et al., 2006; Mittra et al., 2010; Sankaranarayanan et al., 2011). One of them closed after the first round of intervention, due to poor compliance (Pisani et al., 2006), and the other two have not yet reported their results on breast cancer mortality (Mittra et al., 2010; Sankaranarayanan et al., 2011).

(a) CBE versus no screening

See Table 4.6.

(i) Mumbai study

The Mumbai study (Mittra et al., 2010) is a cluster RCT that was initiated in 1998 by investigators from the Tata Memorial Hospital, Mumbai, India. Approximately 150 000 women underwent CBE at 24-month intervals, followed by 8 years of active monitoring for breast cancer incidence and mortality in the screening arm and one round of health education at entry, followed by active monitoring for self-reported cases and deaths from breast cancer in the control arm. The screening positivity rates for CBE were 0.46%, 0.77%, and 0.94% for the first, second, and third rounds of screening, respectively. Compliance rates for diagnostic confirmation ranged from 68% for the first round to 78% for the third round. Cancers were confirmed by histology in about 0.04% of women who underwent CBE. The mean age at detection was 49.8 years for both the screen-detected breast cancer cases and women in the control group.

During the corresponding period, in the control arm, there were 18 symptomatic referrals with 3 histologically confirmed cases at the first round, 61 symptomatic referrals with 39 histologically confirmed cases at the second round, and 76 symptomatic referrals with 45 histologically confirmed cases at the third round. Cohen's kappa for the agreement rates for CBE between the expert and the primary health workers was 0.849. In the screening arm, during the first, second, and third screening rounds, respectively, 21, 15, and 12 breast cancers were detected at early stages (stages 0, I, and II), 9, 7, and 9 cases were detected at advanced stages (stages III and IV), and for 2, 2, and 4 cases, staging information was unavailable. In the screening arm overall, [62.4% (78/125)] cancers were diagnosed at early stages and [25.6% (32/125)] at advanced stages, whereas in the control arm, [43.7%] were diagnosed at early stages and [42.5%] at advanced stages. The shift to a lower stage in the screening arm compared with the control arm was statistically significant (P = 0.0082; RR, 1.45; 95% CI, 1.09–1.93) (Table 4.6). The results on breast cancer mortality are awaited.

(ii) Trivandrum study

The Trivandrum cluster randomized study (Sankaranarayanan et al., 2011) began in 2006 in the Trivandrum District of Kerala, India, to evaluate whether three rounds of 3-yearly CBE would reduce advanced disease incidence rates and breast cancer mortality rates. A total of 115 652 healthy women aged 30–69 years in 275 electoral

Table 4.6 Randomized controlled studies of clinical breast examination: performance characteristics and tumour detection

Study Reference	Age range	Performance		Cancers in screening arm[a]		No./% of tumours, by stage[b]	
		Sensitivity	Specificity	Screen-detected cancers	Interval cancers	Screened group	Control group
Mumbai study Mittra et al. (2010)	35–64	57.4%	91.9%	73 (81)	37 (44)	Early lesion, 78 Advanced lesion, 32	Early lesion, 38 Advanced lesion, 37
Trivandrum study Sankaranarayanan et al. (2011)	30–69	51.7%	94.3%	80	28	Early lesion, 43.8% Advanced lesion, 45.0%	Early lesion, 25.4% Advanced lesion, 68.3%
Philippines study Pisani et al. (2006)	35–64	53.2%	100%	68	NA	[17% more advanced lesions in control group]	

[a] Number of tumours with available staging (total number of tumours).
[b] Early lesion included tumour size < 5 cm (T1 and T2), and advanced lesion included T3 and T4.

wards (clusters) were randomly allocated to the intervention group (CBE) or the control group (no screening). An intention-to-treat analysis was performed for comparison of incidence rates between the two groups. Preliminary results for incidence are based on follow-up until 2009, when the first round of screening was completed. Among the 2880 CBE-positive women, 1767 were judged to have a palpable lump and the remaining 1113 to have other abnormalities. The sensitivity was 51.7%, and the specificity was 94.3%. Among the intervention and control groups, 80 and 63 women, respectively, were diagnosed with breast cancer. The percentage of early-stage (stage IIA or lower) breast cancer was 43.8% (95% CI, 32.9–54.6%) in the intervention group versus 25.4% (95% CI, 14.6–36.1%) in the control group ($P = 0.023$), and the percentage of advanced-stage (stage IIB or higher) breast cancer was 45.0% (95% CI, 34.1–55.9%) in the intervention group versus 68.3% (95% CI, 56.8–79.7%) in the control group ($P = 0.005$). This indicates a shift to a lower stage of cancers in the CBE arm.

(iii) Philippines study

The randomized trial in the Philippines (Pisani et al., 2006) began in 1995. Women aged 35–64 years from urban Manila were randomized to five annual CBEs (carried out by trained nurses or midwives) or no screening. The first round of CBE took place in 1996–1997 (over 24 months) and included 151 168 women, who were also instructed in the technique of BSE; 8% of these women refused CBE. Of those examined, 2.5% had palpable lesions and were referred for investigation; of these, 1293 (37.2%) received further investigation. Complete diagnostic follow-up was achieved for only 1220 women (35% of those who were positive on screening); 42.4% refused further investigation, even with a home visit, and 22.6% were lost to follow-up. The sensitivity of annual CBE was 53.2%, and the positive predictive value (PPV) was 1.2%. In the control arm, 17% of the cases presented with advanced disease. Because of the poor compliance with follow-up of screen-positive women, even with home visits, the active intervention

was discontinued after the first screening round was completed, in December 1997.

All three studies evaluating CBE versus no screening showed a shift to a lower stage of the tumours detected.

(b) Mammography plus CBE versus no screening

Table 4.7 and Table 4.8 present the study characteristics and the outcome, respectively, of RCTs and other studies evaluating the efficacy of mammography plus CBE compared with no screening or compared with CBE alone.

(i) Health Insurance Plan trial

The Health Insurance Plan trial was the first RCT of breast cancer screening and was designed to assess the role of screening in reducing mortality from breast cancer, using mammography and CBE performed by trained surgeons. Approximately 61 000 women aged 40–64 years were included in the study (Shapiro et al., 1971). The results after 18 years from entry reported a relative risk for death from breast cancer of 0.77 (95% CI, 0.61–0.97). The proportion of cases detected with mammography was low, especially in younger women; also, the benefit appeared to be more due to the earlier detection of advanced rather than early disease (Shapiro, 1994; Miller, 2004). [The individual contribution of each intervention remained ambiguous.] The contribution of CBE in the detection of breast cancer was 67% (Table 4.8).

(ii) Edinburgh trial

The Edinburgh randomized trial of breast cancer screening (Alexander et al., 1994; Alexander, 1997) recruited 44 288 women aged 45–64 years into the initial cohort of the trial during 1978–1981. A total of 22 944 women were randomized into the study group and were offered screening for 7 years; the remaining women constituted the control group. After 10 years, breast cancer mortality was 21% lower in the study group than in the control group (not statistically significant) in women older than 50 years. The relative risk of death from breast cancer in all women was 0.82 (95% CI, 0.61–1.11). The contribution of CBE in the detection of breast cancer was 74% (Table 4.8).

(c) Mammography plus CBE versus CBE alone

The CNBSS 2 trial (Miller et al., 1992a, b; Barton et al., 1999) compared annual CBE plus mammography versus CBE in a randomized setting (Table 4.7 and Table 4.8). Mammography plus CBE detected more node-negative and small breast cancers compared with screening with CBE alone, but there was no impact on breast cancer mortality. Mammography showed no added value to CBE, with a relative risk of 0.97 (95% CI, 0.62–1.52). [The Working Group noted that this study does not allow an evaluation of the efficacy of CBE in reducing breast cancer mortality.]

4.3.2 Nested case–control study

The DOM project, a population-based, non-randomized breast cancer screening programme with physical examination and xeromammography, was started in 1974 in the city of Utrecht, The Netherlands (Table 4.7). A total of 116 cases of breast cancer were detected with screening, of which 55.6% were detected with mammography alone, 9.7% with CBE alone, and 34.6% with combined-modality screening (De Waard et al., 1984). A protective effect of screening against breast cancer mortality was found in a nested case–control study after 8 years of follow-up (odds ratio [OR], 0.30; 95% CI, 0.13–0.70) (Collette et al., 1984), which decreased after 14 years of follow-up (Collette et al., 1992). Analysis within different age subgroups showed the effect to be more pronounced for older women (OR, 0.38; 95% CI, 0.18–0.83) than for younger women (OR, 0.91; 95% CI, 0.39–2.13) (Collette et al., 1992).

Table 4.7 Characteristics of studies evaluating combined mammography and clinical breast examination

Study, country References	Design	Years of recruitment	CBE examiners	Age at entry (years)	No. of women		Screening modality (intervention vs control)
					Intervention	Control	
Randomized controlled trials							
Health Insurance Plan trial, USA Shapiro et al. (1988)	Randomized	1963–1966	Surgeons	40–64	30 131	30 565	CBE annually + mammography annually vs none
Edinburgh trial, United Kingdom Alexander et al. (1994)	Cluster randomized	1979–1988	Physicians, nurses	45–64	22 944	21 344	CBE annually + mammography every 2 years vs none
CNBSS 1 trial, Canada Miller et al. (1992a)	Randomized	1980–1988	Nurses	40–49	25 214	25 216	CBE annually + mammography annually vs CBE at entry
CNBSS 2 trial, Canada Miller et al. (1992b)	Randomized	1980–1985	Nurses	50–59	19 711	19 694	CBE annually + mammography annually vs CBE annually
Nested case–control study							
DOM study, Netherlands Collette (1985), Collette et al. (1992)	Nested case–control	1974–1981	Medical assistants	50–64	14 796 invited: 54 cases, 162 controls	–	CBE annually; mammography annually
Observational studies							
Breast Cancer Detection Demonstration Project Baker (1982), Morrison et al. (1988)	Prospective	1973–1981	Nurses	35–74[a]	283 222[a]	–	CBE + mammography + thermography[b] annually
West London study, United Kingdom Chamberlain et al. (1979)	Prospective	1973–1977	Nurses, then doctors	> 40	2484	–	CBE + mammography at 0, 6, 12, and 24 months
United Kingdom Trial of Early Detection of Breast Cancer Moss et al. (1993), UK Trial of Early Detection of Breast Cancer Group (1993)	Prospective, non-randomized	1979–1988	Physicians, nurses	45–64	45 956	127 109	CBE annually + mammography every 2 years vs none
Data analysis from the National Breast and Cervical Cancer Early Detection Programme Bobo et al. (2000)	Prospective	1995–1998	Doctors	[c]	564 708	–	CBE annually; mammography annually

Table 4.7 (continued)

Study, country References	Design	Years of recruitment	CBE examiners	Age at entry (years)	No. of women		Screening modality (intervention vs control)
					Intervention	Control	
Data analysis of four Canadian breast cancer screening programmes Bancej et al. (2003)	Prospective	1996–1998	Nurses, technologists	50–69	300 303	–	CBE and mammography in alternate years
Breast Cancer Screening Programme at Group Health Cooperative of Puget Sound Oestreicher et al. (2005)	Prospective	1996–2000	Nurses	≥ 40	61 688	–	CBE and mammography every 1–2 years based on breast cancer risk factors
Well Women Clinics, opportunistic breast screening in Hong Kong Special Administrative Region, China Lui et al. (2007)	Prospective	1998–2002	Doctors	≥ 40	29 028	–	CBE + mammography every 2 years (188 women aged 35–39 years also screened based on family history)
Breast care centre, Hong Kong Sanatorium and Hospital, Hong Kong Special Administrative Region, China Kwong et al. (2008)	Prospective	1999–2006	Family physicians	[c]	11 408	–	BSE training; CBE; mammography
Breast screening comparative study in Chengdu, China Huang et al. (2012)	Prospective	2009–2011	Breast surgeon	25–80	3 028	–	CBE, mammography, and ultrasonography annually (2 rounds)

[a] 99.4% of screenees were aged 35–74 years at entry, although any woman seeking screening could participate. At least 283 222 women had been screened as of September 1981.

[b] CBE, mammography, and thermography were used from the start of the project until 1977, when thermography was dropped and mammography was restricted to women aged 50 years and older and women at high risk who were younger than 50 years.

[c] The age range of women who were offered breast screening is not specified. However, some data are presented for women aged ≤ 40 years and for those aged ≥ 65 years.

BSE, breast self-examination; CBE, clinical breast examination; CNBSS, Canadian National Breast Screening Study.

Table 4.8 Outcome of studies of combined mammography and clinical breast examination

Study, country References	No. of rounds	Duration of follow-up (years)	Mortality reduction, RR (95% CI)	No. of cancers detected	
				Total	CBE only No. (%)
Randomized controlled trials					
Health Insurance Plan trial, USA Shapiro et al. (1988), Barton et al. (1999)	4	18	0.77 (0.61–0.97)	132	59 (45%)
Edinburgh trial, United Kingdom Alexander et al. (1994), Barton et al. (1999)	7	10	0.82 (0.61–1.11)	88	3 (3%)[a]
CNBSS 1 trial, Canada[b] Miller et al. (1992a), Barton et al. (1999)	5	7	0.86 (0.73–1.01)	255	61 (24%)
CNBSS 2 trial, Canada[c] Miller et al. (1992b), Barton et al. (1999)	5	7	0.29 (0.14–0.62)	325	39 (12%)
Nested case–control study					
DOM study, Netherlands Collette (1985), Collette et al. (1992)	4	14	0.52 (0.32–0.83)[d]	116[e]	(9.7%)[e]
Observational study					
United Kingdom Trial of Early Detection of Breast Cancer UK Trial of Early Detection of Breast Cancer Group (1993), Barton et al. (1999)	7	10	0.73 (0.63–0.84)	432	24 (6%)

[a] Results based only on data from first round screening.
[b] Mammography + CBE vs CBE at entry.
[c] Mammography + CBE vs CBE annually.
[d] Odds ratio estimated after adjusting for confounding and extending follow-up to 14 years.
[e] Values for the entire cohort.
CBE, clinical breast examination; CI, confidence interval; CNBSS, Canadian National Breast Screening Study; RR, relative risk.

4.3.3 Observational studies

See Table 4.7.

After the success of the Health Insurance Plan trial, several population-based implementation projects and case–control studies evaluated the role of CBE plus mammography for the detection of breast cancer.

In the USA, the Breast Cancer Detection Demonstration Project was initiated by the American Cancer Society and the National Cancer Institute in 1973 (Beahrs & Smart, 1979; Baker, 1982; Morrison et al., 1988). After 5 years of follow-up, 3557 cases of breast cancer had been diagnosed in the screened group, of which 41.6% were detected with mammography alone, 8.7% with CBE alone, and the remainder with both modalities. There was a slight shift to a lower stage; less than 20% of women were diagnosed node-positive, compared with 24% nodal positivity in interval cancers. [Although the Breast Cancer Detection Demonstration Project shows benefit with population-based screening using two modalities and an incremental benefit obtained with CBE, it does not provide effective evidence for the efficacy of CBE in the population.]

The West London study, aiming to screen women older than 40 years in Ealing, London, United Kingdom, began in 1973. Initial screening consisted of two independent CBEs, one by a nurse and one by a doctor, and mammography. Repeat screening was offered after 6, 12, and 24 months to women who had not been diagnosed with breast cancer. Over 3 years, 2484 women were screened, and 83%, 65%, and 53% had repeated screens at 6, 12, and 24 months, respectively. Overall, 34 breast cancers were

detected, of which 5 were interval cancers. Of the 29 cases detected by screening, 80% were at an early stage; 10 (29%) of them were detected with mammography alone, 9 with CBE alone (27%), and 10 with both modalities (Chamberlain et al., 1979).

A multicentre project to assess the effect of breast cancer screening with mammography, CBE, and BSE on mortality was started in 1979 by the UK Trial of Early Detection of Breast Cancer Group (1988). The sensitivity of combined-modality screening (mammography plus CBE) was 92% (197/213) and 91% (235/259) for the Edinburgh and Guildford screening centres, respectively, whereas the sensitivity of CBE screening alone was estimated to be 64% (74/115) for both centres; the incremental detection of CBE over mammography was estimated as 8% (Moss et al., 1993). In the 16-year update on mortality (UK Trial of Early Detection of Breast Cancer Group, 1999), in the cohort offered combined-modality breast cancer screening, breast cancer mortality was 27% lower than in the national population (rate ratio, 0.73; 95% CI, 0.63–0.84). [The Working Group noted that this result could be due to a healthy volunteer effect rather than to reduced mortality from screening.]

In the USA, the National Breast and Cervical Cancer Early Detection Program was started to provide screening to poor and uninsured women in a community setting, using combined CBE and mammography (Bobo et al., 2000). Of 752 081 CBEs performed, 6.9% were abnormal. A total of 2852 invasive and 928 in situ cancers were diagnosed; the diagnostic yield was 5 cancers per 1000 CBEs. Across all ages, the sensitivity, specificity, and PPV of CBE were 58.8%, 93.4%, and 4.3% respectively, based on 1-year survival (consistent with results from most RCTs). About 5.1% of cancers were detected with CBE but not with mammography. [The Working Group noted that the CBE practices varied across medical centres (Bobo & Lee, 2000); however, it was felt that this study provides a real-world outcome of implementing CBE as a screening procedure.]

Bancej et al. (2003) analysed the contribution of CBE in four Canadian organized breast cancer screening programmes. CBE detected 45% of cancers in the first screen, and of these, 11% were detected with CBE alone. In rescreening, CBE detected 39% of cancers, and of these, 16% were detected with CBE alone. Without CBE, the programmes would have missed 3 cancers for every 10 000 screens and 3–10 small invasive cancers for every 100 000 screens. The PPV of CBE was 0.9–1.1%.

Oestreicher et al. (2005) prospectively followed 61 688 women aged 40 years and older who were enrolled in the Breast Cancer Screening Program at Group Health Cooperative of Puget Sound, in Seattle, USA, and underwent at least one screening examination with mammography and/or CBE in 1996–2000. The sensitivity of mammography was 78% and that of combined mammography and CBE was 82%, showing an incremental value of CBE in addition to mammography of 4% (Oestreicher et al., 2005). CBE generally added incrementally more to sensitivity among women with dense breasts.

The effect of breast cancer screening using CBE and mammography has also been evaluated more recently in several settings in Asia. The Well Women Clinics in Hong Kong Special Administrative Region, China, offered breast cancer screening with CBE and mammography to women older than 40 years (and to women aged 35–40 years with a family history of breast cancer) in Hong Kong Special Administrative Region every 2 years. In 1998–2002, 29 028 women were screened, and breast cancer was detected in 232 of them; 83 (36%) cancers were detected with CBE, and 15 of them (6.5% of all detected cancers) were not detected with mammography (Lui et al., 2007). Another breast cancer service was set up at the Hong Kong Sanatorium and Hospital in 1999. Over 8 years, 11 408 asymptomatic women were screened with

CBE and mammography and were given instructions on how to perform BSE. A total of 26 breast cancers were diagnosed; 8 of them (31%) were detected with CBE alone (Kwong et al., 2008).

A screening study to compare CBE, mammography, and ultrasonography was carried out in Chengdu, China, in 2009–2011. Among 3028 women aged 25 years and older who were screened with the three techniques, 33 breast cancers were identified after an average follow-up of 1.3 years; 28 (85%) cancers were detected with mammography, 22 (67%) with CBE, and 24 (73%) with ultrasonography. No cases were detected with CBE that were not detected with mammography, whereas three cancers were detected with ultrasonography that were not detected with the other two methods (Huang et al., 2012).

4.4 Breast self-examination

4.4.1 Randomized trials

Two randomized trials of BSE with breast cancer mortality as the primary end-point have been conducted.

(a) St Petersburg trial

The first randomized trial began in Moscow and St Petersburg, Russian Federation, in 1985. Results on deaths from breast cancer have been reported only from the St Petersburg portion of the study (Semiglazov et al., 1999a, b, 2003). In that city, women aged 40–64 years who received medical care at 18 polyclinics and 10 large industrial businesses with health care services were eligible to participate. Nine polyclinics and five businesses were randomly selected as intervention facilities, and the remainder were control facilities. Women who received medical care at the intervention facilities were invited to participate in the trial. Medical personnel in the clinics examined each woman's breasts, and then the women were given detailed BSE instruction in groups of 5–20 women. Each woman was given a calendar to serve as a reminder to practise BSE monthly and to record the dates of her BSEs. All women were also asked to return annually for reinforcement sessions. Women in the control clinics received CBE at entry into the trial and at annual clinic visits, so this was a trial of the additional benefit of BSE in reducing breast cancer mortality in women screened by annual CBE.

The results are summarized in Table 4.9. Approximately 60 000 women were enrolled in each arm of the study (the exact numbers vary in different reports). Significantly more women in the instruction group than in the control group were referred for evaluation of a breast lump ($P < 0.05$), and more were found to have a benign lesion. Somewhat more women in the instruction group than in the control group were also diagnosed with breast cancer, but the difference could be due to chance ($P > 0.05$), and the malignant tumours in the two groups of women did not differ appreciably in size or percentage with axillary node involvement, suggesting that BSE instruction did not result in breast cancer diagnosis at an earlier, less-advanced stage than would be expected in the absence of BSE instruction. Although survival after diagnosis was somewhat more favourable for cases in the instruction group than those in the control group (65% vs 55% at 9 years; relative survival, 0.77 in log-rank test; 53.9% vs 45.3% at 15 years based on 70–75% follow-up), the difference was not statistically significant ($P > 0.05$). After approximately 10 years of follow-up, almost equal percentages of women in the two groups had died of breast cancer.

[In addition to the possibility that BSE would not be efficacious under any circumstances, there are three possible explanations for the results of this study. One is poor compliance with the BSE instruction. Based on a sample of the participants 1 year after BSE training, 82% of the women interviewed reported practising BSE more than 5 times per year, and 53% reported monthly BSE practice. However, by year 4, these percentages

Table 4.9 Results of randomized trials of breast self-examination

Characteristic	St Petersburg trial[a]		Shanghai trial[b]	
	Intervention	Control	Intervention	Control
Age range (years)	40–64	40–64	30–64	30–64
No. of women	57 712	64 759	1 329 769	133 085
No. (%) referred for evaluation/ benign breast lesions[c]	4300 (7.5%)	2438 (3.8%)	2387 (1.8%)	1296 (1.0%)
No. (%) with breast cancer[d]	493 (0.9%)	446 (0.7%)	864 (0.7%)	896 (0.7%)
No. (%) of deaths from breast cancer	157 (0.27%)	167 (0.26%)	135 (0.1%)	131 (0.1%)

[a] From Semiglazov et al. (1999a, b).
[b] From Thomas et al. (2002).
[c] Number referred for further evaluation in the St Petersburg trial, and number of histologically confirmed benign lesions in the Shanghai trial.
[d] After about 10 years in the St Petersburg trial and after 10–11 years in the Shanghai trial.

had dropped to 52% and 18%, respectively. After a re-education programme in 1994, these percentages increased to 76% and 32%, respectively, 8 years after the trial was initiated. Also in 1994, medical personnel observed a random sample of about 400 women practising BSE and recorded their proficiency. Although the reported frequency of correctly practising various aspects of BSE was high, there is no evidence that these observations accurately reflect the routine practice of BSE outside of the clinic setting by all of the women in the instruction group. A second possible reason for the results is that BSE is not effective in reducing mortality from breast cancer in women who are also screened by CBE. A third possible explanation is that women in both groups had easy access to medical care at the polyclinics, and women in the control group tended to present with tumours that were small and at an early stage. Of women in the control group, 17.4% presented with tumours less than 2 cm in diameter, and 46.4% with tumours that had not spread to the axillary lymph nodes.]

(b) Shanghai trial

The second randomized trial was conducted in Shanghai, China (Thomas et al., 1997, 2002). In 1989–1991, more than 266 000 women aged 30–64 years who were current or retired employees of the Shanghai Textile Industry Bureau, working in 519 different factories, were randomized by factory to a BSE instruction group or a control group. Women in the instruction group received initial BSE instruction in groups of about 10 women and two subsequent reinforcement sessions, 1 year and 3 years later, consisting of videos and discussion groups, as well as multiple reminders to practise BSE. Nearly 80% of the women attended all three sessions. In addition, women were asked to attend periodic practice sessions supervised by factory medical workers about every 6 months for 4–5 years. During the first year of the study, 92% of the women attended these sessions; this percentage gradually declined to 74% in the fourth year and 49% in the fifth and last year of the intervention. The women thus practised BSE under supervision on average once every 4–5 months during the first 4–5 years of the trial. The quality of the BSEs at these sessions was high. Women were encouraged to practise BSE monthly, but the frequency and quality of the practice outside of the clinic setting are unknown. No breast cancer screening was offered to women in the control group. A higher level of proficiency in detecting lumps in silicone breast models was demonstrated by randomly selected women in the instruction group compared with the control group.

The results after 10–11 years of follow-up are summarized in Table 4.9. More women were diagnosed with benign breast lesions in the instruction group than in the control group. The numbers of women with breast cancer were similar in the two groups. The breast cancers in the two groups did not differ appreciably in size (44.9% vs 41.6% were ≤ 2 cm in diameter) or stage (47.0% vs 48.3% had no axillary nodal involvement). Also, the numbers of deaths from breast cancer and the cumulative breast cancer mortality rates were nearly identical in the two groups, as were survival rates in women with breast cancer, both from entry into the trial and from date of diagnosis. Evidence was presented that these results cannot be readily explained by the absence of statistical power, insufficient duration or completeness of follow-up, failure of the randomization procedure to select two groups at equal risk of breast cancer, selective exclusions of women after randomization, incomplete or differential ascertainment of breast cancer cases or deaths, screening in the control group, or insufficient breast cancer treatment. The most likely reason for the absence of an effect of BSE instruction on breast cancer mortality in this study is that proficient BSE practice at least once every 5 months for 4–5 years did not result in breast cancer being diagnosed at a sufficiently less advanced stage of progression for appropriate therapy to have altered the course of the disease. There is suggestive evidence that more frequent BSE might have resulted in a more favourable trial result. Among women who attended all of the supervised BSE sessions and those who attended fewer than 70% of the sessions, the percentages with tumours that were less than 2 cm in diameter were 52.3% and 45.3%, respectively, in current workers, and 48.7% and 44.4%, respectively, in retired women.

In summary, the results from both randomized controlled trials provided little evidence that risk of death or of advanced disease is reduced by BSE instruction. In both studies, the women in the control group had easy access to medical care and tended to present with relatively small tumours without regional lymph-node involvement. The efficacy of BSE in populations in which women typically present with more-advanced tumours remains unknown.

4.4.2 Observational studies

(a) Methodological considerations

In evaluating the evidence for the efficacy of BSE from observational studies, several methodological issues must be considered.

BSE must be distinguished from breast awareness. BSE is a screening method used to attempt to detect *asymptomatic* breast cancer before it is clinically apparent (see Section 2.4 for technical details). Breast awareness consists of the education and encouragement of women to seek medical attention for *symptomatic* changes in their breasts that may be due to the presence of breast cancer (see Section 1.5.1 for additional details). These two concepts of breast cancer detection are not always clearly defined or distinguished (Thornton & Pillarisetti, 2008; Mark et al., 2014). Self-reports of BSE practice may include breast awareness, and some cancers that are reported as being detected by BSE may have been symptomatic cancers found by the women themselves through breast awareness.

There are two components to BSE compliance: frequency (typically once a month) and proficiency; these are not consistently considered and reported in observational studies. In addition, there may be underreporting or misclassification of BSE practice. These reporting errors would lead to underestimation of the efficacy of BSE in cohort studies. In case–control studies, if the magnitudes of the reporting errors are different for cases and controls, spurious associations would arise. Finally, the practice of BSE may be related to risk factors for breast cancer, or to other methods of screening, and lead to spurious

results if the potential confounding effect of these associations is not taken into account.

There have been a large number of clinical studies of tumour size and stage at diagnosis, and of survival from date of diagnosis, in relation to whether the patient reported that the tumour was detected by BSE, and in relation to reported frequency of BSE practice (IARC, 2002). In most studies, the proportion of women who had early-stage cancer was slightly higher in women who reported detecting their cancer by BSE than in women whose cancer was detected by other means (excluding mammography screening). However, it is not clear whether the women who reported detecting their tumour by BSE were actually practising BSE or whether they were women who simply reported having found their tumour by themselves. Among cases who reported a history of practising BSE, tumour stage was not consistently related to reported BSE frequency. Most studies did show a tendency towards slightly smaller tumour size in women who reported practising BSE monthly than in women who reported practising BSE less frequently, but differential reporting of BSE frequency by women with small and large tumours cannot be ruled out. Survival tended to be somewhat longer in women reporting a history of BSE practice, or who were taught BSE or accepted an invitation to attend a BSE instruction session, than in women not reporting any of these factors, but the magnitude of the differences varied widely among the studies, the differences were not consistently statistically significant, and enhanced lead-time or length bias sampling cannot be ruled out as alternative explanations for the observations. The results of these observational studies of intermediate end-points may thus all be due to bias, confounding, or chance, and the Working Group therefore concluded that they do not contribute meaningful information in formulating an assessment of the efficacy of BSE. These studies will therefore not be considered further in this review. One more-recent study in the USA (Tu et al., 2006) assessed BSE practice before the development of breast cancer, thus avoiding possible reporting bias, and found no association between the quality of BSE practice and either tumour size or stage of disease.

The two randomized trials evaluated the efficacy of BSE instruction, not the actual practice of BSE. The evidence from observational studies that BSE can reduce mortality from breast cancer and detect interval cancers between periodic screenings is reviewed in this section.

(b) Cohort studies

Reports are available from three studies in which breast cancer mortality rates were compared in women who did and did not practise BSE.

Holmberg et al. (1997) calculated breast cancer mortality rates in a cohort of women in the USA who in 1959 were asked a single question: "Many doctors recommend that women examine their breasts monthly. Do you do so?" A "yes" answer presumably indicated that the women practised BSE monthly, and a "no" answer indicated that BSE either was practised less frequently or was not practised. After a 13-year follow-up period, no association was observed between breast cancer mortality and the answer to this question. [The major strengths of this study are its large size, long duration of follow-up, strong statistical power, and control for multiple possible confounders. However, the absence of any detailed information on the frequency or manner of BSE practice by the women in the study reduces the usefulness of the negative findings, since many of the women who reported practising BSE may not have done so adequately.]

In the Mama Program for Breast Screening in Finland (Gastrin et al., 1994), beginning in 1973 women were given detailed BSE instruction in groups of 20–50 women, followed by periodic reminders and annual mailings of calendars for the women to record their BSE

practice. Mortality rates in the participants were compared with those in the general population of Finland. The breast cancer mortality rate in the participants was significantly lower than expected (mortality rate ratio, 0.71). This occurred in spite of a higher incidence rate of breast cancer in the participants than expected (incidence rate ratio, 1.19). The reduced rates of death from breast cancer were observed in most age groups of women and were most pronounced in years 3–4 after entry into the study. However, mortality rates from all causes were also significantly lower by the same amount as for breast cancer mortality (standardized mortality ratio, 0.70), suggesting that the participants were healthier than women in the general population, and that their lower breast cancer mortality may have been due to factors related to improved survival, other than early diagnosis resulting from BSE practice, that were not controlled for in the analysis. This contention is supported by the observation that the stage of disease at diagnosis was no different in the women in the study cohort than in other cases in the country. [There is no mention of CBE or mammography screening in the published report, and these screening methods were presumably not taken into account in the data analysis, although the frequency of their use was probably low.] A large majority of the women in the cohort reported on their calendars that they had practised BSE monthly. [This information was not validated and is therefore questionable, and proficiency of BSE practice was not assessed.]

As part of the United Kingdom Trial of Early Detection of Breast Cancer (Ellman et al., 1993; UK Trial of Early Detection of Breast Cancer Group, 1999), women in the cities of Huddersfield and Nottingham were invited to attend BSE education sessions. The sessions included a talk and a film demonstrating BSE. In Huddersfield, calendars were mailed annually, as reminders and as a means to record monthly BSE practice. No further BSE instruction was provided in either city. Breast cancer mortality rates in the women invited to the BSE training session (whether or not they attended) were compared with those in four comparison centres in which women received no breast cancer screening or BSE instruction. No overall difference in breast cancer mortality rates was observed between the women in the two BSE instruction centres combined and the women in the four comparison centres (rate ratio, 0.99; 95% CI, 0.87–1.12). However, the rate ratio in Huddersfield was significantly less than 1 (0.79; 95% CI, 0.65–0.96) and was similar to that observed in the Mama Program for Breast Screening in Finland; at the Huddersfield centre, as in the programme in Finland, calendars were mailed annually, suggesting that the difference could be due to more intensive BSE practice in Huddersfield than in Nottingham (rate ratio, 1.09; 95% CI, 0.95–1.26). In addition, more women in Huddersfield than in Nottingham also received breast-conserving surgery, chemotherapy, and tamoxifen, whereas participation rates in the BSE instruction sessions were higher in Nottingham than in Huddersfield, suggesting that differences in treatment or other factors could explain the discrepant results. No information on compliance was reported.

In summary, although the cohort studies in Finland and the United Kingdom (Huddersfield component) showed that BSE instruction with periodic reminders was associated with a small reduction in breast cancer mortality, it is more likely that these observations are due to factors unrelated to BSE practice. No reliable information on compliance was provided for any of the studies. In the study in the USA, BSE practice was defined by a single question, and in the studies in Finland and the United Kingdom, BSE instruction was given in a single session with no reinforcement sessions. It is therefore reasonable to assume that the frequency and proficiency of BSE practice by the women in these three studies was lower than those in the two randomized trials, which provided more intensive BSE

instruction and encouragement to practise, and that the results provide no information on the efficacy of BSE in women who practise BSE regularly and competently.

(c) Case–control studies

Two case–control studies that were nested in prospective studies, and thus did not rely on self-reported BSE practice, have been conducted.

Locker et al. (1989) performed a case–control analysis of data from women invited to enrol in the United Kingdom Trial of Early Detection of Breast Cancer in Nottingham. Of 180 women who died of breast cancer more than 3 months after invitation, 68 (37.8%) had attended the BSE instruction class, compared with 258 (42.8%) of 603 age-matched control women at the Nottingham centre, for an estimated relative risk of 0.70 (95% CI, 0.50–0.97). The comparable relative risk estimate in premenopausal women was 0.85 (95% CI, 0.45–1.60) and in postmenopausal women was 0.66 (95% CI, 0.45–0.97). [These estimates were not controlled for factors other than age that may have been associated with a decision to attend the BSE instruction class, or for treatment or other factors that could influence survival.]

Harvey et al. (1997) conducted a case–control study nested within the CNBSS. Answers to questions about frequency of BSE obtained before enrolment in the trial and during the trial and results of annual assessment of BSE proficiency were compared in 220 cases with fatal or metastatic disease and 2200 age-matched controls selected from trial enrollees. All of the information on BSE was obtained before the development of breast cancer in the cases. Compared with women who practised BSE before enrolment, those who did not had a relative risk of fatal or advanced breast cancer of 1.27 (95% CI, 0.96–1.68), and relative risk estimates decreased with increasing frequency of BSE practice before enrolment. The relative risk of fatal or advanced disease also increased slightly with decreasing frequency of BSE practice during the trial, but none of the estimates or trends were statistically significant ($P > 0.05$). However, there was a significant decrease in estimates of relative risk of fatal or advanced disease with increasing BSE proficiency as observed in clinics by trained examiners 2 years before diagnosis in the cases (Table 4.10). The level of proficiency was defined according to the exclusion of one, two, or three key elements of a proper BSE (visual inspection, use of three middle fingers, and use of finger pads) that were weakly associated with a reduction in risk. Similar but weaker trends in risk were observed in relation to these same levels of proficiency at 1 year and 3 years before diagnosis, but none of the relative risk estimates had 95% confidence limits that excluded 1.0. Also, other elements of BSE practice (systematic search, circular palpation, complete coverage of the breast, and examination of the axilla) were not associated with changes in risk estimates. The relative risk estimates were not found to be confounded by family history of breast cancer, age at menarche or menopause, education level, occupation, or the trial arm to which the woman was allocated.

Two additional case–control studies, which were conducted in the general population and relied on results of interviews with women to obtain information on BSE practice, have been conducted. Both included women with advanced disease (as a surrogate for death from breast cancer) as cases.

In the USA, Newcomb et al. (1991) compared BSE practice in 209 enrollees in a prepaid health plan who developed late-stage (stage III or IV) breast cancer during a defined period of time with BSE practice in 433 age-matched controls selected randomly from enrollees in the same plan. Personal interviews with the women were conducted in which specific questions were asked about various components of the recommended techniques and frequency of practice. Both an open-ended technique and a structured

Table 4.10 Relative risk of death from breast cancer or of advanced disease in relation to proficiency of breast self-examination

Reference, country	Years before diagnosis that assessment was performed	Measure of proficiency	RR (95% CI)
Harvey et al. (1997), Canada	1	All 3 practices included[a]	1.00 (ref)
		1 practice omitted	1.52 (0.93–2.48)
		2 practices omitted	1.53 (0.83–2.84)
		3 practices omitted	1.40 (0.58–3.39)
	2	All 3 practices included[a]	1.00 (ref)
		1 practice omitted	1.82 (1.00–3.29)
		2 practices omitted	2.84 (1.44–5.59)
		3 practices omitted	2.95 (1.19 –7.30)
	3	All 3 practices included[a]	1.00 (ref)
		1 practice omitted	1.21 (0.65–2.28)
		2 practices omitted	0.92 (0.38–2.22)
		3 practices omitted	1.68 (0.59–4.76)
Newcomb et al. (1991), USA	After diagnosis[b]	High proficiency[c]	0.65 (0.33–1.31) [ref][d]
		Moderate proficiency	1.00 (0.56–1.80) [1.53]
		Low proficiency	1.33 (0.83–2.12) [2.05]
		No BSE practice	1.00 (ref) [1.53]

[a] Includes visual inspection, use of three middle fingers, and use of finger pads.
[b] Women were asked about BSE practice 1 year before the date of diagnosis in cases or a comparable reference date in controls.
[c] Proficiency based on a 10-point scoring system of items included in responses to an open-ended questionnaire.
[d] Relative risks in square brackets with high proficiency as the reference category were calculated by the Working Group.
BSE, breast self-examination; ref, reference; RR, relative risk.

interview were used to classify BSE as to level of proficiency. The relative risk of advanced disease in women who ever practised BSE was 1.15 (95% CI, 0.73–1.81), and the relative risk unexpectedly increased with the frequency of BSE practice. However, the women who practised BSE frequently were found to practise it with the lowest level of proficiency, and the relative risk of advanced disease decreased with increasing level of proficiency (Table 4.10). This trend was observed in women with all levels of BSE frequency. [Although the influence of the presence of the disease on responses could have biased this study, it seems unlikely that cases would underreport frequency of BSE practice and overreport proficiency during the same detailed interviews. The relative risk estimates were controlled for age and frequency of CBE. Other risk factors for breast cancer were considered as possible confounders but were found not to alter the values of the estimates.]

Muscat & Huncharek (1991) compared 435 women in Connecticut, USA, with regional or distant breast cancer at diagnosis with 887 control women selected by random-digit dialling. Frequency of BSE practice was ascertained during detailed interviews as part of a larger study on steroid hormones and cancer. No information on proficiency was obtained. BSE practice at least once a month was reported by 27.4% of the cases and 20.5% of the controls. After controlling for family history of breast cancer, age at first birth, race, and frequency of mammograms, a relative risk of 1.27 (95% CI, 0.77–2.07) was estimated, but it is not clear from the report whether this estimate is for women who practised BSE monthly or also less frequently. [As in the study by Newcomb et al. (1991), risk increased with the frequency of

BSE practice, but unlike that study, no information on proficiency was obtained, so it is not known whether this trend is due to confounding by proficiency.]

In summary, the results from case–control studies provided little evidence that risk of death from breast cancer or of advanced disease is reduced by frequent practice of BSE as it is generally practised by women in North America and the United Kingdom. Two of the case–control studies provided evidence to suggest that risk of fatal or advanced disease could be reduced if BSE were practised with a high degree of proficiency. It can be assumed that the documented practice of BSE in the Shanghai trial was performed with a high degree of proficiency, because it was observed by health workers and was the result of intensive instruction over a period of several years; however, such practice about once every 4–5 months for 4–5 years was insufficient to reduce mortality from breast cancer. The efficacy of more frequent, high-proficiency BSE in reducing mortality remains unknown.

(d) Detection of interval cancers

The previous IARC Working Group on breast cancer screening (IARC, 2002) recommended that studies be conducted to assess the efficacy of BSE in detecting interval cancers between periodic mammography screenings. Results of only one such study have been published (Wilke et al., 2009). It involved women who were at high risk of breast cancer (estimated average lifetime risk, > 20%) and therefore probably more highly motivated to practise BSE than other women. A high-risk breast clinic at Duke University, USA, recruited 147 women who had a 5-year Gail-model risk of at least 1.7% and followed them up for an average of 23 months (range, 6–36 months). Risk factors included: a previous histologically confirmed diagnosis of atypical hyperplasia or lobular carcinoma in situ or DCIS; a contralateral invasive breast cancer; a *BRCA1/2* mutation; radiation treatment for Hodgkin lymphoma to the chest, neck, and axilla; or one or more first-degree relatives with premenopausal breast cancer. The women were screened annually with mammography and magnetic resonance imaging (MRI). They also received 6–15 minutes of BSE instruction in conjunction with CBE two or three times a year, and their self-reported home practice of BSE was recorded at each of these sessions. Breast cancer was detected in 12 women, 1 during initial training and 11 during the follow-up period. All 12 women with breast cancer were judged to have complied with the recommendations to practise BSE monthly. Six of the cancers were initially found by BSE (sensitivity, 50%), as were 18 additional masses that were confirmed as not being breast cancer (PPV, 25%). The 5 cases detected by BSE during the follow-up period were detected 6–11 months after the last annual screening.

These results suggest that BSE may be useful in detecting interval cancers in women at high risk of breast cancer who are highly motivated to practise BSE regularly and competently. No information is available to determine whether this would contribute to a reduction in mortality from breast cancer.

References

Alexander F, Roberts MM, Lutz W, Hepburn W (1989). Randomisation by cluster and the problem of social class bias. *J Epidemiol Community Health*, 43(1):29–36. doi:10.1136/jech.43.1.29 PMID:2592888

Alexander FE (1997). The Edinburgh Randomized Trial of Breast Cancer Screening. *J Natl Cancer Inst Monogr*, (22):31–5. PMID:9709272

Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE et al. (1994). The Edinburgh randomised trial of breast cancer screening: results after 10 years of follow-up. *Br J Cancer*, 70(3):542–8. doi:10.1038/bjc.1994.342 PMID:8080744

Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE et al. (1999). 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet*, 353(9168):1903–8. doi:10.1016/S0140-6736(98)07413-3 PMID:10371567

Allgood PC, Warwick J, Warren RM, Day NE, Duffy SW (2008). A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. *Br J Cancer*, 98(1):206–9. doi:10.1038/sj.bjc.6604123 PMID:18059396

Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F et al. (1988). Mammographic screening and mortality from breast cancer: the Malmö mammographic screening trial. *BMJ*, 297(6654):943–8. doi:10.1136/bmj.297.6654.943 PMID:3142562

Andersson I, Janzon L (1997). Reduced breast cancer mortality in women under age 50: updated results from the Malmö Mammographic Screening Program. *J Natl Cancer Inst Monogr*, 22(22):63–7. PMID:9709278

Autier P, Héry C, Haukka J, Boniol M, Byrnes G (2009). Advanced breast cancer and breast cancer mortality in randomized controlled trials on mammography screening. *J Clin Oncol*, 27(35):5919–23. doi:10.1200/JCO.2009.22.7041 PMID:19884547

Bailar JC 3rd, MacMahon B (1997). Randomization in the Canadian National Breast Screening Study: a review for evidence of subversion. *CMAJ*, 156(2):193–9. PMID:9012720

Baines CJ, Christen A, Simard A, Wall C, Dean D, Duncan L et al. (1989). The National Breast Screening Study: pre-recruitment sources of awareness in participants. *Can J Public Health*, 80(3):221–5. PMID:2743247

Baker LH (1982). Breast Cancer Detection Demonstration Project: five-year summary report. *CA Cancer J Clin*, 32(4):194–225. doi:10.3322/canjclin.32.4.194 PMID:6805867

Baker SG, Kramer BS, Prorok PC (2002). Statistical issues in randomized trials of cancer screening. *BMC Med Res Methodol*, 2(1):11. doi:10.1186/1471-2288-2-11 PMID:12238954

Bancej C, Decker K, Chiarelli A, Harrison M, Turner D, Brisson J (2003). Contribution of clinical breast examination to mammography screening in the early detection of breast cancer. *J Med Screen*, 10(1):16–21. doi:10.1258/096914103321610761 PMID:12790311

Barton MB, Harris R, Fletcher SW (1999). The rational clinical examination. Does this patient have breast cancer? The screening clinical breast examination: should it be done? How? *JAMA*, 282(13):1270–80. doi:10.1001/jama.282.13.1270 PMID:10517431

Beahrs OH, Smart CR (1979). Diagnosis of minimal breast cancers in the BCDDP: the 66 questionable cases. *Cancer*, 43(3):848–50. doi:10.1002/1097-0142(197903)43:3<848::AID-CNCR2820430310>3.0.CO;2-1 PMID:427726

Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L (2007). Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol*, 8(12):1129–38. doi:10.1016/S1470-2045(07)70380-7 PMID:18054882

Bjurstam N, Björneld L, Duffy SW, Smith TC, Cahlin E, Eriksson O et al. (1997). The Gothenburg breast screening trial: first results on mortality, incidence, and mode of detection for women ages 39–49 years at randomization. *Cancer*, 80(11):2091–9. doi:10.1002/(SICI)1097-0142(19971201)80:11<2091::AID-CNCR8>3.0.CO;2-# PMID:9392331

Bjurstam N, Björneld L, Warwick J, Sala E, Duffy SW, Nyström L et al. (2003). The Gothenburg Breast Screening Trial. *Cancer*, 97(10):2387–96. doi:10.1002/cncr.11361 PMID:12733136

Black WC, Haggstrom DA, Welch HG (2002). All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*, 94(3):167–73. doi:10.1093/jnci/94.3.167 PMID:11830606

Bobo J, Lee N (2000). Factors associated with accurate cancer detection during a clinical breast examination. *Ann Epidemiol*, 10(7):463. doi:10.1016/S1047-2797(00)00099-5 PMID:11018380

Bobo JK, Lee NC, Thames SF (2000). Findings from 752,081 clinical breast examinations reported to a national screening program from 1995 through 1998. *J Natl Cancer Inst*, 92(12):971–6. doi:10.1093/jnci/92.12.971 PMID:10861308

Boyd NF, Jong RA, Yaffe MJ, Tritchler D, Lockwood G, Zylak CJ (1993). A critical appraisal of the Canadian National Breast Cancer Screening Study. *Radiology*, 189(3):661–3. doi:10.1148/radiology.189.3.8234686 PMID:8234686

Breast Screening Frequency Trial Group (2002). The frequency of breast cancer screening: results from the UKCCCR Randomised Trial. *Eur J Cancer*, 38(11):1458–64. doi:10.1016/S0959-8049(01)00397-5 PMID:12110490

Canadian Task Force on Preventive Health Care (2011). Recommendations on screening for breast cancer in average-risk women aged 40–74 years. *CMAJ*, 183(17):1991–2001. doi:10.1503/cmaj.110334 PMID:22106103

Chamberlain J, Clifford RE, Nathan BE, Price JL, Burn I (1979). Error-rates in screening for breast cancer by clinical examination and mammography. *Clin Oncol*, 5(2):135–46. PMID:466892

Cole P, Morrison AS (1980). Basic issues in population screening for cancer. *J Natl Cancer Inst*, 64(5):1263–72. PMID:6767876

Collette HJ, Day NE, Rombach JJ, de Waard F (1984). Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study. *Lancet*, 1(8388):1224–6. doi:10.1016/S0140-6736(84)91704-5 PMID:6144934

Collette HJ, de Waard F, Rombach JJ, Collette C, Day NE (1992). Further evidence of benefits of a (non-randomised) breast cancer screening programme: the DOM project. *J Epidemiol Community Health*, 46(4):382–6. doi:10.1136/jech.46.4.382 PMID:1431712

Collette HJA (1985). Attempts to evaluate a non-randomized breast cancer screening programme (the 'DOM-project'). *Maturitas*, 7(1):43–50. doi:10.1016/0378-5122(85)90033-7 PMID:4021828

Cuzick J, Edwards R, Segnan N (1997). Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med*, 16(9):1017–29. doi:10.1002/(SICI)1097-0258(19970515)16:9<1017::AID-SIM508>3.0.CO;2-V PMID:9160496

de Gelder R, Heijnsdijk EAM, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ (2011). Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev*, 33(1):111–21. doi:10.1093/epirev/mxr009 PMID:21709144

de Koning HJ, Boer R, Warmerdam PG, Beemsterboer PM, van der Maas PJ (1995). Quantitative interpretation of age-specific mortality reductions from the Swedish breast cancer-screening trials. *J Natl Cancer Inst*, 87(16):1217–23. doi:10.1093/jnci/87.16.1217 PMID:7563167

de Koning HJ (2009). The mysterious mass(es). [Inaugural address, Professor of Screening Evaluation.] Rotterdam, Netherlands: Erasmus MC. Available from: http://repub.eur.nl/res/pub/30689/oratie.pdf.

De Waard F, Collette HJ, Rombach JJ, Baanders-van Halewijn EA, Honing C (1984). The DOM project for the early detection of breast cancer, Utrecht, The Netherlands. *J Chronic Dis*, 37(1):1–44. doi:10.1016/0021-9681(84)90123-1 PMID:6690457

Dean PB (2007). A withdrawn prepublication. *Lancet*, 369(9565):901. doi:10.1016/S0140-6736(07)60434-6 PMID:17368138

Duffy SW, Parmar D (2013). Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. *Breast Cancer Res*, 15(3):R41. doi:10.1186/bcr3427 PMID:23680223

Duffy SW, Tabár L, Vitak B, Day NE, Smith RA, Chen HH et al. (2003b). The relative contributions of screen-detected in situ and invasive breast carcinomas in reducing mortality from the disease. *Eur J Cancer*, 39(12):1755–60. PMID:12888371

Duffy SW, Tabár L, Vitak B, Yen MF, Warwick J, Smith RA et al. (2003a). The Swedish Two-County Trial of mammographic screening: cluster randomisation and end point evaluation. *Ann Oncol*, 14(8):1196–8. doi:10.1093/annonc/mdg322 PMID:12881376

Ellman R, Moss SM, Coleman D, Chamberlain J (1993). Breast self-examination programmes in the Trial of Early Detection of Breast Cancer: ten year findings. *Br J Cancer*, 68(1):208–12. doi:10.1038/bjc.1993.315 PMID:8318415

Fagerberg CJG, Tabár L (1988). The results of periodic one-view mammography screening in a randomized, controlled trial in Sweden. In: Day NE, Miller AB, editors. Screening for breast cancer. Toronto, Canada: Hans Huber; pp. 33–8.

Frisell J, Eklund G, Hellström L, Lidbrink E, Rutqvist LE, Somell A (1991). Randomized study of mammography screening – preliminary report on mortality in the Stockholm trial. *Breast Cancer Res Treat*, 18(1):49–56. doi:10.1007/BF01975443 PMID:1854979

Frisell J, Glas U, Hellström L, Somell A (1986). Randomized mammographic screening for breast cancer in Stockholm. Design, first round results and comparisons. *Breast Cancer Res Treat*, 8(1):45–54. doi:10.1007/BF01805924 PMID:3790749

Frisell J, Lidbrink E (1997). The Stockholm Mammographic Screening Trial: risks and benefits in age group 40–49 years. *J Natl Cancer Inst Monogr*, 22(22):49–51. PMID:9709275

Frisell J, Lidbrink E, Hellström L, Rutqvist LE (1997). Followup after 11 years – update of mortality results in the Stockholm mammographic screening trial. *Breast Cancer Res Treat*, 45(3):263–70. doi:10.1023/A:1005872617944 PMID:9386870

Gastrin G, Miller AB, To T, Aronson KJ, Wall C, Hakama M et al. (1994). Incidence and mortality from breast cancer in the Mama Program for Breast Screening in Finland, 1973–1986. *Cancer*, 73(8):2168–74. doi:10.1002/1097-0142(19940415)73:8<2168::AID-CNCR2820730822>3.0.CO;2-V PMID:8156521

Gøtzsche PC, Jørgensen KJ (2013). Screening for breast cancer with mammography. *Cochrane Database Syst Rev*, 6:CD001877. doi:10.1002/14651858.CD001877.pub5 PMID:23737396

Gøtzsche PC, Olsen O (2000). Is screening for breast cancer with mammography justifiable? *Lancet*, 355(9198):129–34. doi:10.1016/S0140-6736(99)06065-1 PMID:10675181

Gulati R, Tsodikov A, Wever EM, Mariotto AB, Heijnsdijk EA, Katcher J et al. (2012). The impact of PLCO control arm contamination on perceived PSA screening efficacy. *Cancer Causes Control*, 23(6):827–35. doi:10.1007/s10552-012-9951-8 PMID:22488488

Hanley JA (2011). Measuring mortality reductions in cancer screening trials. *Epidemiol Rev*, 33(1):36–45. PMID:21624962

Harvey BJ, Miller AB, Baines CJ, Corey PN (1997). Effect of breast self-examination techniques on the risk of death from breast cancer. *CMAJ*, 157(9):1205–12. PMID:9361639

Holmberg L, Duffy SW, Yen AMF, Tabár L, Vitak B, Nyström L et al. (2009). Differences in endpoints between the Swedish W-E (two county) trial of mammographic screening and the Swedish overview: methodological consequences. *J Med Screen*, 16(2):73–80. doi:10.1258/jms.2009.008103 PMID:19564519

Holmberg L, Ekbom A, Calle E, Mokdad A, Byers T (1997). Breast cancer mortality in relation to self-reported use of breast self-examination. A cohort study

of 450,000 women. *Breast Cancer Res Treat*, 43(2):137–40. doi:10.1023/A:1005788729145 PMID:9131269

Holmberg LH, Tabár L, Adami HO, Bergström R (1986). Survival in breast cancer diagnosed between mammographic screening examinations. *Lancet*, 328(8497):27–30. doi:10.1016/S0140-6736(86)92569-9 PMID:2873324

Huang Y, Kang M, Li H, Li JY, Zhang JY, Liu LH et al. (2012). Combined performance of physical examination, mammography, and ultrasonography for breast cancer screening among Chinese women: a follow-up study. *Curr Oncol*, 19(Suppl 2):eS22–30. doi:10.3747/co.19.1137 PMID:22876165

IARC (2002). Breast cancer screening. *IARC Handb Cancer Prev*, 7:1–229. Available from: http://www.iarc.fr/en/publications/pdfs-online/prev/handbook7/Handbook7_Breast.pdf.

Independent UK Panel on Breast Cancer Screening (2012). The benefits and harms of breast cancer screening: an independent review. *Lancet*, 380(9855):1778–86. doi:10.1016/S0140-6736(12)61611-0 PMID:23117178

Johns LE, Moss SM, Cuckle H, Bobrow L, Evans A, Kutt E et al.; Age Trial Management Group (2010a). False-positive results in the randomized controlled trial of mammographic screening from age 40 ("Age" trial). *Cancer Epidemiol Biomarkers Prev*, 19(11):2758–64. doi:10.1158/1055-9965.EPI-10-0623 PMID:20837718

Johns LE, Moss SM; Trial Management Group (2010b). Randomized controlled trial of mammographic screening from age 40 ('Age' trial): patterns of screening attendance. *J Med Screen*, 17(1):37–43. doi:10.1258/jms.2010.009091 PMID:20356944

Kerr M (1991). A case-control study of treatment adequacy and mortality from breast cancer for women age 40–49 years at entry into the National Breast Screening Study [dissertation]. Toronto, Canada: University of Toronto.

Kingston N, Thomas I, Johns L, Moss S; Trial Management Group (2010). Assessing the amount of unscheduled screening ("contamination") in the control arm of the UK "Age" Trial. *Cancer Epidemiol Biomarkers Prev*, 19(4):1132–6. doi:10.1158/1055-9965.EPI-09-0996 PMID:20233850

Kopans D (2014). Re: Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. http://www.bmj.com/content/348/bmj.g366?tab=responses

Kopans DB (1990). The Canadian screening program: a different perspective. *AJR Am J Roentgenol*, 155(4):748–9. doi:10.2214/ajr.155.4.ajronline_155_4_001

Kopans DB (1993). Mammography. *Lancet*, 341(8850):957 doi:10.1016/0140-6736(93)91246-I PMID:8096288

Kopans DB, Feig SA (1993). The Canadian National Breast Screening Study: a critical review. *AJR Am J Roentgenol*, 161(4):755–60. doi:10.2214/ajr.161.4.8372752 PMID:8372752

Kwong A, Cheung PS, Wong AY, Hung GT, Lo G, Tsao M et al. (2008). The acceptance and feasibility of breast cancer screening in the East. *Breast*, 17(1):42–50. doi:10.1016/j.breast.2007.06.005 PMID:17720500

Locker AP, Caseldine J, Mitchell AK, Blamey RW, Roebuck EJ, Elston CW (1989). Results from a seven-year programme of breast self-examination in 89,010 women. *Br J Cancer*, 60(3):401–5. doi:10.1038/bjc.1989.294 PMID:2789950

Lui CY, Lam HS, Chan LK, Tam KF, Chan CM, Leung TY et al. (2007). Opportunistic breast cancer screening in Hong Kong; a revisit of the Kwong Wah Hospital experience. *Hong Kong Med J*, 13(2):106–13. PMID:17406037

Magnus MC, Ping M, Shen MM, Bourgeois J, Magnus JH (2011). Effectiveness of mammography screening in reducing breast cancer mortality in women aged 39–49 years: a meta-analysis. *J Womens Health (Larchmt)*, 20(6):845–52. doi:10.1089/jwh.2010.2098 PMID:21413892

Mark K, Temkin SM, Terplan M (2014). Breast self-awareness: the evidence behind the euphemism. *Obstet Gynecol*, 123(4):734–6. doi:10.1097/AOG.0000000000000139 PMID:24785598

Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M (2013). The benefits and harms of breast cancer screening: an independent review. *Br J Cancer*, 108(11):2205–40. doi:10.1038/bjc.2013.177 PMID:23744281

Miller AB (2001). Screening for breast cancer with mammography. *Lancet*, 358(9299):2164, author reply 2167–8. doi:10.1016/S0140-6736(01)07189-6 PMID:11784651

Miller AB (2004). Commentary: a defence of the Health Insurance Plan (HIP) study and the Canadian National Breast Screening Study (CNBSS). *Int J Epidemiol*, 33(1):64–5, discussion 69–73. doi:10.1093/ije/dyh015 PMID:15075146

Miller AB (2014). Implications of the Canadian National Breast Screening Study. *Womens Health (Lond Engl)*, 10(4):345–7. doi:10.2217/whe.14.25 PMID:25259895

Miller AB, Baines CJ, Sickles EA (1990). Canadian National Breast Screening Study. *AJR Am J Roentgenol*, 155(5):1133–4. doi:10.2214/ajr.155.5.2120947 PMID:2120947

Miller AB, Baines CJ, To T, Wall C (1992a). Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *CMAJ*, 147(10):1459–76. PMID:1423087

Miller AB, Baines CJ, To T, Wall C (1992b). Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *CMAJ*, 147(10):1477–88. PMID:1423088

Miller AB, Howe GR, Wall C (1981). The National Study of Breast Cancer Screening Protocol for a Canadian

randomized controlled trial of screening for breast cancer in women. *Clin Invest Med*, 4(3–4):227–58. PMID:6802546

Miller AB, To T, Baines CJ, Wall C (2000). Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years. *J Natl Cancer Inst*, 92(18):1490–9. doi:10.1093/jnci/92.18.1490 PMID:10995804

Miller AB, To T, Baines CJ, Wall C (2002). The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med*, 137(5 Part 1):305–12. doi:10.7326/0003-4819-137-5_Part_1-200209030-00005 PMID:12204013

Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA (2014a). Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. *BMJ*, 348:g366. doi:10.1136/bmj.g366 PMID:24519768

Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA (2014b). Re: Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. http://www.bmj.com/content/348/bmj.g366?tab=responses

Mittra I, Mishra GA, Singh S, Aranke S, Notani P, Badwe R et al. (2010). A cluster randomized, controlled trial of breast and cervix cancer screening in Mumbai, India: methodology and interim results after three rounds of screening. *Int J Cancer*, 126(4):976–84. doi:10.1002/ijc.24840 PMID:19697326

Morrison AS, Brisson J, Khalid N (1988). Breast cancer incidence and mortality in the Breast Cancer Detection Demonstration Project [published erratum appears in *J Natl Cancer Inst* (1989) Oct 4; 81(19):1513]. *J Natl Cancer Inst*, 80(19):1540–7. doi:10.1093/jnci/80.19.1540 PMID:3193469

Moskowitz M (1992). Guidelines for screening for breast cancer. Is a revision in order? *Radiol Clin North Am*, 30(1):221–33. PMID:1732929

Moss S (1999). A trial to study the effect on breast cancer mortality of annual mammographic screening in women starting at age 40. *J Med Screen*, 6(3):144–8. doi:10.1136/jms.6.3.144 PMID:10572845

Moss S (2005). Overdiagnosis and overtreatment of breast cancer: overdiagnosis in randomised controlled trials of breast cancer screening. *Breast Cancer Res*, 7(5):230–4. doi:10.1186/bcr1314 PMID:16168145

Moss S, Thomas I, Evans A, Thomas B, Johns L; Trial Management Group (2005b). Randomised controlled trial of mammographic screening in women from age 40: results of screening in the first 10 years. *Br J Cancer*, 92(5):949–54. doi:10.1038/sj.bjc.6602396 PMID:15726102

Moss S, Waller M, Anderson TJ, Cuckle H; Trial Management Group (2005a). Randomised controlled trial of mammographic screening in women from age 40: predicted mortality based on surrogate outcome measures. *Br J Cancer*, 92(5):955–60. PMID:15726103

Moss SM, Coleman DA, Ellman R, Chamberlain J, Forrest AP, Kirkpatrick AE et al. (1993). Interval cancers and sensitivity in the screening centres of the UK Trial of Early Detection of Breast Cancer. *Eur J Cancer*, 29A(2):255–8. doi:10.1016/0959-8049(93)90187-K PMID:8422291

Moss SM, Cuckle H, Evans A, Johns L, Waller M, Bobrow L; Trial Management Group (2006). Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. *Lancet*, 368(9552):2053–60. doi:10.1016/S0140-6736(06)69834-6 PMID:17161727

Muscat JE, Huncharek MS (1991). Breast self-examination and extent of disease: a population-based study. *Cancer Detect Prev*, 15(2):155–9. PMID:2032258

Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L; U.S. Preventive Services Task Force (2009). Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med*, 151(10):727–37, W237–42. doi:10.7326/0003-4819-151-10-200911170-00009 PMID:19920273

Newcomb PA, Weiss NS, Storer BE, Scholes D, Young BE, Voigt LF (1991). Breast self-examination in relation to the occurrence of advanced breast cancer. *J Natl Cancer Inst*, 83(4):260–5. doi:10.1093/jnci/83.4.260 PMID:1994055

Njor SH, Garne JP, Lynge E (2013). Over-diagnosis estimate from The Independent UK Panel on Breast Cancer Screening is based on unsuitable data. *J Med Screen*, 20(2):104–5. doi:10.1177/0969141313495190 PMID:24065032

Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE (2002a). Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet*, 359(9310):909–19. doi:10.1016/S0140-6736(02)08020-0 PMID:11918907

Nyström L, Andersson I, Bjurstam N, Frisell J, Rutqvist LE (2002b). Update on effects of screening mammography. Authors' reply. *Lancet*, 360(9329):339–40. doi:10.1016/S0140-6736(02)09527-2

Oestreicher N, Lehman CD, Seger DJ, Buist DS, White E (2005). The incremental contribution of clinical breast examination to invasive cancer detection in a mammography screening program. *AJR Am J Roentgenol*, 184(2):428–32. doi:10.2214/ajr.184.2.01840428 PMID:15671358

Olsen O, Gøtzsche PC (2001). Cochrane review on screening for breast cancer with mammography. *Lancet*, 358(9290):1340–2. doi:10.1016/S0140-6736(01)06449-2 PMID:11684218

Paap E, Verbeek A, Puliti D, Broeders M, Paci E (2011). Minor influence of self-selection bias on the effectiveness of breast cancer screening in case-control studies in the Netherlands. *J Med Screen*, 18(3):142–6. doi:10.1258/jms.2011.011027 PMID:22045823

Paap E, Verbeek AL, Botterweck AA, van Doorne-Nagtegaal HJ, Imhof-Tas M, de Koning HJ et al. (2014). Breast cancer screening halves the risk of breast cancer death: a case-referent study. *Breast*, 23(4):439–44. doi:10.1016/j.breast.2014.03.002 PMID:24713277

Pisani P, Parkin DM, Ngelangel C, Esteban D, Gibson L, Munson M et al. (2006). Outcome of screening by clinical examination of the breast in a trial in the Philippines. *Int J Cancer*, 118(1):149–54. doi:10.1002/ijc.21343 PMID:16049976

Porta M, editor (2014). A dictionary of epidemiology, 5th edition. Online version. Oxford University Press. Available from: http://www.oxfordreference.com/view/10.1093/acref/9780195314496.001.0001/acref-9780195314496.

Puliti D, Miccinesi G, Paci E (2011). Overdiagnosis in breast cancer: design and methods of estimation in observational studies. *Prev Med*, 53(3):131–3. doi:10.1016/j.ypmed.2011.05.012 PMID:21658405

Rijnsburger AJ, van Oortmarssen GJ, Boer R, Draisma G, To T, Miller AB et al. (2004). Mammography benefit in the Canadian National Breast Screening Study-2: a model evaluation. *Int J Cancer*, 110(5):756–62. doi:10.1002/ijc.20143 PMID:15146566

Sankaranarayanan R, Ramadas K, Thara S, Muwonge R, Prabhakar J, Augustine P et al. (2011). Clinical breast examination: preliminary results from a cluster randomized controlled trial in India. *J Natl Cancer Inst*, 103(19):1476–80. doi:10.1093/jnci/djr304 PMID:21862730

Semiglazov VF, Manikhas AG, Moiseenko VM, Protsenko SA, Kharikova RS, Seleznev 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

Semiglazov VF, Moiseenko VM, Manikhas AG, Protsenko SA, Kharikova RS, Popova RT et al. (1999a). Interim results of a prospective randomized study of self-examination for early detection of breast cancer (Russia (St Petersburg)/WHO) [in Russian]. *Vopr Onkol*, 45(3):265–71. PMID:10443229

Semiglazov VF, Moiseyenko VM, Manikhas AG, Protsenko SA, Kharikova RS, Ivanov VG et al. (1999b). Role of breast self-examination in early detection of breast cancer: Russia/WHO prospective randomized trial in St. Petersburg. *Cancer Strategy*, 1:145–51.

Shapiro S (1977). Evidence on screening for breast cancer from a randomized trial. *Cancer*, 39(6 Suppl): 2772–82. doi:10.1002/1097-0142(197706)39:6<2772::AID-CNCR2820390665>3.0.CO;2-K PMID:326378

Shapiro S (1994). Screening: assessment of current studies. *Cancer*, 74(Suppl 1):231–8. doi:10.1002/cncr.2820741306 PMID:8004592

Shapiro S (1997). Periodic screening for breast cancer: the HIP randomized controlled trial. Health Insurance Plan. *J Natl Cancer Inst Monogr*, (22):27–30. PMID:9709271

Shapiro S, Strax P, Venet L (1966). Evaluation of periodic breast cancer screening with mammography. Methodology and early observations. *JAMA*, 195(9):731–8. doi:10.1001/jama.1966.03100090065016 PMID:5951878

Shapiro S, Strax P, Venet L (1971). Periodic breast cancer screening in reducing mortality from breast cancer. *JAMA*, 215(11):1777–85. doi:10.1001/jama.1971.03180240027005 PMID:5107709

Shapiro S, Venet W, Strax P, Venet L (1988). Screening for breast cancer: the Health Insurance Plan project and its sequelae, 1963–1986. Baltimore (MD), USA: The Johns Hopkins University Press.

Smith RA (2000). Breast cancer screening among women younger than age 50: a current assessment of the issues. *CA Cancer J Clin*, 50(5):312–36. PMID:11075240

Tabár L (2014). Re: Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. http://www.bmj.com/content/348/bmj.g366?tab=responses

Tabár L, Chen H-HT, Duffy SW, Krusemo UB (1999). Primary and adjuvant therapy, prognostic factors and survival in 1053 breast cancers diagnosed in a trial of mammography screening. *Jpn J Clin Oncol*, 29(12):608–16. doi:10.1093/jjco/29.12.608 PMID:10721943

Tabár L, Chen HH, Yen AM, Chen SL, Fann JC, Chiu SY et al. (2015b). Response to Miller et al. *Breast J*, 21(4):459–61. doi:10.1111/tbj.12439 PMID:26010345

Tabár L, Duffy SW, Yen MF, Warwick J, Vitak B, Chen HH et al. (2002). All-cause mortality among breast cancer patients in a screening trial: support for breast cancer mortality as an end point. *J Med Screen*, 9(4):159–62. doi:10.1136/jms.9.4.159 PMID:12518005

Tabár L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Gröntoft O et al. (1985). Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. *Lancet*, 1(8433):829–32. doi:10.1016/S0140-6736(85)92204-4 PMID:2858707

Tabár L, Fagerberg G, Chen HH, Duffy SW, Smart CR, Gad A et al. (1995). Efficacy of breast cancer screening by age. New results from the Swedish Two-County Trial. *Cancer*, 75(10):2507–17.

doi:10.1002/1097-0142(19950515)75:10<2507::AID-CNCR2820751017>3.0.CO;2-H PMID:7736395

Tabár L, Fagerberg G, Duffy SW, Day NE, Gad A, Gröntoft O (1992). Update of the Swedish two-county program of mammographic screening for breast cancer. *Radiol Clin North Am*, 30(1):187–210. PMID:1732926

Tabár L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF et al. (2000). The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. *Radiol Clin North Am*, 38(4):625–51. doi:10.1016/S0033-8389(05)70191-3 PMID:10943268

Tabár L, Vitak B, Chen TH, Yen AM, Cohen A, Tot T et al. (2011). Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology*, 260(3):658–63. doi:10.1148/radiol.11110469 PMID:21712474

Tabár L, Yen AM-F, Wu WY-Y, Chen SL-S, Chiu SY-H, Fann JC-Y et al. (2015a). Insights from the breast cancer screening trials: how screening affects the natural history of breast cancer and implications for evaluating service screening programs. *Breast J*, 21(1):13–20. doi:10.1111/tbj.12354 PMID:25413699

Thomas DB, Gao DL, Ray RM, Wang WW, Allison CJ, Chen FL et al. (2002). Randomized trial of breast self-examination in Shanghai: final results. *J Natl Cancer Inst*, 94(19):1445–57. doi:10.1093/jnci/94.19.1445 PMID:12359854

Thomas DB, Gao DL, Self SG, Allison CJ, Tao Y, Mahloch J et al. (1997). Randomized trial of breast self-examination in Shanghai: methodology and preliminary results. *J Natl Cancer Inst*, 89(5):355–65. doi:10.1093/jnci/89.5.355 PMID:9060957

Thornton H, Pillarisetti RR (2008). 'Breast awareness' and 'breast self-examination' are not the same. What do these terms mean? Why are they confused? What can we do? *Eur J Cancer*, 44(15):2118–21. doi:10.1016/j.ejca.2008.08.015 PMID:18805689

Tu SP, Reisch LM, Taplin SH, Kreuter W, Elmore JG, Legge Muilenburg J (2006). Breast self-examination: self-reported frequency, quality, and associated outcomes. *J Cancer Educ*, 21(3):175–81. doi:10.1207/s15430154jce2103_18 PMID:17371185

UK Trial of Early Detection of Breast Cancer Group (1988). First results on mortality reduction in the UK Trial of Early Detection of Breast Cancer. *Lancet*, 2(8608):411–6. PMID:2900351

UK Trial of Early Detection of Breast Cancer Group (1993). Breast cancer mortality after 10 years in the UK Trial of Early Detection of Breast Cancer. *Breast*, 2(1):13–20. doi:10.1016/0960-9776(93)90031-A

UK Trial of Early Detection of Breast Cancer Group (1999). 16-year mortality from breast cancer in the UK Trial of Early Detection of Breast Cancer. *Lancet*, 353(9168):1909–14. doi:10.1016/S0140-6736(98)07412-1 PMID:10371568

van Schoor G, Moss SM, Otten JD, Donders R, Paap E, den Heeten GJ et al. (2011a). Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer*, 104(6):910–4. doi:10.1038/bjc.2011.44 PMID:21343930

van Schoor G, Paap E, Broeders MJ, Verbeek AL (2011b). Residual confounding after adjustment for age: a minor issue in breast cancer screening effectiveness. *Eur J Epidemiol*, 26(8):585–8. doi:10.1007/s10654-011-9584-3 PMID:21519892

Weiss NS (2014). All-cause mortality as an outcome in epidemiologic studies: proceed with caution. *Eur J Epidemiol*, 29(3):147–9. PMID:24729152

Wilke LG, Broadwater G, Rabiner S, Owens E, Yoon S, Ghate S et al. (2009). Breast self-examination: defining a cohort still in need. *Am J Surg*, 198(4):575–9. doi:10.1016/j.amjsurg.2009.06.012 PMID:19800471

Yen AM, Duffy SW, Chen TH, Chen LS, Chiu SY, Fann JC et al. (2012). Long-term incidence of breast cancer by trial arm in one county of the Swedish Two-County Trial of mammographic screening. *Cancer*, 118(23):5728–32. doi:10.1002/cncr.27580 PMID:22605639

Yen MF, Tabár L, Vitak B, Smith RA, Chen HH, Duffy SW (2003). Quantifying the potential problem of over-diagnosis of ductal carcinoma in situ in breast cancer screening. *Eur J Cancer*, 39(12):1746–54. doi:10.1016/S0959-8049(03)00260-0 PMID:12888370

Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP (2006). Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ*, 332(7543):689–92. doi:10.1136/bmj.38764.572569.7C PMID:16517548

Zahl P-H, Gøtzsche PC, Andersen JM, Maehlen J (2006). Results of the Two-County trial of mammography screening are not compatible with contemporaneous official Swedish breast cancer statistics. *Dan Med Bull*, 53(4):438–40. PMID:17150148