ORIGINAL ARTICLE

The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies

Mireille Broeders, Sue Moss, Lennarth Nyström, Sisse Njor, Håkan Jonsson, Ellen Paap, Nathalie Massat, Stephen Duffy, Elsebeth Lynge and Eugenio Paci, for the EUROSCREEN Working Group

J Med Screen 2012;**19 Suppl 1**:14–25
DOI: 10.1258/jms.2012.012078

See end of article for authors' affiliations

Correspondence to: Mireille Broeders, Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands; m.broeders@ebh.umcn.nl

Accepted for publication 14 June 2012

Objectives To assess the impact of population-based mammographic screening on breast cancer mortality in Europe, considering different methodologies and limitations of the data.
Methods We conducted a systematic literature review of European trend studies ($n = 17$), incidence-based mortality (IBM) studies ($n = 20$) and case-control (CC) studies ($n = 8$). Estimates of the reduction in breast cancer mortality for women invited versus not invited and/or for women screened versus not screened were obtained. The results of IBM studies and CC studies were each pooled using a random effects meta-analysis.
Results Twelve of the 17 trend studies quantified the impact of population-based screening on breast cancer mortality. The estimated breast cancer mortality reductions ranged from 1% to 9% per year in studies reporting an annual percentage change, and from 28% to 36% in those comparing post- and prescreening periods. In the IBM studies, the pooled mortality reduction was 25% (relative risk [RR] 0.75, 95% confidence interval [CI] 0.69–0.81) among invited women and 38% (RR 0.62, 95% CI 0.56–0.69) among those actually screened. The corresponding pooled estimates from the CC studies were 31% (odds ratio [OR] 0.69, 95% CI 0.57–0.83), and 48% (OR 0.52, 95% CI 0.42–0.65) adjusted for self-selection.
Conclusions Valid observational designs are those where sufficient longitudinal individual data are available, directly linking a woman's screening history to her cause of death. From such studies, the best 'European' estimate of breast cancer mortality reduction is 25–31% for women invited for screening, and 38–48% for women actually screened. Much of the current controversy on breast cancer screening is due to the use of inappropriate methodological approaches that are unable to capture the true effect of mammographic screening.

INTRODUCTION

Mammographic breast cancer screening has been the subject of controversy, despite or perhaps due to the fact that it is one of the most scrutinized public health interventions. Randomized controlled trials (RCTs), conducted in the 1970s and 1980s, have shown that mammographic screening can reduce breast cancer mortality by 25–30% after 7–12 years from entry in the trials.[1] Nevertheless, since 2000, concerns have been raised about the validity of these trials because of supposed 'flaws' in randomization and ascertainment of cause of death,[2,3] although these issues have been addressed.[4,5] More recently, observational studies reporting on the impact of population-based screening programmes have also been questioned.[6,7] The debate that followed, in academic journals as well as the lay press, has not helped women and their physicians to have a clear view of the benefit of mammographic screening.[8] Concern has also been expressed that women are not fully informed about the potential harms of screening, in particular, possible over-diagnosis of cancers that might not have been diagnosed clinically.[9–12]

Many countries implemented population-based screening following the results of the RCTs.[13] There are several reasons why the effectiveness of population-based service screening mammography may differ from that observed in the RCTs, including the wider base of professionals who are involved in screening and the improvement of mammographic and other techniques since the trials were conducted.[14,15] In RCTs and in some observational designs the effect of screening is measured by comparing women invited with women not invited. This comparison is influenced by the attendance rate and therefore reflects the performance of the programme, rather than the screening test itself. The effect estimate will be larger when comparing breast cancer mortality in screened women with that in non-screened women.[16] Service screening effectiveness will also be influenced by the extent of opportunistic screening. Although data on opportunistic screening are scarce, the increased use of mammography outside organized screening programmes may contribute to a reduction in breast cancer mortality.[17]

The emphasis for evaluation has now shifted to population-based screening services, and observational studies will become the main contributors of new information on the impact of breast cancer screening as a public health policy. In this review, we focus on the reduction in breast cancer mortality as the principal benefit of screening, which is by definition a long-term commitment. Several studies corroborate that well-designed observational studies produce results that are similar to those from RCTs.[18] There are, however, specific difficulties in determining the impact of breast cancer screening.

A common first step in the evaluation of screening is to study trends in breast cancer mortality over time. However, the impact of service screening on breast cancer mortality observed in routine population statistics will take many years to emerge.[19] Firstly, with improved treatment, breast cancer survival is generally much higher than in the past while breast cancer incidence has increased in most countries. In combination, the number of deaths in the short-term will be lower, but in the long-term the absolute number of potentially preventable breast cancer deaths has increased. Secondly, it usually takes a number of years before a screening programme is fully implemented. Thirdly, most trend studies are not able to allow for breast cancers diagnosed in women before the start of the screening programme.[20,21] Finally, when there is no individual data, no corrections can be made for the varying participation behaviour of women invited.[22] Potential confounding, where factors other than screening may also contribute to changes in breast cancer mortality, also presents a complication. Therefore, service-based screening programmes cannot be evaluated using only analyses of trends.

A further difficulty in determining the impact of screening is the typical absence of a readily available control population. Studies which were able to identify, albeit for a limited time period, a group of contemporaneous controls that were not (yet) invited for screening have mostly used the incidence-based mortality (IBM) approach. IBM studies estimate the impact of screening by calculating mortality rates based on breast cancer deaths occurring in women with breast cancer diagnosed after their first invitation to screening.[23] Using individual data in IBM studies can overcome many of the problems that affect trend analyses.

Case-control (CC), or case-referent, studies have also been used to evaluate the impact of service screening.[24–29] A CC study compares breast cancer deaths (cases) with a sample of women who have not died from breast cancer, in terms of individual screening exposure. There is an efficiency gain in taking a sample of the population invited to be screened, rather than observing the entire population.[30] If correctly designed and analysed, the CC approach offers a valid and efficient method for estimating the impact of service screening programmes.[25]

Our objective is to assess the impact of population-based screening with mammography on breast cancer mortality in Europe. A best estimate for the effectiveness of population-based screening in Europe will be provided, acknowledging the different methodologies and the limitations of the available data.

METHODS

A systematic search of PubMed was performed based on all papers published up to February 2011 (details in the Appendix A). We identified 5011 English-language articles evaluating the effect of mammographic screening on breast cancer mortality in Europe. After inspection of titles and abstracts, 122 studies were considered to be relevant. These were reviewed and further selected using the following criteria: (a) the study represents original data on a population-based screening programme in Europe, (b) breast cancer mortality is reported, (c) the analysis includes at least some of the age groups between 50 and 69, and (d) one of the following observational research designs was used: trend, IBM or CC study. In addition, we only considered studies estimating the impact of current breast cancer screening programmes, and therefore excluded those which had less than three years' overlap with the relevant current regional or national population screening programme. Based on these criteria, 83 studies were excluded on the following grounds: data from RCTs ($n = 17$), outcome measure is not breast cancer mortality ($n = 20$), insufficient overlap with current population-based programme ($n = 11$), data limited to younger or older women ($n = 9$), study reporting no new data or no analysis with regard to screening ($n = 15$), modelling study ($n = 6$), full paper not in English ($n = 2$), study on opportunistic screening ($n = 2$) and study on benign breast disease ($n = 1$).

In addition to the literature search, the Working Group added publications fulfilling the inclusion criteria but not identified by the search and new publications that became available after February 2011 ($n = 5$). Studies were summarized according to the three designs (see Table 1): trend studies[21,31,41,44,46,48,52–56,64,67–71] IBM studies[22,32–39,42,45,51,53,57–59,60–63] and CC studies.[15,40,43,47,49,50,65,66]

Trend studies

Relevant papers were those that reported on trends in breast cancer mortality rates in a population as a whole in relation to the introduction and/or extent of population based mammographic screening ($n = 17$). They are described in detail elsewhere in this supplement. These studies were usually based on aggregated data obtained from routine sources, such as cancer registries. Trend studies were either classified into (a) descriptions of the trend over time in breast cancer mortality in relation to the timing of the introduction of population-based screening ($n = 5$), or (b) those which included a more detailed analysis with the aim of quantifying the impact of screening on mortality ($n = 12$). Methods of analysis in the latter category included Poisson regression (with or without age cohort modelling), and the use of joinpoint regression to identify 'break points' at which changes in mortality trends occurred (see Table 2). Due to the varied methodology and comparisons in the studies, no attempt was made to produce a pooled estimate of the effect of screening.

IBM studies

In an IBM study all breast cancer deaths occurring in a dynamic or cohort population over a period of time are

Table 1 Publications on the impact of population-based screening with mammography in Europe according to observational study design

Country	Start of national programme	Age group targeted (years)	Study design (first author, journal and year of publication): Trend studies ($n = 17$)	IBM studies ($n = 20$)	Case-control studies ($n = 8$)
Denmark	2007	50–69	Jørgensen, *BMJ* (2010)[31]	Olsen, *BMJ* (2005)[32]	
Finland	1987	50–63		Wu, *Breast Cancer Res Treat* (2010)[33] Anttila, *BMC Public Health* (2008)[34] Sarkeala, *Int J Cancer* (2008)[35] Sarkeala, *Br J Cancer* (2008)[36] Parvinen, *J Med Screen* (2006)[37] Anttila, *J Med Screen* (2002)[38] Hakama, *BMJ* (1997)[39]	
Iceland	1987	40–69			Gabe, *Acta Radiol* (2007)[40]
Italy	1995	50–69	Gorini, *Br J Cancer* (2004)[41] Barchielli, *Cancer Causes Control* (2001)[44]	Paci, *Eur J Cancer* (2002)[42] Paci, *Br J Cancer* (2002)[45]	Puliti, *Br J Cancer* (2008)[43]
The Netherlands	1989	50–74	Otten, *Int J Cancer* (2008)[46] Otto, *Lancet* (2003)[48]		Otto, *Cancer Epidemiol Biomarkers Prev* (2012)[47] van Schoor, *Br J Cancer* (2011)[15] Paap, *Cancer Causes Control* (2010)[49] Broeders, *J Med Screen* (2002)[50]
Norway	1996	50–69		Kalager, *N Engl J Med* (2010)[51]	
Spain	1990	45–69	Ugarte, *Ann Epidemiol* (2010)[52] Cabanes, *Cancer Epidemiol* (2009)[54] Pons-Vigués, *Cancer Detect Prev* (2008)[55] Ascunce, *J Med Screen* (2007)[53]	Ascunce, *J Med Screen* (2007)[53]	
Sweden	1986	40–74*	Haukka, *PLoS One* (2011)[56]	Chen, *Ann Epidemiol* (2010)[57] Jonsson, *J Med Screen* (2007)[58] *SOSSEG, Cancer Epidemiol Biomarkers Prev* (2006)[59] Baker, *BMC Med Res Methodol* (2004)[60] Jonsson, *Breast* (2003)[61] Duffy, *Cancer* (2002)[62] Tabar, *Cancer* (2001)[63] Jonsson, *J Med Screen* (2001)[22]	
United Kingdom	1988	50–64†	Duffy, *J Med Screen* (2010)[64] Blanks, *BMJ* (2000)[21] Quinn, *BMJ* (1995)[67]		Allgood, *Br J Cancer* (2008)[65] Fielder, *J Med Screen* (2004)[66]
16 European countries			Botha, *Eur J Cancer* (2003)[68]		
Nordic capitals			Törnberg, *Acta Oncol* (2006)[69]		
30 European countries			Autier, *BMJ* (2010)[70]		
Northern Ireland, the Netherlands, Sweden‡			Autier, *BMJ* (2011)[71]		

*In half of the Swedish counties, the lower age limit is 40; in the other half screening starts at age 50
†Current age limits are 50–70 but will be extended to 47–73
‡Northern Ireland (UK) compared with the republic of Ireland, the Netherlands compared with Belgium/Flanders and Sweden compared with Norway

Table 2 Summary of European trend studies that report an estimate of the effect of screening

Reference	Study area	Service screening programme: Start	Service screening programme: 100% coverage	Service screening programme: Age group invited	Time period studied	Age range studied	Method	Reference group	Reduction in breast cancer mortality
Denmark									
Jørgensen (2010)[31]	Copenhagen and Funen	1991, 1993		50–69	1971–2006	35–84	Poisson	Rest of Denmark 55–74 versus other age groups	1977–2006: 1% (95% CI −1 to 4) reduction per annum versus 2% (95% CI 1–3) in non-screening areas
Italy									
Gorini (2004)[41]	Florence	1970 (ES), 1990 (LS)		(40)50–69, 50–69	1985–2000	35+	Log linear regression	Early (ES) versus late starting (LS) areas	1985–2000: 30% reduction in age standardized rate; 41% (95% CI 21–56) in ES versus 11% (95% CI 0–21) in LS
Barchielli (2001)[44]	Florence	(1970), 1990	1995 (72%)	50–69	1970–1997	25+	Poisson	Remainder of Tuscany	Similar reduction in both areas
The Netherlands									
Otten (2008)[46]	The Netherlands	1989	1997	50–69	1975–2006	35–85	Joinpoint	Before/after in age group 55–74	1994–2006: 2.3% (95% CI 1.6–3.0) −2.8% (95% CI: 2.2–3.4) reduction
Otto (2003)[48]	The Netherlands	1989	1997	50–69	1980–2001	55–74	Poisson	Before/after; clustered by start date.	1989–2001: 1.7% (95% CI 1.0–2.4) reduction pa after start of screening
Spain									
Cabanes (2009)[54]	Spain	1990–99	2001	(45)50–64	1980–2006	25+	Joinpoint	Age group	Changepoint 1993. Reduction pa (95% CI) 1993–2006 by age group: 4% (95% CI: 3.5–4.4) pa 25–44 years; 3.1% (95% CI: 2.9–3.4) 45–64 years, 1.3% 65+ years
Pons-Vigués (2008)[55]	Barcelona Spain	1995	2004	50–69	1984–2004	50–74	Poisson	Before/after; grouped by start date.	1995–2004: 5% (95% CI 1–8) reduction pa versus 1% (95% CI 1–2%) before start
Ascunce (2007)[53]	Navarre, Spain	1990		45–65	1975–2004	30+	Joinpoint	Before/after: 50–69 versus other age groups	1987/9–2002/4: 36% reduction (95% CI 21–48) all ages. Changepoint 1995. 1995–2004: 9% reduction pa 50–69 years, no significant trend in other age groups
Sweden									
Haukka (2011)[56]	Sweden, 9 counties	1986–1990	1996	40–69	1974–2003	40(59)–69(74)	Poisson	Screening effect allowing for lead time and secular trend. Grouped by start date.	16% reduction (40–69) (95% CI 9–22). 11% reduction (70–79) (95% CI 2–20). After start of screening
United Kingdom									
Duffy (2010)[64]	England	1988	1993	50–64(69)	1974–2000	All ages	Poisson	Before/after 50–69 versus other age groups	1995–2004 versus 1974–1988: 28% reduction (95% CI 26–30)
Blanks (2000)[21]	England and Wales	1988	1993	50–64	1969–1998	40–79	Poisson/APC	Observed versus expected in 55–69 versus other age groups	Reduction in 1998 of 6.4% (21.3% ages 55–69 versus 14.9% other age groups). Range 5.4–11.8%
Multicountry									
Autier (2011)[71]	Northern Ireland, Sweden, the Netherlands*	1990, 1986, 1989	1993, 1996, 1997	50–64 (40)–69(74), 50–69(74)	1980–2006, 1980–2006, 1980–2006	40–79	Joinpoint	50–69y, Republic of Ireland, Norway, Belgium	1989–2006: % change age-standardized rate, 36.7 versus 27.7, 28.0 versus 21.4, 16.0 versus 22.8

CI, confidence interval

*Northern Ireland (UK) compared with the republic of Ireland, Sweden compared with Norway and the Netherlands compared with Belgium/Flanders enrolled in the study only if the breast cancer diagnosis occurred in a certain time/age window (taking into account eligibility and opportunity to be screened) and the population is classified by screening or by invitation to screening. Thus, for example, breast cancer deaths in the 15 years after screening is initiated in one region, *from tumours diagnosed in that 15-year period*, may be compared with the corresponding deaths from tumours diagnosed in the same period in a region without screening. The selection of IBM studies contributing to this overall review is described in detail elsewhere in this supplement.[23] There were 20 IBM studies – one each from Denmark, Norway and Spain, two from Italy, seven from Finland and eight from Sweden. A key issue in these studies is how the breast cancer mortality expected in the absence of screening is estimated. Another methodological concern is how the study deals with potential biases in the estimated mortality reduction due to screening. Because breast cancer cases are diagnosed earlier in screened women than in those who are not screened, a longer follow-up period for breast cancer deaths than the accrual period for cases will confer an artificial increase in mortality in the screening period due to fatal cases whose diagnosis is moved to the accrual period due to lead time. The same consideration applies to age at diagnosis. If mortality includes deaths from tumours diagnosed within a certain age range, but with no upper limit on age at death, there will be a number of fatal cancers diagnosed by screening within the age range, which would otherwise have been excluded as diagnosed symptomatically above the age range.[23]

Table 3 presents some basic characteristics of the IBM studies. Where there was overlapping data, the study used in this review was selected on the basis of follow-up time, judgement of quality of the comparison group and study size. We calculated a pooled estimate of the effect on breast cancer mortality in women invited versus not invited, as well as a pooled estimate for women screened versus not screened, using the formula described by Duffy *et al.*[72] The effect sizes were pooled using the inverse variance method (random effects model) and heterogeneity between the studies was assessed.[14,73]

CC studies

A CC study is embedded in a cohort or a dynamic population and based on sampling of the population experience. Breast cancer deaths (cases) in the population are collected over the period of interest and controls who have not died of breast cancer are selected from the same population, often closely matched by temporal factors. Breast cancer cases and control subjects are then compared with respect to screening history before the date of diagnosis of the breast cancer case. The eight CC studies used in this review (Table 4) came from a recently published methodological overview, but we excluded non-European studies[26] and added publications by Broeders *et al.*,[50] van Schoor *et al.*[15] and Otto *et al.*[47]

The results were pooled to obtain estimates of the effect on breast cancer mortality for women screened versus not screened, based on the crude odds ratios (ORs) as well as ORs adjusted for self-selection. In addition, intention to treat estimates were calculated, using the formula described by Duffy *et al.*,[72] in order to compare the women invited with those not invited. Because the studies by Broeders *et al.* and van Schoor *et al.* were both conducted in Nijmegen, with overlap in the included cases, the former was excluded from the meta-analysis. The effect sizes were pooled as above.[14,73]

Breast cancer mortality as an outcome measure

Breast cancer mortality is the most appropriate primary endpoint for evaluating screening, although its use has been questioned.[74,75] An outcome parameter which avoids problems with cause of death classification is (refined) excess mortality from breast cancer, which includes all mortality associated with breast cancer, even indirectly caused deaths, such as treatment-induced mortality, or deaths caused by the stress imposed by the cancer.[76] However, this method, so far, has only been used in Sweden.

Potential limitations of using breast cancer mortality as an outcome measure are that there could be an increase in deaths attributed to breast cancer because more breast cancer cases are diagnosed in screened women, and the misclassification of breast cancer as the underlying cause of death because the treating physician is influenced by the screening history of the patient. Screening may also affect mortality from other causes, for example, due to complications arising from procedures triggered by screening.[75] However, several studies explicitly assessed the quality of cause-of-death determination in relation to mammographic screening and found no significant evidence of bias.[77–80]

RESULTS

Trend studies

Of the 12 trend studies, three used joinpoint regression, and nine Poisson regression (Table 2). Five papers were based on all of an individual country (England, the Netherlands and Spain), two studied the programme in the city of Florence (Italy), two studied different regions in Spain and one studied two regions of Denmark. One paper included Northern Ireland, the Netherlands and Sweden in comparison with the Republic of Ireland, Belgium/Flanders and Norway, respectively. The most recent paper studied nine counties in Sweden.

Authors of several studies estimated the annual percentage change in mortality, while others presented a comparison between two distinct time periods. Of the former, estimates ranged from reductions of 1% to 9% per year; for those studies with adequate follow-up (at least 10 years from the date of full coverage by invitation) the estimates were 1%, 2.3–2.8% and 9%.[31,46,48,52–55] Of the three studies comparing time periods within a single country, all had adequate follow-up, and the estimates of mortality reduction compared with a prescreening period ranged from 28% to 36%.[41,53,64]

IBM studies

Table 3 shows the design characteristics of the IBM studies. The outcomes were generally compatible when differences

Table 3 Design characteristics of European IBM studies, excluding those with overlapping data, and estimate of effect

Reference	Region and age group	Maximum follow-up (years)	Accrual period = follow-up	Age at diagnosis = age at breast cancer death	Contemporaneous uninvited comparison group	Balanced follow-up Comparison	Lead time adjustment	Relative risk, invited to screening (95% CI)
Denmark								
Olsen (2005)[32]	Copenhagen, 50–69	11	Yes	Yes	Yes	Yes	NR	0.75 (0.63–0.89)
Finland								
Sarkeala, *Br J Cancer* (2008)[36]	8 municipalities, 50–69	12	Yes	Yes	Yes	Yes	NR	0.72 (0.51–0.97)
Hakama (1997)[39]	2/3 of municipalities, 50–64	6	Yes	Yes	Yes	Yes	NR	0.76 (0.53–1.09)
Italy								
Paci, *Eur J Cancer* (2002)[42]	Florence, 50–69	10	No	Yes	Yes	No	NA	0.81 (0.64–1.01)
Norway								
Kalager (2010)[51]	4 counties, 50–69	10	Yes	No	Yes	No	NA	0.88 (0.73–1.05)
Spain								
Ascunce (2007)[53]	Navarra, 45–65	11	Yes	Yes	No	No	NA	0.58 (0.44–0.75)
Sweden								
SOSSEG (2006)[59]	13 areas, 50–69	22	Yes	Yes	No	Yes	No	0.73 (0.69–0.77)

NR, not required; NA, not adjusted; IBM, incidence-based mortality; CI, confidence interval

Table 4 Design characteristics of European case-control studies, and estimate of effect

Study	Geographical area	Study population Cases/controls	Age group	Time period for including cases (breast cancer deaths)	Screening exposure	Crude odds ratio (95% CI)	Adjusted estimates (95% CI): Factor Duffy 1.36	Adjusted estimates (95% CI): Using own correction factor	Adjusted estimates (95% CI): SES	Participation rate*	Intention to treat estimate (95% CI)†
Iceland											
Gabe (2007)[40]	Iceland	226/902	≥40	1990–2002	Ever/never	0.59 (0.41–0.84)		0.65 (0.39–1.09) (factor = 1.17, 95% CI 1.08–1.26) → also adjusted for screening-opportunity bias		62%	0.87 (0.72–1.06)
Italy											
Puliti (2008)[43]	Five regions	657/2058	50–74	1988–2002	Ever/never	0.46 (0.38–0.56)		0.55 (0.36–0.85), (factor = 1.11, 95% CI 0.87–1.40)		65%	0.72 (0.56–0.93)
The Netherlands											
Otto (2012)[47]	Southwest region	755/3739	49–75	1990–2003	Ever/never in index invitation and/or preceding 4 years (maximum 3 screens)	0.45 (0.37–0.54)		0.51 (0.40–0.66), (factor = 1.11, 95% CI 0.99–1.25)		75%	0.65 (0.56–0.77)
Van Schoor (2011)[15]	Nijmegen	52/260	50–69	1992–2008	Ever/never in index invitation‡ + previous invitation (4 years, maximum 2 screens)	0.35 (0.19–0.64)		0.28 (0.12–0.60), (factor = 0.84, 95% CI 0.58–1.21; from Paap *et al.*)[49]		68%	0.47 (0.30–0.74)
Paap (2010)[49]	Mid-South Limburg	118/118	50–75	2004–2005	Index invitation	0.30 (0.14–0.63)		0.24 (0.10–0.58), (factor = 0.84, 95% CI 0.58–1.21)		82%	0.36 (0.20–0.64)
Broeders (2002)[50]	Nijmegen	26/192, 36/231, 48/194, 34/118, 12/39	40–49, 50–59, 60–69, 70–79, >79	1987–1997	Ever/never in index invitation + previous invitation (4 years, maximum 2 screens)	0.90 (0.38–2.14), 0.71 (0.35–1.46), 0.80 (0.42–1.54), 1.13 (0.50–2.58), 2.92 (0.55–15.40)				59%	0.85 (0.63–1.15)††
United Kingdom											
Allgood (2008)[65]	East Anglia	284/568	50–70	1995–2004	Ever/never	0.35 (0.24–0.50)	0.52 (0.32–0.84)		0.35 (0.23–0.51)	80%	0.65 (0.48–0.88)
Fielder (2004)[66]	Wales	419/717	50–75	1998–2001	Ever/never	0.62 (0.47–0.82)	0.75 (0.49–1.14)**		No difference	77%	0.96 (0.73–1.27)

*Based on original publication, except for van Schoor (personal communication) and Broeders (personal communication)
†Calculated using the formula by Duffy *et al.* (*Appl Stat* 2002) and using the crude OR
‡Index invitation is the most recent invitation before diagnosis of the breast cancer case
§And in NETB Report XII
**Limited to cancers diagnosed in 1995–2001, where the crude OR was 0.49 (0.36–0.66)
††Based on an overall crude OR of 0.64 (0.44–0.92) (Broeders, personal communication) and self-selection factor of 1.08, 95% CI 0.85–1.37 (Paap *et al.*[87])

in methodology and local circumstances were taken into account. Details are given elsewhere in this supplement.[23] Those with the strongest designs had (a) expected breast cancer mortality estimated from a cohort of women not yet invited[39] or from historical and contemporaneous control groups;[32,36] and (b) an accrual period equal to the follow-up period for breast cancer deaths.[23] Using all IBM studies, excluding overlapping data-sets, produced a pooled relative risk (RR) estimate of 0.75 (95% confidence interval [CI] 0.69–0.81) for invitation to screening, with no significant heterogeneity ($P = 0.23$). The combined RR for women actually screened was 0.62 (95% CI 0.56–0.69), again with no significant heterogeneity ($P = 0.40$). Figure 1 shows the forest plots.

CC studies

Of the eight CC studies included, one came from Iceland, one from Italy, four from the Netherlands and two from the UK (Table 4), but their designs were very similar.[26] The definition of exposure to screening was based on a comparison of women 'ever' screened versus women 'never' screened in four studies. All Dutch studies adopted the concept of the index invitation, defined as the invitation date closest to the date of diagnosis of the case. The comparison in these studies was between women screened in an exposure period which varied from one to three screening examinations versus women not screened in this period. All studies reported ORs adjusted for self-selection bias, either using the correction factor estimated by Duffy *et al.*[72] or their own correction factor, all closer to 1 than the Duffy factor. Based on the results in the original publications, we also calculated the reduction in breast cancer mortality for women invited versus not invited.[72]

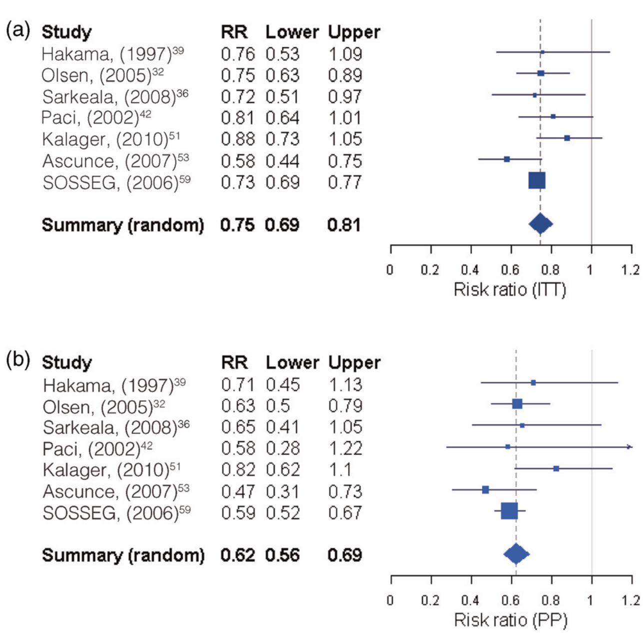

Figure 1 Incidence-based mortality studies excluding overlapping data: (a) estimates for breast cancer mortality reduction in women invited versus not invited; (b) estimates for breast cancer mortality reduction in women screened versus not screened. ITT = intention to treat; PP = per protocol

Seven CC studies were included in a pooled analysis (see Methods). The combined unadjusted OR was 0.46 (95% CI 0.40–0.54), a significant 54% reduction in breast cancer mortality for screened versus not screened women. This became a 48% reduction after adjusting for self-selection (OR 0.52, 95% CI 0.42–0.65). There was no evidence of heterogeneity in either analysis ($P = 0.10$ and 0.17, respectively). The combined mortality reduction for invitation to screening was 31% (OR 0.69, 95% CI 0.57–0.83), but with significant heterogeneity ($P = 0.005$). Figure 2 shows the forest plots. The squares representing the point estimates in the individual CC studies are proportional to the precisions of the log ORs. The order of these may vary when adjusted for self-selection bias as after adjustment the precision also depends on the standard error of the self-selection correction. This in turn depends on the participation rate in each study.

DISCUSSION

Our overview indicates that the estimates from observational studies, using different study designs, are consistent with a

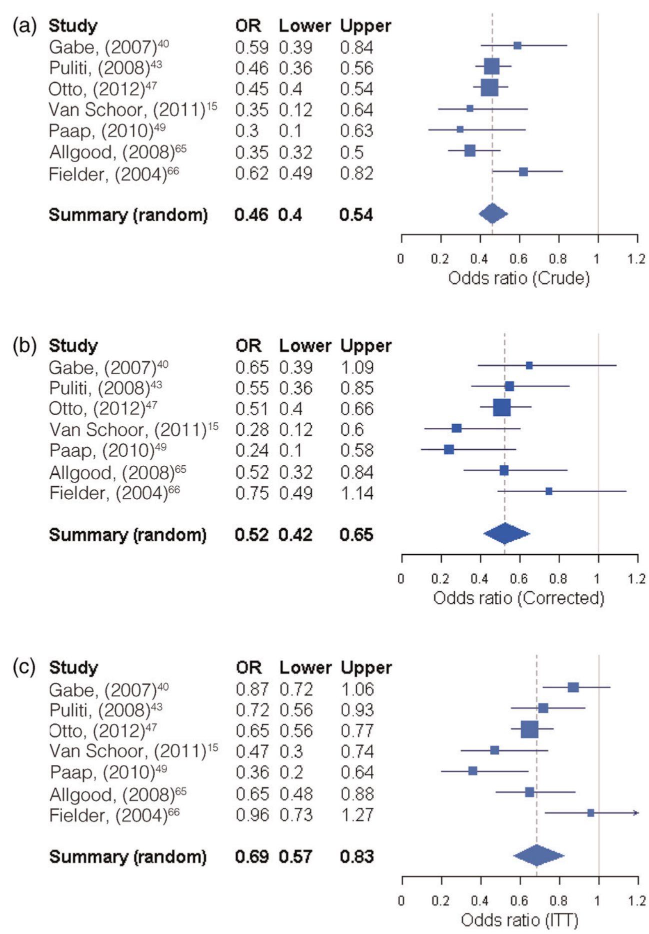

Figure 2 Case-control studies excluding overlapping data: (a) crude odds ratios for breast cancer mortality reduction in women screened versus not screened; (b) crude odds ratios for breast cancer mortality reduction, corrected for self-selection, in women screened versus not screened; (c) crude odds ratios for breast cancer mortality reduction translated to intention to treat estimates for women invited versus not invited

breast cancer mortality reduction of 25–31% for women in Europe invited for population-based screening. The current best estimate of the effectiveness of European screening programmes is therefore at least as large as that observed in the long-term follow-up of the Swedish RCTs[81] or more recent meta-analyses.[74,82]

Given the methodological limitations inherent in observational studies, and the differences in designs, the similarity in the effect estimates from trend, IBM and CC studies is noteworthy. Using all IBM studies without overlapping data, the reduction in breast cancer mortality for women invited was 25%. The corresponding intention to treat estimate in the CC studies was 31%. The relative reduction in breast cancer mortality for women who actually participated in screening was 38% based on IBM studies and 48% based on CC studies. Of the three trend studies comparing time periods within a single country, all had adequate follow-up, and the estimates of mortality reduction compared with a prescreening period ranged from 28% to 36%.

The choice of IBM studies to include in the case of overlapping data was not crucial to the estimated mortality reduction, because pooling all studies, including those with overlapping data, gave a mortality reduction of 24%, and selection of three studies on the basis of both historical and contemporaneous comparison groups gave a reduction of 26%.[23] The heterogeneity among studies of the intention to treat estimate from the CC studies is likely to be due to differing uptake rates between studies, because there was no significant heterogeneity when the effect of actually being screened was assessed.

The study and analysis of population breast cancer mortality rates can be a first step in evaluating the impact of screening on mortality. However, such analyses should be restricted to the age ranges likely to demonstrate a benefit from screening; they should attempt to exclude time periods where dilution due to deaths in women diagnosed preinvitation will be evident; and they should attempt to take account of past underlying trends. We do not support the recommendation of Harris *et al.*[83] to focus on a trend or ecological approach.

The most valid observational designs are those where longitudinal individual data are available, directly linking screening history to the cause of death, achieved using either an IBM or a CC approach. IBM studies and CC studies have one major feature in common – they typically take as clinical endpoint deaths from cancers which have been diagnosed in the age range and time period in which screening is offered. This avoids dilution bias associated with deaths from breast cancers in a given period from tumours diagnosed before that period began.[62] The most obvious difference between the two is that the CC study is retrospective and the IBM study prospective.

In the CC study, data on deaths from the cancer in question are collected along with that from subjects who have not died of the disease, and screening histories retrieved retrospectively. There are a number of well-known potential biases associated with this design, some conservative and some anticonservative.[24,84] However, these can be minimized by appropriate design or corrected for in the statistical analysis.[25,85] Some biases, such as residual confounding after adjusting for age, tend to be very small.[86]

Typically in the IBM studies, rates of death from cancers diagnosed in a population and period of invitation to screening are compared with the corresponding rates in a population or period without such invitation.[59] This too has potential biases. There is likely to be confounding of some variables between populations and periods if individual data on invitation and screening are not available. For example, if a before-after comparison of IBM is carried out, the time cut-off will inevitably incur some misclassification of exposure to invitation, because screening is usually phased in over a period of years.[86] In the CC approach, individual screening histories are retrieved so there is no misclassification of exposure.[25]

In principle, screening exposure can be ascertained for all subjects in the population in the IBM approach, but this involves retrieval of data on tens or even hundreds of thousands of subjects, whereas the CC design typically involves much smaller numbers.[15] Therefore, the CC approach is a more economic research strategy, even though it may involve more complex design or analytic procedures. However, if exposure to screening is ascertained for all study subjects on an individual basis in both study designs, the intention-to-treat estimate from CC studies should be similar to that from the IBM studies, as indeed is observed in this review.

CONCLUSION

After considering all published data from European studies, the reduction in breast cancer mortality associated with mammographic population-based service screening programmes is in the range of 25–31% for women invited for screening and 38–48% for women actually screened with sufficient follow-up time. It appears that much of the current controversy surrounding the value of mammography screening is due to the use of inappropriate methodological approaches that are unable to capture the true effect of mammographic screening.

EUROSCREEN WORKING GROUP

Coordinators:
E Paci, M Broeders, S Hofvind, S W Duffy

Members:
Ancelle-Park R (F) [1], Armaroli P (I)[2], Ascunce N (E)[3], Bisanti, L (I)[4], Bellisario C (I)[2], Broeders M (NL)[5], Cogo C (I)[6], De Koning H (NL)[7], Duffy S W (UK) [8], Frigerio A (I)[2], Giordano L (I) [2], Hofvind S (N) [9], Jonsson H (S)[10], Lynge E (DK)[11], Massat N (UK)[8], Miccinesi G (I)[12], Moss S (UK)[8], Naldoni C (I)[13], Njor S (DK)[11], Nyström L (S) [14], Paap E (NL)[5], Paci E (I)[12], Patnick J (UK) [15], Ponti A (I)[2], Puliti D (I)[12], Segnan N (I)[2], Von Karsa L (D)[16], Törnberg S (S)[17], Zappa M (I)[12], Zorzi M (I)[6]

Affiliations:
[1]Ministère du travail de l'emploi et de la santé, Paris, France
[2]CPO-Piedmont, Turin, Italy
[3]Navarra Breast Cancer Screening Programme. Pamplona, Spain
[4]S.C. Epidemiologia, ASL di Milano, Italy

[5]Radboud University Nijmegen Medical Centre & National Expert and Training Centre for Breast Cancer Screening, Nijmegen, The Netherlands
[6]Veneto Tumor Registry, Padua, Italy
[7]Erasmus MC, Dept. of Public Health, Rotterdam, The Netherlands
[8]Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
[9]Cancer Registry of Norway, Research Department and Oslo and Akershus University College of Applied Science, Oslo, Norway
[10]Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden
[11]Centre for Epidemiology and Screening, Universtiy of Copenhagen, Copenhagen, Denmark
[12]ISPO Cancer Research and Prevention Institute, Florence, Italy
[13]Regional Cancer Screening Center, Emilia-Romagna Region, Bologna, Italy
[14]Department of Public Health and Clinical Medicine, Division of Epidemiology and Global Health, Umeå University, Umeå, Sweden
[15]NHS Cancer Screening Programmes and Oxford University, UK
[16]International Agency for Research on Cancer, Lyon, France
[17]Stockholm Cancer Screening, Stockholm, Sweden

Authors' affiliations
Mireille Broeders, Senior Epidemiologist, Department of Epidemiology, Biostatistics and HTA, Radboud University Nijmegen Medical Centre & National Expert and Training Centre for Breast Cancer Screening, Nijmegen, The Netherlands
Sue Moss, Professor of Cancer Epidemiology, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
Lennarth Nyström, Associate Professor of Epidemiology, Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden
Sisse Njor, Post Doc, Centre for Epidemiology and Screening, University of Copenhagen, Denmark
Håkan Jonsson, Associate Professor of Cancer Epidemiology, Department of Radiation Sciences, Umeå University, Sweden
Ellen Paap, Epidemiologist, Department of Epidemiology, Biostatistics and HTA, Radboud, University Nijmegen Medical Centre & National Expert and Training Centre for Breast Cancer Screening, Nijmegen, The Netherlands
Nathalie Massat, Senior Statistician, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
Stephen Duffy, Professor of Cancer Screening, Centre for Cancer Prevention, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
Elsebeth Lynge, Professor of Epidemiology, Department of Public Health, University of Copenhagen, Copenhagen, Denmark
Eugenio Paci, MD MPH, Director, Clinical and Descriptive Epidemiology Unit, ISPO Cancer Research and Prevention Institute, Florence, Italy
The EUROSCREEN Working Group

ACKNOWLEDGEMENTS

Financial support was provided by the National Monitoring Italian Centre (ONS) to host the EUROSCREEN meetings in Florence in November 2010 and in March 2011 and the supplement publication and the National Expert and Training Centre for Breast Cancer Screening, Nijmegen, The Netherlands to host a meeting of the EUROSCREEN mortality working group in July 2011.

REFERENCES

1 Shapiro S. Screening: Assessment of current studies. *Cancer* 1994;**74**(Suppl):231–8
2 Gotzsche P, Olsen O. Is screening for breast cancer with mammography justifiable? *Lancet* 2000;**355**:129–34
3 Olsen O, Gøtzsche PC. Cochrane review on screening for breast cancer with mammography. *Lancet* 2001;**358**:1340–2
4 Freedman DA, Pettiti DB, Robins J. On the efficacy of screening for breast cancer. *Int J Epidemiol* 2004;**33**:43–55
5 Duffy SW, Tabár L, Viták B, *et al*. The Swedish Two-County Trial of mammographic screening: Cluster randomisation and end point evaluation. *Ann Oncol* 2003;**14**:1196–8
6 Jørgensen KJ, Keen JD, Gøtzsche PC. Is mammographic screening justifiable considering its substantial overdiagnosis and minor effect on mortality? *Radiology* 2011;**260**:621–7
7 Corder AP. Screening for breast cancer. 'Mega-trial' needed. *BMJ* 2010;**341**:c4453
8 Gelmon KA, Olivotto I. The mammography screening debate: Time to move on [Commentary]. *Lancet* 2002;**359**:904–5
9 Gøtzsche PC, Jørgensen KJ. The breast screening programme and misinforming the public. *J R Soc Med* 2011;**104**:361–9
10 Fletcher SW, Elmore JG. Mammographic screening for breast cancer. *N Engl J Med* 2003;**348**:1672–80
11 Baines CJ. Mammography screening: Are women really giving informed consent? *J Natl Cancer Inst* 2003;**95**:1508–11
12 Thornton H, Edwards A, Baum M. Women need better information about routine mammography. *BMJ* 2003;**327**:101–3
13 von Karsa L, Anttila A, Ronco G, *et al*. Cancer Screening in the European Union. Report on the Implementation of the Council Recommendation on Cancer Screening – First report. European Commission, 2008
14 Gabe R, Duffy SW. Evaluation of service screening mammography in practice: The impact on breast cancer mortality. *Ann Oncol* 2005;**16**(Suppl 2):ii153–562
15 van Schoor G, Moss SM, Otten JDM, *et al*. Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer* 2011;**104**:910–4
16 Schopper D, de Wolf C. How effective are breast cancer screening programmes by mammography? Review of the current evidence. *Eur J Cancer* 2009;**45**:1916–23
17 Lynge E, Braaten T, Njor SH, *et al*. Mammographic activity in Norway 1983 to 2008. *Acta Oncol* 2011;**50**:1062–7
18 Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies and the hierarchy of research designs. *N Engl J Med* 2000;**342**:1887–92
19 IARC Working Group on the Evaluation of Cancer Preventive Strategies. Breast cancer screening. In: Vainio H, Bianchini F, eds. IARC handbooks of cancer prevention. Vol 7. Lyon: IARC Press, 2002
20 Jatoi I, Miller AB. Why is breast-cancer mortality declining? *Lancet Oncol* 2003;**4**:251–4
21 Blanks RG, Moss SM, McGahan CE, Quinn MJ, Babb PJ. Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990–8: Comparison of observed with predicted mortality. *BMJ* 2000;**321**:665–9
22 Jonsson H, Nystrom L, Törnberg S, Lenner P. Service screening with mammography of women aged 50–69 years in Sweden: Effects on mortality from breast cancer. *J Med Screen* 2001;**8**:152–160
23 Njor S, Nyström L, Moss S, *et al*. Breast cancer mortality after start of service mammography screening in Europe. A review of incidence based mortality studies. *J Med Screen* 2012; supplement
24 Walter SD. Mammographic screening: Case-control studies. *Ann Oncol* 2003;**14**:1190–2
25 Verbeek AL, Broeders MJ. Evaluation of cancer service screening: Case referent studies recommended. *Stat Methods Med Res* 2010;**19**:487–505
26 Paap E, Verbeek ALM, Puliti D, Paci E, Broeders MJM. Breast cancer screening case-control study design: Impact on breast cancer mortality. *Ann Oncol* 2011;**22**:863–9
27 Verbeek ALM, Hendriks JHCL, Holland R, Mravunac M, Sturmans F, Day NE. Reduction of breast cancer mortality through mass screening with modern mammography; first results of the Nijmegen Project, 1975–1981. *Lancet* 1984;**I**:1222–4
28 Moss SM, Summerley ME, Thomas BT, Ellman R, Chamberlain JO. A case-control evaluation of the effect of breast cancer screening in the United Kingdom trial of early detection of breast cancer. *J Epidemiol Community Health* 1992;**46**:362–4
29 Palli D, Rosselli del Turco M, Buiatti E, *et al*. A case-control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy). *Int J Cancer* 1986;**38**:501–4

30 Rothman KJ. A first course in epidemiologic principles and methods. In: Olsen J, Saracci R, Trichopoulos D, eds. *Teaching epidemiology. A guide for teachers in epidemiology, public health and clinical medicine*. Oxford: Oxford University Press, 2001:63–75

31 Jørgensen KJ, Zahl PH, Gøtzsche PC. Breast cancer mortality in organised mammography screening in Denmark: Comparative study. *BMJ* 2010;**340**:c1241

32 Olsen AH, Njor SH, Vejborg I, *et al.* Breast cancer mortality in Copenhagen after the introduction of mammography screening: A cohort study. *BMJ* 2005;**330**:220

33 Wu JC, Anttila A, Yen AM, *et al.* Evaluation of breast cancer service screening programme with a Bayesian approach: Mortality analysis in a Finnish region. *Breast Cancer Res Treat* 2010;**121**:671–8

34 Anttila A, Sarkeala T, Hakulinen T, Heinävaara S. Impacts of the Finnish service screening programme on breast cancer rates. *BMC Public Health* 2008;**8**:38

35 Sarkeala T, Heinävaara S, Anttila A. Organised mammography screening reduces breast cancer mortality: A cohort study from Finland. *Int J Cancer* 2008;**122**:614–9

36 Sarkeala T, Heinävaara S, Anttila A. Breast cancer mortality with varying invitational policies in organised mammography. *Br J Cancer* 2008;**98**:641–5

37 Parvinen I, Helenius H, Pylkkänen L, *et al.* Service screening mammography reduces breast cancer mortality among elderly women in Turku. *J Med Screen* 2006;**13**:34–40

38 Anttila A, Koskela J, Hakama M. Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. *J Med Screen* 2002;**9**:153–8

39 Hakama M, Pukkala E, Heikkilä M, Kallio M. Effectiveness of the public health policy for breast cancer screening in Finland: Population based cohort study. *BMJ* 1997;**314**:864–7

40 Gabe R, Tryggvadóttir L, Sigfússon BF, Olafsdóttir GH, Sigurdsson K, Duffy SW. A case-control study to estimate the impact of the Icelandic population-based mammography screening program on breast cancer death. *Acta Radiol* 2007;**48**:948–55

41 Gorini G, Zappa M, Miccinesi G, Paci E, Seniori Constantini A. Breast cancer mortality trends in two areas of the province of Florence, Italy, where screening programmes started in the 1970s and 1990s. *Br J Cancer* 2004;**90**:1–4

42 Paci E, Giorgi D, Bianchi S, *et al.* Assessment of the early impact of the population-based breast cancer screening programme in Florence (Italy) using mortality and surrogate measures. *Eur J Cancer* 2002;**38**:568–73

43 Puliti D, Miccinesi G, Collina N, *et al.* Effectiveness of service screening: A case-control study to assess breast cancer mortality reduction. *Br J Cancer* 2008;**99**:423–7

44 Barchielli A, Paci E. Trends in breast cancer mortality, incidence, and survival, and mammographic screening in Tuscany, Italy. *Cancer Causes Control* 2001;**12**:249–55

45 Paci E, Duffy SW, Giorgi D, *et al.* Quantification of the effect of mammographic screening on fatal breast cancers: The Florence Programme 1990–96. *Br J Cancer* 2002;**87**:65–9

46 Otten JDM, Broeders MJM, Fracheboud J, Otto SJ, de Koning HJ, Verbeek ALM. Impressive time-related influence of the Dutch screening programme on breast cancer incidence and mortality, 1975–2006. *Int J Cancer* 2008;**123**:1929–34

47 Otto SJ, Fracheboud J, Verbeek ALM, *et al.* Mammography screening and risk of breast cancer death: A population-based case-control study. *Cancer Epidemiol Biomarkers Prev* 2012;**21**:66–73

48 Otto SJ, Fracheboud J, Looman CWN, *et al.* Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: A systematic review. *Lancet* 2003;**361**:1411–7

49 Paap E, Holland R, den Heeten GJ, *et al.* A remarkable reduction of breast cancer deaths in screened vs unscreened women: A case-referent study. *Cancer Causes Control* 2010;**21**:1569–73

50 Broeders MJM, Verbeek ALM, Straatman H, *et al.* Repeated mammographic screening reduces breast cancer mortality along the continuum of age. *J Med Screen* 2002;**9**:163–7

51 Kalager M, Zelen M, Langmark F, Adami HO. Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* 2010;**363**:1203–10

52 Ugarte MD, Goicoa T, Etxeberria J, Militino AF, Pollán M. Age-specific spatio-temporal patterns of female breast cancer mortality in Spain (1975–2005). *Ann Epidemiol* 2010;**20**:906–16

53 Ascunce EN, Moreno-Iribas C, Barcos Urtiaga A, *et al.* Changes in breast cancer mortality in Navarre (Spain) after introduction of a screening programme. *J Med Screen* 2007;**14**:14–20

54 Cabanes A, Vidal E, Pérez-Gómez B, Aragonés N, López-Abente G, Pollán M. Age-specific breast, uterine and ovarian cancer mortality trends in Spain: Changes from 1980 to 2006. *Cancer Epidemiol* 2009;**33**:169–75

55 Pons-Vigués M, Puigpinós R, Cano-Serral G, Marí-Dell'Olmo M, Borrell C. Breast cancer mortality in Barcelona following implementation of a city breast cancer-screening program. *Cancer Detect Prev* 2008;**32**:162–7

56 Haukka J, Byrnes G, Boniol M, Autier P. Trends in breast cancer mortality in Sweden before and after implementation of mammography screening. *PLoS One* 2011;**6**:e22422

57 Chen LS, Yen AM, Duffy SW, Lin WC, Chen HH. Computer-aided system of evaluation for population-based all-in-one service screening (CASE-PASS): From study design to outcome analysis with bias adjustment. *Ann Epidemiol* 2010;**20**:786–96

58 Jonsson H, Bordás P, Wallin H, Nyström L, Lenner P. Service screening with mammography in Northern Sweden: Effects on breast cancer mortality – an update. *J Med Screen* 2007;**14**:87–93

59 Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. *Cancer Epidemiol Biomarkers Prev* 2006;**15**:45–51

60 Baker SG, Kramer BS, Prorok PC. Comparing cancer mortality rates before-and-after a change in availability of screening in different regions: Extension of the paired availability design. *BMC Med Res Methodol* 2004;**4**:1–23

61 Jonsson H, Nyström L, Törnberg S, Lundgren B, Lenner P. Service screening with mammography. Long-term effects on breast cancer mortality in the county of Gävleborg, Sweden. *Breast* 2003;**12**:183–93

62 Duffy SW, Tabar L, Chen HH, *et al.* The impact of organized mammography service screening on breast cancer mortality in seven Swedish counties. *Cancer* 2002;**95**:458–69

63 Tabár L, Vitak B, Chen HHT, Yen MF, Duffy SW, Smith RA. Beyond randomized controlled trials. Organized mammographic screening substantially reduces breast carcinoma mortality. *Cancer* 2001;**91**:1724–31

64 Duffy SW, Tabár L, Olsen AH, *et al.* Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen* 2010;**17**:25–30

65 Allgood PC, Warwick J, Warren RM, Day NE, Duffy SW. A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. *Br J Cancer* 2008;**98**:206–9

66 Fielder HM, Warwick J, Brook D, *et al.* A case-control study to estimate the impact on breast cancer death of the breast screening programme in Wales. *J Med Screen* 2004;**11**:194–8

67 Quinn M, Allen E. Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening. United Kingdom Association of Cancer Registries. *BMJ* 1995;**311**:1391–5

68 Botha JL, Bray F, Sankila R, Parkin DM. Breast cancer incidence and mortality trends in 16 European countries. *Eur J Cancer* 2003;**39**:1718–29

69 Törnberg S, Kemetli L, Lynge E, *et al.* Breast cancer incidence and mortality in the Nordic capitals, 1970–1998. Trends related to mammography screening programmes. *Acta Oncol* 2006;**45**:528–35

70 Autier P, Boniol M, La Vecchia C, *et al.* Disparities in breast cancer mortality trends between 30 European countries: Retrospective analysis of WHO mortality database. *BMJ* 2010;**341**:c3620

71 Autier P, Boniol M, Gavin A, Vatten LJ. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: Trend analysis of WHO mortality database. *BMJ* 2011;**343**:d4411. doi: 10.1136/bmj.d4411

72 Duffy SW, Cuzick J, Tabar L, *et al.* Correcting for non-compliance bias in case-control studies to evaluate cancer screening programmes. *Appl Statist* 2002;**51**:235–43

73 Egger M, Davey Smith G, Altman DG. *Systematic reviews in health care: Meta-analysis in context.* 2nd edn. London: BMJ Books, 2001

74 Gøtzsche PC, Nielsen M. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2011;(1):CD001877

75 Black WC, Haggstrom DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst* 2002;**94**:167–73

76 Lenner P, Jonsson H. Excess mortality from breast cancer in relation to mammography screening in northern Sweden. *J Med Screen* 1997;**4**:6–9

77 Nystrom L, Larsson LG, Rutqvist LE, *et al.* Determination of cause of death among breast cancer cases in the Swedish randomized mammography screening trials. A comparison between official statistics and validation by an endpoint committee. *Acta Oncol* 1995;**34**:145–52

78 Holmberg L, Duffy SW, Yen AMF, *et al.* Differences in endpoints between the Swedish W-E (two county) trial of mammographic screening and the Swedish overview: Methodological consequences. *J Med Screen* 2009;**16**:73–80

79 Tabar L, Duffy S, Yen MF, *et al.* All-cause mortality among breast cancer patients in a screening trial: Support for breast cancer mortality as an end point. *J Med Screen* 2002;**9**:159–62

80 Goldoni CA, Bonora K, Ciatto S, *et al.* Misclassification of breast cancer as cause of death in a service screening area. *Cancer Causes Control* 2009;**20**:533–8

81 Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: Updated overview of the Swedish randomised trials. *Lancet* 2002;**359**:909–19

82 Nelson HD, Tyne K, Nalk A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: An update for the US Preventive Services Task Force. *Ann Intern Med* 2009;**151**:727–37

83 Harris R, Yeatts J, Kinsinger L. Breast cancer screening for women ages 50 to 69 years of age: A systematic review of observational evidence. *Prev Med* 2011;**53**:108–14
84 Connor RJ, Boer R, Prorok PC, Weed DL. Investigation of design and bias issues in case-control studies of cancer screening using micro-simulation. *Am J Epidemiol* 2000;**151**:991–8
85 Duffy SW. Case-control studies to evaluate the effect of mammographic screening on mortality from breast cancer. *Semin Breast Disease* 2007;**10**:61–3
86 van Schoor G, Paap E, Broeders MJ, Verbeek AL. Residual confounding after adjustment for age: A minor issue in breast cancer screening effectiveness. *Eur J Epidemiol* 2011;**26**:585–8
87 Paap E, Verbeek ALM, Puliti D, Broeders MJM, Paci E. Minor influence of self-selection bias on the effectiveness of breast cancer screening in case-control studies in the Netherlands. *J Med Screen* 2011;**18**:142–6

APPENDIX A

Search Strategy

Evaluation of the effect of service screening programmes with mammography on the breast cancer mortality in Western Europe

SEARCH STRATEGY

We searched the National Library of Medicine PubMed with the following search terms:

(1) 'Mortality'[Mesh]
(2) 'Mass Screening'[Mesh]
(3) 'Mammography'[Mesh]
(4) 'Breast Neoplasms/mortality*'[Mesh]
(5) breast cancer mortality
(6) screening
(7) mammography
(8) ((#1) OR #4) OR #5
(9) (#2) OR #6
(10) (#7) OR #3
(11) **((#8) AND #9) AND #10**

This search strategy retrieved a total of 2462 papers.

(12) effect*
(13) evaluation
(14) impact
(15) trend
(16) service screening
(17) programme screening
(18) mass screening
(19) breast cancer
(20) mortality
(21) survival
(22) (((#12) OR #13) OR #14) OR #15
(23) ((#16) OR #17) OR #18
(24) (#20) OR #21
(25) **(((#22) AND #23) AND #19) AND #24**

This search strategy retrieved a total of 1680 papers.

(26) 'Mortality/trends'[Mesh]
(27) 'Survival Analysis'[Mesh]
(28) 'Survival Rate/trends'[Mesh]
(29) ((#26) OR #27) OR #28
(30) **((#29) AND #2) AND #4**

This search strategy retrieved a total of 193 papers.

(31) PubMed 'related articles' to the following article suggested by experts in the field, not retrieved by the previous search strategies:

Otten JDM, Broeders MJM, Fracheboud J, Otto SJ, de Koning HJ, Verbeek ALM. Impressive time-related influence of the Dutch screening programme on breast cancer incidence and mortality, 1975–2006. *Int J Cancer* 2008;123:1929–34.

This search strategy retrieved a total of 726 papers.

These searches were supplemented with suggestion by experts in the field.

The results were sorted by Europe Western Countries: The Netherlands, Finland, Sweden, Norway, Iceland, Denmark, UK, Ireland, Germany, Austria, Italy, Spain, Greece, Nordic countries, Europe (not specified).

We considered all articles published in English language up to February 2011 (no date restriction); the articles were imported into ProCite to select the papers considered relevant after the reading of title and abstracts.