TYTTI SARKEALA

Performance and Effectiveness of Organised Breast Cancer Screening in Finland

■

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine of the University of Tampere,
for public discussion in the Auditorium of
Tampere School of Public Health, Medisiinarinkatu 3,
Tampere, on May 16th, 2008, at 12 o'clock.

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere, School of Public Health
Doctoral Programs in Public Health (DPPH)
Finnish Cancer Registry
Finland

Supervised by
Docent Ahti Anttila
University of Tampere
Professor emeritus Matti Hakama
University of Tampere

Reviewed by
Professor Peter B. Dean
University of Turku
Docent Pia Verkasalo
University of Kuopio

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 3 3551 6055
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
http://granum.uta.fi

Cover design by
Juha Siro

Acta Universitatis Tamperensis 1309
ISBN 978-951-44-7299-2 (print)
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 717
ISBN 978-951-44-7300-5 (pdf)
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2008

CONTENTS

ABBREVIATIONS

Bcd	= breast cancer death
Bcm	= breast cancer mortality
B:M ratio	= benign to malignant biopsy ratio
CBE	= clinical breast examination
CGHFBC	= Collaborative Group on Hormonal Factors in Breast Cancer
CI	= confidence interval
CSF	= Cancer Society of Finland
DCIS	= ductal carcinoma in situ
EU	= European Union
FCR	= Finnish Cancer Registry
GISMa	= Italian Group of Mammography Screening
HIP	= Health Insurance Plan of Greater New York
HRT	= hormone replacement therapy
IARC	= International Association for Research on Cancer
IR	= background incidence rate
LCIS	= lobular carcinoma in situ
M	= mammography test
MISCAN	= MIcrosimulation SCreening ANalysis
P	= teaching of practice of self-examination
PPV	= positive predictive value
pT	= pathological classification for primary tumour
pN	= pathological classification for lymph nodes
pM	= pathological classification for distant metastasis
RR	= risk ratio
SOSSEG	= Swedish Organised Service Screening Evaluation Group
UK	= United Kingdom

ABSTRACT

Background and objective. Randomised trials have shown that invitation to screening reduces mortality from breast cancer. Since the results from rigorously conducted trials may not be directly convertible into routine service, and the validity of some trials has been questioned, evaluation of population-based service screening is essential. This study retrospectively evaluated the performance and the effectiveness of organised breast cancer screening in Finland by estimating the process and validity indicators of screening, the effect of screening on breast cancer mortality, and the relationships between process, validity and outcome of screening.

Material and methods. The study subjects were 40-74 year old women who had been invited to breast cancer screening in ten centres of the Cancer Society of Finland (CSF) from 1991 onwards. Invitations from CSF centres accounted for 50-60% of screening invitations in Finland in the study period.

The indicators for the screening process, e.g. screening coverage and attendance, recall rate, breast cancer detection rate, and positive predictive value of mammography, were examined among the 40-74 year old invitees and participants. The indicators were measured separately for the first and subsequent screens by five-year age groups, invitational year, and screening centre in 1991-2000. The indicator for screening validity, episode sensitivity by detection and incidence methods, was examined among the 50-64 year old participants. The sensitivity was measured separately for the first and for the subsequent screens over the whole study period 1991-2001, during the first and the second year after the screening visit, and by five-year age groups, three-year periods, and three centre categories. The categorisation of screening centres was based on recall rates and visits for each centre at the subsequent screens.

The overall incidence-based breast cancer mortality was analysed in 1992-2003 among the invitees and participants of screening aged 50-69 at death. Mortality was measured by five-year age groups at death and by the same centre categories that were used in the validity analysis. Observed deaths from breast cancer were compared with expected breast cancer deaths without screening. The observed deaths were obtained from a cohort of individual invitees (n=361 848). The expected deaths were defined by modelling population level breast cancer mortality from 1974 to 1985 and 1992 to 2003. The population data were derived from the same municipalities (n=260) that were incorporated into the cohort. The analysis of incidence-based breast cancer mortality was further extended to ages 60-79 by comparing observed and expected breast cancer deaths among all female and among screened and non-screened women in three invitational policies (regular screening at ages 50-59; regular screening at ages 50-69; irregular screening at ages 50-69).

Results. In 1991-2000, the CSF centres sent more than one million invitations, and processed 930 000 visits. The coverage of invitations was more than 95% among women aged 50-59 years, 20-40% among women aged 60-69 years, and less than 10% among 40-49 and 70-79 year old women. The overall attendance at ages 50-64 was 90%. Per one subsequent screen, less than 3% of attendees were sent for further examinations. Their breast cancer detection rate was 0.4%. The positive predictive value of mammography increased towards the end of the study period and with age.

In 1991-2001, the overall episode sensitivity at the subsequent screens was 54% by incidence, and 65% by detection method. The sensitivity estimates 0-11 and 12-23 months after the screening visit

were 70% and 38% by the incidence method. The sensitivity increased 13% per 1 % absolute increase in the recall rate.

In 1992-2003, the overall reduction in the incidence-based breast cancer mortality among the screening invitees aged 50-69 at death was 22% (relative risk 0.78, 95% confidence interval 0.70-0.87). After adjusting for self-selection, the reduction among the participants became 28% (0.72, 0.56-0.88). Among all females aged 60-79 at death, the greatest reduction in breast cancer mortality, 28% (0.72, 0.51-0.97), was estimated in municipalities inviting 50-69 year old women on a regular basis. In municipalities restricting their invitations to 50-59 year old women, no reduction in breast cancer mortality at ages 60-79 was observed.

The overall performance indicators in Finland were at a level similar to those in other European screening programmes, but the centre-specific differences were wide. Variability in centre-specific episode sensitivity was associated with the variability in recall. In further analyses, no association between the level of recall and the screening outcome could be found.

Conclusion. Organised breast cancer screening in Finland has been effective. Extension of invitations to women aged 60-69 will decrease breast cancer deaths among the elderly. Due to ambiguity in the relationships between the estimates of performance and outcome, evaluation of breast cancer mortality remains the most reliable means to ascertain the success of service screening.

LIST OF ORIGINAL PUBLICATIONS

The thesis is based on the following publications, referred to by Roman numerals (I-V) in the text:

I Sarkeala T, Anttila A, Forsman H, Luostarinen T, Saarenmaa I, Hakama M (2004): Process indicators from ten centres in the Finnish breast cancer screening programme from 1991 to 2000. Eur J Cancer 40**:** 2116-25.

II Sarkeala T, Anttila A, Saarenmaa I, Hakama M (2005): Validity of process indicators of screening for breast cancer to predict mortality reduction. J Med Screen 12: 33-7.

III Sarkeala T, Hakama M, Saarenmaa I, Hakulinen T, Anttila A (2006): Episode sensitivity in association with process indicators in the Finnish breast cancer screening program. Int J Cancer 118**:** 174-9.

IV Sarkeala T, Heinavaara S, Anttila A (2008a): Organised mammography screening reduces breast cancer mortality: A cohort study from Finland. Int J Cancer 122**:** 614-9.

V Sarkeala T, Heinävaara S, Anttila A (2008b): Breast cancer mortality with varying invitational policies in organised mammography. Br J Cancer 98: 641-5.

1. INTRODUCTION

Breast cancer is the commonest cancer among women worldwide, with the highest incidence rates in the high income countries (www-dep.iarc.fr/). In Finland, breast cancers accounted for 32% of all incident cancers among women in 2005. The five-year relative survival ratio for breast cancer in 2003-2005 was 89% (www.cancerregistry.fi/eng/statistics, FCR 2007).

The risk for breast cancer is connected to female hormones and reproductive patterns, and also to certain characteristics of life-style (Key et al. 2001). Since breast cancers occur in women having none clearly identifiable risk factors, gaps remain in the current aetiology of the disease.

Attempts to reduce breast cancer incidence by primary prevention have been discouraging (Clamp et al. 2002). Screening provides a secondary means to move forward. The goal of screening is to reduce mortality from breast cancer by detecting tumours at pre-clinical stage: early breast cancers are considered more responsive to treatment than tumours detected by clinical signs and symptoms (Morrison 1991, Vainio and Bianchini 2002). Besides mortality reduction, screening should promote quality of life by minimising false-positive and false-negative findings that may cause delays in tumour diagnosis and treatment.

Several randomised trials have been designed to study the effect of screening on breast cancer mortality (Tabar and Gad 1981, Roberts et al. 1984, Tabar et al. 1985, Frisell et al. 1986, Andersson et al. 1988, Roberts et al. 1990, Bjurstam et al. 1997, Hakama et al. 1997, Shapiro 1997, Alexander et al. 1999, Miller et al. 2000, Miller et al. 2002, Bjurstam et al. 2003). The overall results in ages 50-69 at randomisation have shown a 25% reduction in breast cancer mortality among invitees of screening (Vainio and Bianchini 2002). In view of the randomised trials, many European countries

have introduced population-based screening programmes (Shapiro et al. 1998, Ballard-Barbash et al. 1999).

Since the randomised trials represent the effect of screening performed under optimum conditions (Day 2005), and the validity of some trials has been questioned (Gotzsche and Olsen 2000, Olsen and Gotzsche 2001), efforts have been intensified to analyse the performance and effectiveness of screening applied on a routine basis (e.g. Otto et al. 2003, Olsen et al. 2005a, SOSSEG 2006b).

In Finland, population-based breast cancer screening was started in 1987. The effectiveness for the first five years of the programme was examined by randomised birth cohorts, and demonstrated a 24%, non-significant reduction in breast cancer mortality among invitees (Hakama et al. 1997). The screening programme was fully implemented at the beginning of 1992. Thereafter, one study has assessed the delivery of routine screening in Finland (Dean and Pamilo 1999). The effect of routine screening on breast cancer mortality has been examined in two Finnish cities, Helsinki and Turku (Anttila et al. 2002, Parvinen et al. 2006).

This study evaluated the performance of organised breast cancer screening, and the effect of screening invitation as well as that of participation upon breast cancer mortality using individual data on women who were invited to ten centres of the Cancer Society of Finland between 1991 and 2003.

2. REVIEW OF THE LITERATURE

2.1 Epidemiology and development of breast cancer

Incidence and mortality

Breast cancer is the commonest cancer among women worldwide with an estimated number of 1 150 000 new cases and 410 000 deaths in 2002 (www-dep.iarc.fr/). Age-standardised incidence rates are highest in the high income countries, and they are on the increase throughout the world. The mortality from breast cancer has recently levelled off or declined in the high income countries, but breast cancer still causes most deaths from cancer among women in these areas (Hermon and Beral 1996, Vainio and Bianchini 2002, Boyle and Ferlay 2005).

In Finland, breast cancer incidence has increased considerably since the 1960´s. The overall, age-adjusted incidence rate in 2005 was 87.1 per 100 000 person-years, and the corresponding mortality rate 14.8. The number of new breast cancer cases was 4021, and the number of breast cancer deaths 829 . In 2001-2005, breast cancer incidence was highest in the age group 60-64, while the mortality from breast cancer increased steadily with age. Highest overall incidence and mortality rates were reported from the southern part of the country (www.cancerregistry.fi/eng/statistics, FCR 2007).

Biologic and pathologic characteristics

Alterations in breast tissue can roughly be divided into three groups: benign condition, in-situ cancer and invasive cancer. Benign conditions may manifest themselves as fibroadenoma or fibrocystic change. Between these and cancer, there are benign lesions that impart an increased risk of subsequent carcinoma, e.g. proliferative breast disease (atypical ductal or lobular hyperplasia)

and sclerosing adenosis (Jensen et al. 1989, Dupont et al. 1993, Dupont et al. 1994, Marshall et al. 1997). Ductal carcinoma in situ (DCIS) and lobular carcinoma in situ (LCIS) represent the two non-invasive forms of breast carcinoma, and arise from epithelial cell population in the duct or lobule of the breast. Invasive carcinoma is a malignant tumour, part or all of which penetrates the basement membrane of the epithelial site of origin (the duct or lobule) to the adjacent breast stroma (Vainio and Bianchini 2002). The development of breast cancer from atypia to carcinoma in situ, invasive cancer and metastasis is not straightforward: breast cancer can be characterised as a heterogeneous group of diseases with more than one natural history (Buerger et al. 1999, Buerger et al. 2001).

The prognosis for tissue alteration in breast is related to time-dependent and intrinsic variables. Time-dependent variables influence tumour size, lymph node status, and the presence of systemic metastatic disease. Intrinsic characteristics are related to histological grade, growth fraction, hormone and growth factor receptor status, and molecular characteristics, and some of these may also be time-dependent. The variables shown to be related to clinical outcome are tumour size, histological grade, and lymph node status. In tumours measuring 15 mm or more, the frequency of lymph node metastasis has been over 40%. Lymph node status and grade predict the clinical outcome equally (Rosen and Groshen 1990, Galea et al. 1992, Elston et al. 1999).

In benign conditions, the risk for further development to cancer depends on the degree of epithelial proliferation and atypia. High grade ductal carcinoma in situ is associated with an increased risk for invasive cancer. Low grade ductal carcinoma in situ is associated with low-grade invasive cancer, which has been related to good prognosis (Silverstein et al. 1995, Douglas-Jones et al. 1996, Silverstein et al. 1996, Cadman et al. 1997, Burstein et al. 2004).

Risk factors

Breast cancer incidence among women is more than hundred times higher than the incidence among men (Fentiman et al. 2006), and increases with age. The highest rate of increase is seen from early adulthood until about 50 years of age, i.e. between menarche and menopause (e.g. FCR 2007). Breast cancer risk factors are related to female hormones and reproductive life: low parity, late age at first pregnancy, early menarche and late menopause are consistently associated with an increased risk of tissue alteration in breast (de Waard and Trichopoulos 1988, Ewertz et al. 1990, Bernstein and Ross 1993, Kerlikowske et al. 1997, McPherson et al. 2000, Key et al. 2001). Diet (obesity), physical activity, alcohol intake and use of exogenous hormones (oestrogen, progesterone) also have a role in breast cancer aetiology (Brinton and Schairer 1993, CGHFBC 1997, Weisburger 1997, Clemons and Goss 2001, CGHFBC 2002, Chen et al. 2002, Vainio et al. 2002, Chlebowski et al. 2003, Rintala et al. 2003, Bakken et al. 2004, Cade et al. 2007, Tjonneland et al. 2007). Heritable factors explain about one quarter of breast cancers, and 5-10% of breast cancer cases are due to germline mutations in cancer-susceptibility genes showing autosomal dominant inheritance (Easton et al. 1993, Gayther et al. 1997, Struewing et al. 1997, Lichtenstein et al. 2000, Clamp et al. 2002). Nevertheless, breast cancers often occur in women having none clearly identifiable risk factors (Clamp et al. 2002).

Prevention

The aim of primary prevention is to eliminate or modify the risk factors of a disease. The attempts to reduce breast cancer incidence by primary prevention, such as chemoprevention, have been discouraging (Clamp et al. 2002, Cuzick et al. 2003). Moreover, changes in the reproductive

patterns or in the exposure to other risk factors (e.g. controlling weight or physical activity) are difficult to achieve.

Secondary prevention aims to inhibit the progression of localised cancer to advanced, metastasized and incurable cancer, and thus offers an opportunity to reduce the consequences of the increasing incidence of breast cancer by early detection and treatment. In recent decades, one of the means for the secondary prevention has been breast cancer screening.

2.2 Breast cancer screening

The goal of breast cancer screening is to reduce mortality from breast cancer. The means to achieve this goal is to detect breast cancers before they become clinically evident. The core concept of breast cancer screening thus includes an assumption of the ultimate progression of an untreated lesion to a clinical cancer (Vainio and Bianchini 2002). It also assumes that detection of early disease enables treatment at a more tractable stage (Morrison 1991).

A model for breast cancer progression with the intervention of screening is presented in Figure 1. In the model there is first a period in which there is no detectable disease, but early malignant changes take place. The point at which a breast tumour or lesion can be found by screening is the beginning of sojourn time (or preclinical detectable phase) (Zelen and Feinleib 1969, Day et al. 1989). The sojourn time ends when cancer appears through clinical signs and symptoms. Lead time is a period between the detection of preclinical disease and the appearance of a clinical tumour (Day et al. 1989). Length of sojourn time is related to the characteristics of the breast tumour and to the ability of screening to detect these tumours. Length of lead time is also affected by the frequency of screening (Vainio and Bianchini 2002).

Figure 1. Model for breast cancer progression with intervention of screening.

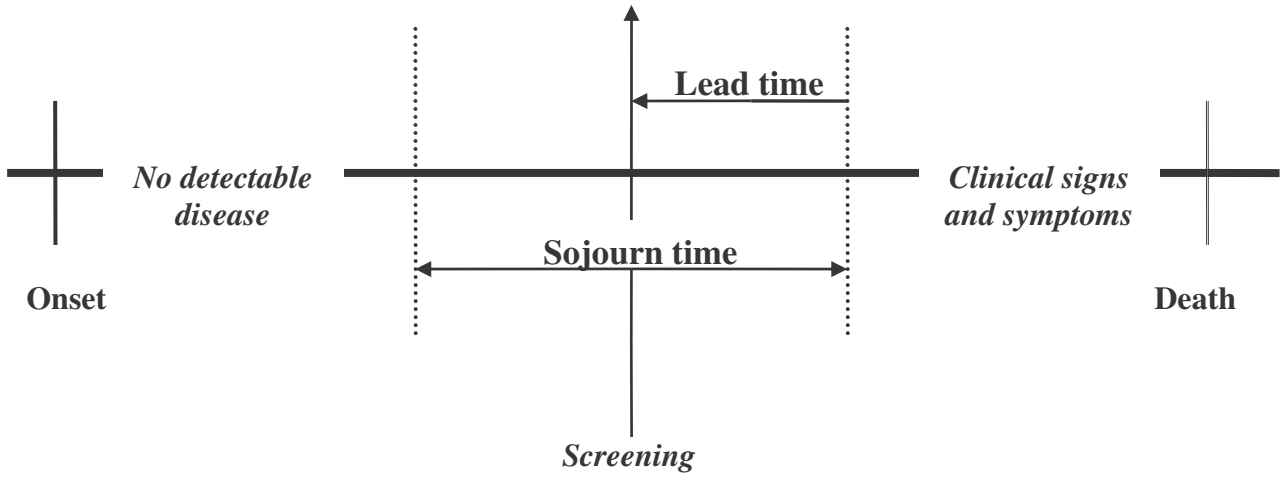
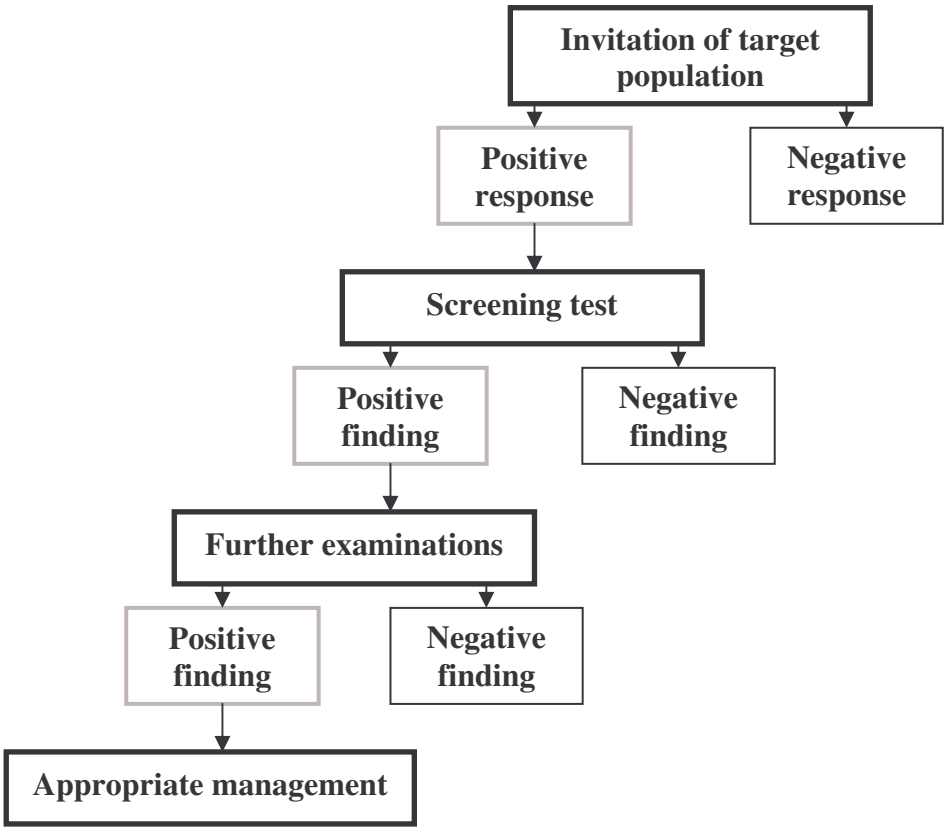


Figure 2. Process of breast cancer screening.



Process and main definitions

Breast cancer (mammography) screening is a multiple-step process, which starts with identification and invitation of the target population (Figure 2). Screening test, x-ray images (mammograms) from both breasts is performed for all attendees. After interpretation of images, those with positive findings are sent for further examinations. For selected women, the screening episode culminates in surgical operation, where the diagnosis of tissue alteration is, or is not, confirmed. Invitations and screenings can be repeated, and a woman may experience many mammograms, occasional recall examinations and one or more surgical operations during her lifetime.

When a population of women goes for screening, a certain number of breast cancers are detected at the first or subsequent screens. Further lesions are diagnosed clinically during the post-screening periods among those with negative results at previous screens. Screening validity, the extent to which screening measures what it is intended to measure, is assessed by sensitivity, specificity and predictive values (Table 1). *Screening sensitivity* measures the ability of screening to detect breast cancer, *screening specificity* refers to the ability to identify an attendee without breast cancer. *Positive predictive value* represents the probability of breast cancer after a positive screening test, *negative predictive value* the probability that someone with a negative screening result does not have a breast cancer.

The impact of screening on breast cancer mortality can be measured among those targeted for screenings (*effectiveness*) and among those actually participating (*efficacy*). The performance parameters of screening, such as *coverage* among those targeted for screening, *attendance* among those invited to screening, and sensitivity and specificity, assess the screening quality and predict the potential impact of screening on breast cancer mortality (Day et al. 1989, Perry et al. 2006).

Table 1. General layout of validity table.

		Preclinical detectable disease	
		Yes	**No**
Result of screening	**Positive**	**tp**	fp
	Negative	fn	**tn**

tp= true positive; tn= true negative; fp= false positive; fn= false negative

Sensitivity = **tp**/(**tp**+fn)
Specificity = **tn**/(**tn**+fp)

Positive predictive value = **tp**/(**tp**+fp)
Negative predictive value = **tn**/(**tn**+fn)

Uncertainties

The probability of a breast cancer being detected at screening depends on the length of time the lesion is detectable pre-clinically. The longer this sojourn time is the greater is the chance that the lesion will be found by screening. Screen-detected breast cancers thus represent a sample of lesions with a bias towards a longer sojourn time and, hence, a better prognosis. This bias is known as the length bias. Breast cancers detected at first screen are most biased, as lesions with a sojourn time that is long in comparison with the inter-screening interval are overrepresented at the first screening test (Vainio and Bianchini 2002). The length bias is counteracted by another bias toward advanced cancers that tend to be detected with greater frequency at the first screen.

Among the slow-growing breast cancers detected at screening may be a proportion that would not have surfaced clinically in the lifetime of the individual. These lesions constitute over-diagnosis. Due to over-diagnosis bias, the individuals undergo unnecessary further examinations and treatment, and the health care system has to bear excessive costs (Day 2005).

Lead time is the length of time the diagnosis is advanced by screening (Day et al. 1989). Lead-time bias refers to the longer lifespan among individuals whose breast cancers are identified by screening compared to those whose breast cancers are detected clinically. It may thus distort the comparison of survival between those screened and those not screened even if there were no true benefits of screening (Vainio and Bianchini 2002, de Koning 2003).

Screening attendees have been shown to be more health-conscious than general population, and are thus more likely to have a better outcome (Hakama et al. 1997, Vainio and Bianchini 2002, Zackrisson et al. 2004). Methods have been developed to assess the magnitude of this selection bias, and to adjust the estimates of screening efficacy accordingly (e.g. Cuzick et al. 1997).

Lower sensitivity and specificity of screening have been reported among women with dense breasts compared to women with more radiolucent, fatty breasts (Mandelson et al. 2000, Ciatto et al. 2004, Chiarelli et al. 2006, McCormack and dos Santos Silva 2006, Boyd et al. 2007, MacKenzie et al. 2007). Young age and use of hormone replacement therapy (HRT) are the most common features associated with breast density (Kerlikowske et al. 1996, Carney et al. 2003, Bremnes et al. 2007). The use of HRT may affect screening sensitivity through other mechanisms also, such as subtle changes in breast tissue or increased surveillance (Kavanagh et al. 2005, Hofvind et al. 2006b).

2.3 Randomised trials of breast cancer screening

The main objective of randomised screening trial is to evaluate whether screening reduces the risk of dying from breast cancer among those free from breast cancer at enrolment. The subjects are randomly allocated to invitees and non-invitees of screening. This should ensure that the assignment of the subject is determined by chance alone, and is not influenced by the investigator, by the

subject, or by other observable characteristics. To confirm equivalence between the randomised invitees and non-invitees, baseline data on variables which are known or thought to affect the outcome are needed. Moreover, mammography of study subjects outside the trial needs to be followed, similar treatment for invitees and non-invitees guaranteed, and confirmation of cause of death derived from a reliable source. Analysis of the final outcome needs to be performed according to the intention to treat principle, in which the invitees include both the attendees and non-attendees of screening (dos Santos Silva 1999, Vainio and Bianchini 2002).

Trial designs

The first randomised trial in breast cancer screening, the Health Insurance Plan of Greater New York (HIP), was started in the United States in 1963 (Shapiro et al. 1971, Shapiro 1997). The HIP study as well as the later established seven randomised trials in Scotland, Canada, Sweden and Finland formulated the scientific basis for the future screening programmes (Tabar and Gad 1981, Roberts et al. 1984, Tabar et al. 1985, Frisell et al. 1986, Andersson et al. 1988, Roberts et al. 1990, Bjurstam et al. 1997, Hakama et al. 1997, Alexander et al. 1999, Miller et al. 2000, Tabar et al. 2000, Miller et al. 2002, Bjurstam et al. 2003). The latest randomised study on breast cancer screening was established between 1991 and 1997 in the United Kingdom to estimate the effectiveness of breast cancer screening among women aged 40-49 years at entry (Moss et al. 2006).

In the trials, the randomisation of invitees was performed on an individual or cluster basis. Age at entry varied between 40-70 years, and attendance between 61 and 100%. The determination of breast cancer death was based on independent or non-independent reviews of death certificates or on official statistics. Table 2 presents the basic characteristics of all randomised designs.

Table 2. Overview of randomised designs.

	HIP USA	Edinburgh Scotland	Canada I	Canada II	Malmö I Sweden	Malmö II Sweden	Kopparberg Sweden	Östergötland Sweden	Stockholm Sweden	Göteborg Sweden	Finland	United Kingdom
Randomisation If cluster, type	Individual	Cluster General practice	Individual	Individual	Individual	Individual	Cluster Municipality, tax district	Cluster Municipality, parish	Cluster Day of birth	Cluster Day of birth	Cluster Birth cohort	Individual
Number of women	60 995	45 130	50 430	39 405	42 283	17 793	56 448	92 872	60 800	52 222	158 755	160 921
Acrual period Invited group	Dec 1963– June 1996	1978-1981 1982-1983 1984-1985	Jan 1980– Mar 1985	Jan 1980– Mar 1985	Oct 1976– Aug 1978	Sept 1978– Nov 1990	Jul 1977– Feb 1980	May 1978– March 1981	March 1981– May 1983	Dec 1982– April 1984	1987- 1988	1991-1997
Control group	-	-	-	-	Oct 1992– Feb 1993	Sept 1991– April 1994	Sept 1982– Dec 1986	April 1986– Feb 1988	Oct 1985– May 1986	Nov 1987– June 1991	1990- 1991	-
Birth cohorts					1908-32	1933-45		1903-40	1917-42	1923-44	1927-30 1932-35 1938-39	
Age at entry	40-64	45-64	40-49	50-59	45-70	43-49	40-74	38-75	39-65	39-59	49-60	39-41
Intervention	M+CBE	M+CBE	M+CBE+P	M+CBE+P	M	M	M	M	M	M	M	M
Number of views	2	Screen 1: 2 Screen 2+: 1	2	2	2	2	1	1	1	Round 1: 2 Round 2: 1-2	2	Screen 1: 2 Screen 2+: 1
Number of readers		1			2	2	1	1	1	Rounds 1-3: 1 Rounds 4-5: 2	2	
Screening interval, months	12	24	12	12	18-24	18-24	24, 33	24, 33	28	18	24	12
Number of rounds	4	2-4	4-5	4-5	1908-17: 6 1918: 7 1919-32: 8	1-7	70-74 yrs: 2 50-69 yrs: 3 40-49 yrs: 4	70-74 yrs: 2 50-69 yrs: 3 40-49 yrs: 4	2	1923-32: 4 1933-44: 5	1-3	6
Attendance rate	67%	61%	*	*	74%	75-80%	89%	89%	82%	84%	85%	68%

M= mammography test; CBE= clinical breast examination; P= teaching of practice of self-examination
*Participants were volunteers who fulfilled the eligibility criteria of the study

HIP (Shapiro 1997); Edinburgh (Roberts et al. 1984, Alexander et al. 1999); Canada I, II (Miller et al. 2000, Miller et al. 2002); Malmö I, II (Anderson et al. 1988, Nyström et al. 2002); Kopparberg (Tabar and Gad 1981, Tabar et al. 1985); Östergötland (Tabar and Gad 1981, Tabar et al. 1985, Nyström et al. 2002); Stockholm (Frisel et al. 1986, Nyström et al. 2002); Göteborg (Bjurstam et al. 1997, Bjurstam et al. 2003, Nyström et al. 2002); Finland (Hakama et al. 1997); United Kingdom (Moss et al. 2006)

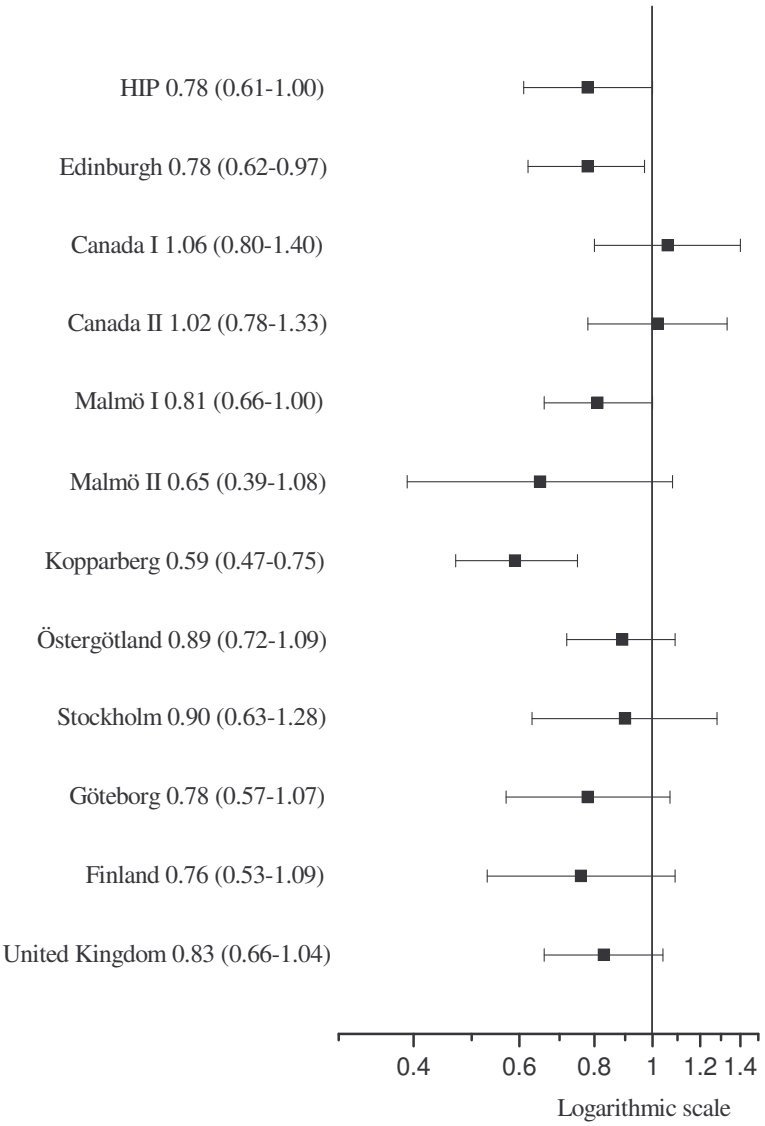
Trial outcomes

The HIP study randomised 30 294 women to the study group and 30 756 women to the control group in ages 40-64 at entry (Shapiro 1997). After 18 years of follow-up, there was 22% lower breast cancer mortality in the study group compared to the control group (Figure 3). In Edinburgh, Scotland the intervention cohort consisted of 26 868 women and the control cohort 26 026 women aged 45-64 at entry (Alexander et al. 1999). After 14 years of follow-up, the breast cancer mortality in the study cohorts was 13-17% lower than in the control cohorts. Cluster randomisation had resulted in differences by socioeconomic status between the intervention and the control groups, however. After adjustment, the reduction in breast cancer mortality was 22% (relative risk 0.78, 95% confidence interval 0.62-0.97). In Canada, trials were organised for women aged 40-49 years at entry (3569/3674 women), and 50-59 years at entry (2164/2207 women) (Miller et al. 2000, Miller et al. 2002). After 11 to 16 years of follow-up, the cumulative rate ratio was 1.06 (0.80-1.40) in the 40-49 years old women at entry, and 1.02 (0.78-1.33) in women aged 50-59 years at entry. In Finland, Hakama and colleagues (1997) compared deaths from breast cancer among women aged 49-64 at entry in a group-randomised design in 1987-1991. The breast cancer mortality rate ratio among screening invitees was 0.76 (0.53-1.09).

Several studies have reported the mortality results of the Swedish randomised trials in Malmö (Malmö I and Malmö II), Kopparberg (Two-County), Östergötland (Two-County), Stockholm, and Göteborg (Figure 3). The first results from Kopparberg were published in 1985 (Tabar et al. 1985), the most recent report in 2000 (Tabar et al. 2000). A 41% reduction in breast cancer mortality was maintained 20 years after randomisation among women aged 40-74 years at entry. An updated, pooled analysis of the Malmö (Malmö I,II), Östergötland, Stockholm and Göteborg trials was published in 2002 (Nystrom et al. 2002). The median follow-up time (time from randomisation to

the end of follow-up) of these trials was 15.8 years. There were 511 breast cancer deaths in 1 864 770 person-years in the invited groups and 584 breast cancer deaths in 1 688 440 person-years in the control groups. A significant 21% (0.79, 0.70-0.89) reduction in breast cancer mortality was observed among women aged 40-74 years at entry.

Figure 3. Relative risks with 95% confidence intervals in the randomised trials in ages 40-74 at entry.



HIP (Shapiro 1997); Edinburgh (Roberts et al. 1984, Alexander et al. 1999); Canada I, II (Miller et al. 2000, Miller et al. 2002)
Malmö I, II (Anderson et al. 1988, Nyström et al. 2002); Kopparberg (Tabar and Gad 1981, Tabar et al. 1985)
Östergötland (Tabar and Gad 1981, Tabar et al. 1985, Nyström et al. 2002); Stockholm (Frisel et al. 1986, Nyström et al. 2002)
Göteborg (Bjurstam et al. 1997, Bjurstam et al. 2003, Nyström et al. 2002); Finland (Hakama et al. 1997); United Kingdom (Moss et al. 2006)

The updated overview of Nyström and colleagues (2002) was published partly due to intensive debate on the value of screening mammography, started by Peter Gotzsche and Ole Olsen (Gotzsche and Olsen 2000). Gotzsche and Olsen did a meta-analysis of randomised trials (Hakama et al. (1997) design was not included) using the Cochrane Library Review Manager, and found imbalances in age distribution and in exclusion of those having breast cancer at the time of randomisation, as well as inconsistencies in the number of women randomised. They were also concerned about the diagnoses of breast cancer deaths, because knowledge of screening status may affect the judgement regarding cause of death. Moreover, the invited cohorts in Sweden seemed to exhibit higher all-cause mortality than the controls. According to Gotzsche and Olsen (2000), only two trials, Canada and Malmö (Malmö I), were adequately randomised with masked assessment of cause of death, and these trials found no effect of screening on breast cancer mortality (pooled relative risk 1.04). Gotzsche and Olsen (2000) concluded that screening for breast cancer with mammography was unjustified.

In their response, Nyström and colleagues pointed out that the population-based trials in Sweden invited women to mammography alone while the HIP, the Edinburgh and the Canadian trials evaluated mammography combined with breast self-examination, clinical breast examination, or both (Nystrom et al. 2002). In addition, the HIP and the Canadian trials were not population-based. The randomisation methods of the Swedish trials were presented in detail. The researchers underlined that the marginal imbalances in age (cases were 0.09-0.18 years younger than controls in Göteborg and Stockholm, and 0.27-0.45 years older than controls in Östergötland and Kopparberg) did not indicate a biased allocation of women. The discrepancy in the reported numbers of women randomised could be explained by the fact that the trial reports referred to all randomised women according to their birth cohort, while the overview figures referred to women aged exactly 40-74 years at randomisation. The determination of cause of death was done according to the Statistics of

Sweden, and resulted in more conservative estimates than a blind determination of cause of death by an independent endpoint committee (Nystrom et al. 1995). According to Nystrom and colleagues (2002), the assumption of higher total mortality among invited women compared to that among controls was based on a misunderstanding about inappropriate age-adjustment, and they insisted, that age-adjustment was both necessary and appropriate, because the cluster randomisation in the Malmö and the Stockholm trials might have resulted in imbalances in age distribution.

In 2002, the working group of the International Association for Research on Cancer (IARC) summarized the results of randomised trials in breast cancer screening (Vainio and Bianchini 2002). The working group took into account the criticism of Gotzche and Olsen (2000), the updated overview by Nystrom et al. (2002), and all the original publications on randomised trials, and reanalysed the results in detail. The HIP study was considered valid, but was excluded from the evaluation because the screening test included clinical breast examination and the trial was not population-based. The Edinburgh trial was excluded because confounding from other variables than socioeconomic status could not be overtaken. The Canadian trials were excluded from the evaluation because both trials employed breast self-examination and the women were recruited through media and physicians. All the Swedish trials were considered valid. The evaluators of the IARC also included the information on the Finnish group-randomised design in their summary. In their conclusion the working group stated: 'There is sufficient evidence from randomised trials that inviting women 50-69 years of age to screening with mammography reduces their mortality from breast cancer; the best current estimate of the average reduction is 25%'.

The evidence regarding mortality reduction among women aged 40-49 remained limited in the summary statement of the IARC. In 2006, the results from a UK randomised trial on the effectiveness of breast cancer screening at ages 40-49 at entry was published (Moss et al. 2006)

(Figure 3). A 17%, non-significant reduction in breast cancer mortality was accompanied by a high proportion of false positive findings and substantial reductions in the positive predictive value compared to values among those invited after the age of 50. The authors concluded that further follow-up was needed to make decisions on screening policy for those aged 40-49. The decisions should take into account the disadvantages as well as the benefits.

There are several studies published on the performance of the Two-County trials (Tabar and Gad 1981, Tabar et al. 1985, Tabar et al. 1987, Day et al. 1989, Duffy et al. 1991), and the reported performance indicators of these trials have been used as a basis for quality assurance criteria of the European Union (Perry et al. 2006) (see Table 4). Comparatively little information on screening performance has been reported from the other randomised trials. The women from the Malmö I trial formed baseline data to assess the level of over-diagnosis in Sweden in 1976-2001 (Zackrisson et al. 2006). The study reported a 10%, non-significant, increase in the cumulative rate of breast cancers fifteen years after the last screen (1.10, 0.99-1.22). The result was reported among women aged 55-69 at randomisation whose control groups were never screened.

2.4 Organised breast cancer screening in Europe

Ways of delivery

In view of the randomised trials, several European countries introduced breast cancer screening in the late 1980's and the 1990's. The introduction was performed in co-operation with the existing health care and financing systems. Currently, breast cancer screening in Europe is delivered in organised and non-organised ways. Organised programmes include policies at national or regional levels, or pilot efforts only, and have administrative structures responsible for screening implementation, registration and evaluation. Non-organised, opportunistic screening involves referral to mammography facilities by the clinicians and/or self-referral by the women themselves (Vainio and Bianchini 2002).

The first breast cancer screening programmes were begun in 1986-1989 in the Nordic countries and in the United Kingdom. Organised screening is nowadays run in at least 18 European countries (Table 3) (Giordano et al. 1996, Shapiro et al. 1998, Ballard-Barbash et al. 1999, Dean and Pamilo 1999, Autier et al. 2002, Vainio and Bianchini 2002, Sasieni 2003, Fracheboud et al. 2004, Hofvind et al. 2004b, SOSSEG 2006b, Giordano et al. 2007, Hofvind et al. 2007a). Most countries have national or regional programmes. The commonest invitational age is 50-69 years, and the screening interval two years. Images are taken from two views and they are mostly read by two radiologists.

Table 3. Main features of organised breast cancer screening in Europe.

Screening	Country	Implementation (nationwide)	Cancer registry	Population coverage (%)	Target age (yrs)	Participation (%)	Screening interval (yrs)	Detection method	Number of views first/subsequent
Nationwide	Finland	1987 (1992)	Yes[n]	100	50-59 (69)	88	2	M	2/2
	France	1989 (2002)	Yes[r]	30	50-69 (74)	50	2	M+CBE	2/2
	Iceland	1987 (1989)	Yes[n]	100	40-69	NA	2	M+CBE	2/1
	Luxembourg	1992	Yes[n]	100	50-65	56	2	M+CBE	2/2
	Netherlands	1989 (1997)	Yes[r]	100	50-74	75	2	M	2/2
	Norway	1995 (2004)	Yes[n]	100	50-69	76	2	M	2/1
	Sweden	1986 (1997)	Yes[r]	100	40-74	70-88	1.5/2	M	2/1
	United Kingdom	1988 (1995)	Yes[n]	100	50-70	76	3	M	2/1
Regional	Belgium	1989/1992	No	25	50-69	NA	2	M+CBE	1,2/1,2
	Denmark	1991-1993	Yes[n]	20	50-69	63-71	2	M	2/1
	Ireland	1989	No	25	50-65	62	NA	M	2/1
	Italy	1985-1993	Yes[r]	76	50-69	38-62	2	M+CBE	2/1
	Portugal	1990	Yes[r]	25-50	40+	34	2	M	2/1
	Spain	1990	Yes[r]	25	45-64	85	2	M	1/1
	Switzerland	1999	Yes[r]	50	50-69	50	2	M	NA
Pilot	Greece	1989	No	25	40-64	40	2	M+CBE	2/2
	Germany	1999	No	NA	50+	NA	1	M	2/2
	Hungary	1991	No	NA	50-64	NA	1	M+CBE	2/2

M= mammography test; CBE= clinical breast examination
[n] national registration
[r] regional registration
NA= not applicable

(Giordano et al. 1996, Shapiro et al. 1998, Ballard-Barbash et al. 1999, Dean and Pamilo 1999, Autier et al. 2002, Vainio and Bianchini 2002, Sasieni 2003, Fracheboud et al. 2004, Hofvind et al. 2004b, SOSSEG 2006a, Giordano et al. 2007, Hofvind et al. 2007a)

Means of evaluation

The main outcome measure for organised breast cancer screening is mortality due to breast cancer. *Performance parameters* assess the quality of screening. As the majority of deaths during the first ten years of screening occur in women whose breast cancers were diagnosed before the introduction of screening, performance parameters provide means to predict the mortality outcome in the early years (Day et al. 1989).

Coverage of screening, namely the percentage of women in the target population that are regularly invited, is an important determinant for the mortality impact at population level (Lynge et al. 2003). *Attendance at screening* predicts the outcome among invitees. *Stage distribution of screen-detected cancers* and *rate of advanced cancers* reflect the success of screening in enabling earlier diagnosis, which should eventually result in reduced mortality from advanced disease (Day et al. 1989, Smith et al. 2004). *Proportion of ductal carcinoma in situ (DCIS)* may, however, also indicate a potential for over-diagnosis (Vainio and Bianchini 2002).

Recall rate, *rate of histological confirmation*, *positive predictive value* and *benign to malignant biopsy ratio* reflect the *screening specificity* by describing the potential of screening to identify attendees without breast cancer instead of causing harm and costs by unnecessary investigations.

Ratio of detection rate to background incidence represents the capability of screening to detect unrecognised breast cancers in the target population. Prevalence of screen-detected breast cancers and incidence of interval breast cancers between the screens provide estimates for *screening sensitivity*. For organised programmes, two measures of sensitivity have been defined (Fletcher et al. 1993). By the detection method, the sensitivity is calculated as a proportion of screen-detected

breast cancers out of the sum of screen-detected and interval breast cancers. By the incidence method, the formula for sensitivity is one minus the ratio of interval breast cancers to the expected breast cancers in the absence of screening. At test and episode levels, screen-detected and interval breast cancers are identified among the attendees of screening, and the identification is based on the mammograms (test sensitivity), or on the mammograms together with all further examinations that culminate in biopsy (episode sensitivity). Programme sensitivity identifies breast cancers among the whole target population (Hakama et al. 2007).

Randomised design is the most convenient way to prove *the impact of screening on breast cancer mortality* (Day et al. 1989). Mortality reduction observed by trials does not guarantee the effectiveness of routine screening, however, because the trials often are performed under optimum conditions (Day 2005). Variation in the screening policies and the diagnostic and therapeutic activities outside screening advocates country-specific evaluation.

Several approaches have been developed to evaluate breast cancer mortality within routine screening. Often the first step is an analysis of population level mortality trends (Hristova and Hakama 1997, Hakama et al. 1999). Cohort and case-control designs estimate the mortality impact at individual level. In cohort studies, breast cancer mortality in a group of individually followed-up screening invitees (cohort) is compared to breast cancer mortality in control population which may be non-invitees of screening in the study area, or women living in regions where screening has not yet been introduced. If the whole target population is covered by screening, breast cancer deaths within a specified cohort can be compared with modelled breast cancer mortality in a hypothetical situation where no screening took place. In case-control studies, cases represent e.g. women who have died from breast cancer. Those alive represent controls. The exposure of interest is ever having had a mammogram (dos Santos Silva 1999, Vainio and Bianchini 2002).

Developments other than screening over the evaluation period, such as changes in breast cancer incidence, improvements in therapy, increased awareness of the signs and symptoms of breast cancer, and diagnostic activity outside the organised screening affect the evaluation of breast cancer mortality, especially if the mortality is compared before and after the introduction of screening. There may also be failures to distinguish breast cancers among those who might have benefited from screening from those who have not benefited. This can be addressed by using incidence-based mortality, when only deaths from breast tumours diagnosed after the introduction of screening (or during a specified follow-up period) are considered (Hakama et al. 1997, Paci et al. 2002a, SOSSEG 2006b). In the cohort and case-control designs, the controls should be as similar as possible to the cohort members or cases with respect to the distribution of factors related to the mortality outcome except the exposure to screening (dos Santos Silva 1999). In total screening coverage, the mortality effect in cohort studies relies largely on the assumptions of models that are used to estimate breast cancer mortality in the absence of screening.

The accuracy of death certificates and official statistics has been questioned, and suggestions on other outcomes than breast cancer mortality have been introduced. All cause mortality as an endpoint would avoid the problems of precise determination of cause of death and the possibility that the death was caused by the treatment for cancer (Olsen and Gotzsche 2001, Black et al. 2002). Due to increase in random error, this endpoint requires large populations, however (Vainio and Bianchini 2002). Another method, excess mortality rate, is defined as ‘death rate in the general population due to excess risk imposed by a specific disease’ (Lenner 1990). In the case of mammography, the expected rate of death from all causes among the screening invitees is subtracted from the observed rate of death from all causes among invitees with breast cancer. Similar subtraction is done among the non-invitees of screening. The excess mortality rate among the invitees with breast cancer is then compared to the excess mortality rate among the non-invitees

with breast cancer (Lenner and Jonsson 1997). This approach, however, is not applicable if screening covers the whole target population.

Estimates for performance

Estimates for screening performance vary in Europe (del Moral Aldaz et al. 1994, Lynge et al. 2003, Broeders et al. 2005). Due to differences in protocol (registration, coverage, screening interval, invitational age, screening technique) and inconsistence in identification and determination of the validity parameters (background incidence, interval cancers, screening sensitivity), comparison of performance indicators between and within the European countries is complex (Vainio and Bianchini 2002, Yankaskas et al. 2004, Bulliard et al. 2006).

Table 4 presents the reported parameters for the screening process and validity from the Two-County trial from Sweden, the desirable levels for the screening performance by the European Union, and the reported performance indicators by selected sources from the service screening programmes in Europe (www.cancerscreening.nhs.uk, Tabar and Gad 1981, Tabar et al. 1985, Tabar et al. 1987, Day et al. 1989, Duffy et al. 1991, Lidbrink et al. 1994, Thurfjell and Lindgren 1994, Woodman et al. 1995, Lenner and Jonsson 1997, Fracheboud et al. 1998, Dean and Pamilo 1999, Fracheboud et al. 1999, Blanks et al. 2000b, Fracheboud et al. 2001, Paci et al. 2002b, Vejborg et al. 2002, Njor et al. 2003, Sasieni 2003, Broeders et al. 2005, Perry et al. 2006, SOSSEG 2006b, von Euler-Chelpin et al. 2006, Giordano et al. 2007, Giorgi et al. 2007, Hofvind et al. 2007a).

Table 4. Performance parameters of the Two-County trial, desirable levels for the screening performance by the European Union, and reported performance parameters of population-based screening programmes in Europe.

	TWO-COUNTY TRIAL Kopparberg, Östergötland	EUROPEAN UNION Desirable level	SWEDEN	NETHERLANDS	DENMARK	NORWAY	UNITED KINGDOM*	ITALY	FINLAND
Population coverage **Target age**	40-74 at entry		100% [b] 40-74¤	100% [b] 50-74¤	20% [a] 50-69¤	100% [b] 50-69¤	68-70% [a] 50-70¤	52% [a] 50-69¤	100% [b] 50-60¤
First screen									
Attendance	89%	**>75%**	69-93%	75%	63-84%	76.2% [1]	71-77%	38-62% [1]	89% [1]
Recall rate	4.8% [2]	**< 5%**	1.7-5.2%	1.3%	2.8-6.8%	4.6%	7.0-8.5%	8.4%	3.3% [1]
Detection rate/background	3.09-4.59 x IR [9]	**> 3 x IR**		2.95 x IR	4-4.4 x IR	3.0 x IR			2.4 x IR [1]
DCIS	9.2% [1]	**10-20%**	11-13% [3]	14%	12-14%	17.6%	20%	11.3%	11% [1]
Stage II+	33.3-34.3% [9]	**<30%**	33% [3]	20%	30-35%				28% [1]
Node negative	78.6% [8]	**>70%**	79-80% [3]	67%	72-81%	74.1%	67-75% [1]	73% [2]	
B:M ratio		**<1:4**	1.6-3.6:4 [3]	2:4	1.6-3.6:4	7.4:4	1.5:4	1.3:4	3.0:4 [1]
Episode sensitivity **(0-11 months)** **(12-23 months)**	 87% [7] 71% [7]	 **>70%** **>50%**		 73% 48%	 81% 32%	 72% 36%	 69% [6] 48% [6]		
Subsequent screens									
Attendance	84%	**>75%**	70-88%	>90%	63-90%	76.2% [1]	84-90%	38-62% [1]	89% [1]
Recall rate		**<3%**	1.9-5.7%	0.7%	1.3-3.2%	2.6%	3.1-3.8%	4.2%	3.3% [1]
Detection rate/ background		**> 1.5 x IR**		1.44 x IR	1.9-2.0 x IR	2.3 x IR			2.4 x IR [1]
DCIS	9.2% [1]	**10-20%**	15% [5]	14%	11%	17.2%	16%	13.7%	11% [1]
Stage II+	16.9-25.0% [9]	**<25%**	18% [5]	17%	27-34%				28% [1]
Node negative	83.9% [8]	**>75%**	83% [5]	71%	72-81%	75.2%	67-75% [1]	64% [4]	
B:M ratio		**<1:4**	3.2:4 [5]	1.2:4	0.4-1.2:4	3.8:4	0.5:4	0.8:4	3.0:4 [1]
Episode sensitivity **(0-11 months)** **(12-23 months)**	 87% [7] 71% [7]	 **>70%** **>50%**		 74% 45%	 68% 39%	 72% 34%	 69% [6] 48% [6]		

DCIS= Proportion of screen-detected breast cancers that are ductal carcinoma in situ; Stage II+= Proportion of screen-detected breast cancers that are stage II+; Node negative= Proportion of screen-detected invasive breast cancers that are node negative; B:M ratio= benign to malignant biopsy ratio; IR =background incidence rate

* Three-year screening interval; [a] Actual coverage; [b] Theoretical coverage; ¤ in 2007

[1] Screens not specified; [2] Kopparberg county only; [3] First screening round; [4] Subsequent screening round; [5] Subsequent screening round, target age 40-69; [6] Screens not specified, target age 50-64; [7] Screens not specified, target age 50-69; [8] Target age 40-69; [9] Target age 50-69

Two-county trial (Tabar and Gad 1981, Tabar et al. 1985, Tabar et al. 1987, Day et al. 1989, Duffy et al. 1991); European Union (Perry et al. 2006); Sweden (Lidbrink et al. 1994, Thurfjel l and Lindgren 1994, Lenner and Jonsson 1997, SOSSEG 2006a); Denmark (Vejborg et al. 2002, Njor et al. 2003, von Euler-Chelpin et al. 2006); Norway (Hofvind et al. 2007a); Netherlands (Fracheboud et al. 1998, Fracheboud et al. 1999, Fracheboud et al. 2001); United Kingdom (Woodman et al. 1995, Blanks et al. 2000b, Sasieni 2003, Broeders et al. 2005, www.cancerscreening.nhs.uk); Italy (Paci et al. 2002b, Giordano et al. 2007, Giorgi et al. 2007); Finland (Dean and Pamilo 1999)

Sweden was the pioneer country of randomised controlled trials in Europe and started service screening in 1986-1996. The first papers on performance of service screening reported results by screening round (Lidbrink et al. 1994, Thurfjell and Lindgren 1994, Lenner and Jonsson 1997). Recently, process parameters by the first and the subsequent screens in 1980-2001 were reported from 13 areas covering 45% of the Swedish women by the Swedish Organised Service Screening Evaluation Group (SOSSEG) (SOSSEG 2006b). Attendance among 40-69 year old invitees was 70-93% at the first screen, and 70-88% at the subsequent screens. Recall rates varied between 1.7-5.2% and 1.9-3.2%, and breast cancer detection rates per hundred women screened from 0.41 to 0.77 and from 0.36 to 0.60, respectively.

The latest report on service screening in Sweden compared stage at presentation in the screening period to that in the pre-screening period (SOSSEG 2007). A 15% reduction in cancers that had spread to the surrounding lymph nodes (0.85, 0.79-0.92), a 30% reduction in tumours of size >2cm (0.70, 0.65-0.76), and a 20% reduction in stage II+ tumours (0.80, 0.75-0.86) was observed among 40-69 years old invitees. Incidence of total invasive cancers in Sweden had been analysed earlier in 11 counties where service screening had been initiated in 1986-1990 (Jonsson et al. 2005). In this study, a substantial increase in breast cancer incidence persisted more than six years after the first screening round. The estimated relative risks adjusted for year, region, age and lead time in ages 40-49, 50-59, 60-69, and 70-79 were 0.96 (0.77-1.21), 1.54 (1.33-1.79), 1.21 (1.04-1.41) and 1.03 (0.82-1.30), respectively. The areas of this study and the study of SOSSEG (2007) overlapped. Differences in the modelling methods and underlying assumptions probably explain the contradicting findings of these two studies.

The episode sensitivity within service screening in Sweden has been reported from Stockholm, and could also be calculated for Östergötland area after the randomisation period. In Stockholm, the

invitational age was 50-69 years and the sensitivity by the detection method 68-79% at the first, and 67% at the subsequent screen in 1989-1993 (Tornberg et al. 2005). In Östergötland, the overall sensitivity by the detection method in 1987-1992 was 57% for women aged 50-59, and 72% for women aged 60-69 (Vitak et al. 1997).

The nationwide breast cancer screening programme in the Netherlands was launched in 1988-1989. At the end of 1997 the programme covered the whole country (Otto et al. 2003). Several studies have assessed the performance of the Dutch programme (Fracheboud et al. 1998, Fracheboud et al. 1999, Fracheboud et al. 2001, Fracheboud et al. 2004). In 1990-1997, over 75% of women invited attended screening and more than 90% of the attendees re-attended in the following round (Fracheboud et al. 2004). The rate of referral was low, 1.3% at the first screen, and 0.7% at the subsequent screens. The corresponding detection rates for the breast cancers per hundred women screened were 0.61 and 0.36 (Fracheboud et al. 2001). The impact of screening on the incidence rates of breast cancer was studied in seven regions where no screening took place before 1990 (Fracheboud et al. 2004). The overall breast cancer incidence rose markedly, but the increase was mainly due to tumours of size <2cm and ductal carcinoma in situ. In women aged 50-69 (target group of mammography screening), the rates of advanced cancers showed a decline of 12% (estimated annual percentage of change -2.14, 95% CI -3.47, -0.80). Interval breast cancers in the Dutch programme were estimated in 1990-1993 by the incidence method (Fracheboud et al. 1999). In the first year after the first and subsequent screens the interval tumours amounted to 26-27% of the underlying incidence, in the second year the proportion was 52-55%.

The organised mammography screening in Denmark has covered Copenhagen and Fyn, approximately one fifth of the Danish women aged 50-69. The overall participation of the programmes was reported recently (von Euler-Chelpin et al. 2006). The attendance at the first round

was 71%, at the fourth round it had decreased to 63%. The process indicators in Copenhagen were reported by two studies (Vejborg et al. 2002, Njor et al. 2003). In 1991-1999, approximately 70% of invitees participated in the first round, and 90% of them re-attended in the following round. Cancer detection rates per hundred women screened were 1.19 at the first screen, and 0.54-0.63 at the subsequent screens. In 1991-1997, the recall rates were 6.8% at the first screen, and 3.2% at the subsequent screens. In Fyn, the recall rates were lower than those in Copenhagen, 2.8% and 1.3% (Njor et al. 2003). The episode sensitivity by the incidence method was 68-81% in the first year of follow-up and 32-39% in the second year of follow-up after screening.

The Norwegian breast cancer screening programme started as a four-year pilot in 1995-1996, and covered the whole country in 2004 (Hofvind et al. 2004b). The process and validity parameters among the 50-69 years old invitees have been reported intensively (Wang et al. 2001a, Wang et al. 2001b, Hofvind et al. 2004a, Hofvind et al. 2004b, Hofvind et al. 2006a, Hofvind et al. 2007a, Hofvind et al. 2007b). The overall attendance rate during the first ten years of the programme was 76%, the recall rates in the first and subsequent screens 4.6% and 2.6%, and the detection rates for the screen-detected cancers 0.64% and 0.49%, respectively. Axillary lymph node metastases were present in 25% of screen-detected breast cancers. The episode sensitivity by the incidence method within the first year after mammography screening was 72%, and within the second year 34-36% (Hofvind et al. 2006a, Hofvind et al. 2007a). The impact of screening on the incidence and tumour characteristics of breast cancer was studied in 1996-2004 (Hofvind et al. 2007b). Breast cancers diagnosed in the screening period had more favourable characteristics than those in the pre-screening period, but the decline in advanced cancers after the first years of screening was not clearly apparent.

Organised breast cancer screening in the United Kingdom (UK) was introduced in April 1987, and achieved national coverage in 1995. Women are invited every three years. Only few papers have been published on the screening performance in the UK (Sasieni 2003). The first parameters for the three-year period 1990-1993 were reported in 1995 (Moss et al. 1995). Results from 1994-1999 were announced to examine the extent to which interim targets had been achieved (Blanks et al. 2000b). According to this report, the attendance at the first screen among the 50-69 years old invitees varied between 75 and 77%, and approximately 90% of the participants re-attended after the following invitation. The corresponding recall rates were 7.0-8.3% and 3.1-3.8%, and the detection rates 0.46-0.63% and 0.38-0.52%. In the North West of England, the rates of advanced disease among attendees, lapsed attendees and non-attendees aged 54 years or younger were compared between two areas, Manchester and Wigan in 1988-2000 (Threlfall et al. 2003). According to the authors, the significantly lower rates of advanced disease among the invitees in Wigan compared to those in Manchester (2.49 vs. 4.73 per 10 000 person years) were due to higher cancer detection rates and more regular attendance in Wigan.

The occurrence of invasive interval cancers in the UK was examined by the incidence method in the North-West region in 138 000 women with negative screening assessment (Woodman et al. 1995). The rate of detection of interval cancers expressed as a proportion of the underlying incidence was 31% during the first, 52% during the second, and 82% during the third year after screen.

Since the beginning of the 1990's, the Italian Group of Mammography Screening (GISMa) has published an annual survey of organised screening in Italy (Giordano et al. 2007, Giorgi et al. 2007). The coverage of breast cancer screening among women aged 50-69 increased from 5% to 76% in the period 1992-2005. In 2005, ten counties had total screening coverage, seven counties partial coverage. Three counties did not invite women to organised mammography screening. The

overall, crude attendance rates varied from 38% to 62%. The referral rates at the first and at the subsequent screens were 8.4% and 4.2%, and the detection rates for screen-detected cancers per hundred women screened 0.67 and 0.51.

In Finland, one study has reported overall performance for the population-based breast cancer screening programme in 1987-1997 for women aged 50-59 at first invitation (Dean and Pamilo 1999). The data were based on information of the Finnish Radiological Society, and reported 89% attendance, and 3.3% recall rate. The breast cancer detection rate per hundred women screened was 0.36. The proportion of stage II+ cancers of all screen-detected breast tumours was 28%, and the proportion of carcinoma in situ 11%. The screening sensitivity in Finland was evaluated in the city of Turku in 1987-1996, where the test sensitivity by the detection method was 76% among women aged 50-74 (Klemi et al. 1998).

Estimates for mortality

Development of breast cancer mortality within the service screening has been evaluated intensively in Sweden (Table 5). The first studies found a range of estimated mortality reduction of 9% to 28% depending on the age group and the region studied (Tornberg et al. 1994, Lenner and Jonsson 1997, Jonsson et al. 2000, Jonsson et al. 2001, Jonsson et al. 2003a, Jonsson et al. 2003b). Törnberg and colleagues (1994) compared the average breast cancer mortality in regions with mammography screening to regions without screening in 1971-1990 in all the 26 Swedish counties. Lenner and Jonsson (1997) used excess mortality to analyse the screening affect among 40-74 years old invitees in two Swedish counties that started mammography screening in 1990. The relative risk estimates were based on the cumulative excess mortality associated with breast cancer in 1990-1995 in the

screened population compared to that in the control population (women from two counties that initiated screening after 1995).

Jonsson and colleagues investigated the effect of screening on breast cancer mortality among women aged 40-49 (screening period 1986-1996; reference period 1976-1986), 50-69 (1986-1997; 1979-1990), and 70-74 years (1986-1998; 1976-1988) using various analysis techniques and by comparing counties introducing screening early to counties starting screening later (Jonsson et al. 2000, Jonsson et al. 2001, Jonsson et al. 2003a). Jonsson and colleagues also reported long-term effects on breast cancer mortality from 1977-1998 in a pilot screening area, Gävleborg, where the study cohort included women aged 40-64 years (Jonsson et al. 2003b). Women of the same age in the neighbouring counties and Sweden as a whole served as controls. Breast cancers diagnosed within the first ten years after the introduction of screening were included in the analysis.

Duffy and colleagues compared incidence-based breast cancer deaths in the screening period to deaths in the pre-screening period in seven Swedish counties in women aged 50-69 (Duffy et al. 2002). In counties with more than 10 years of screening, a 32% decrease in breast cancer mortality was reported. A 18% decrease was estimated in counties with 10 years or less of screening. In 2003, a study with 20 years of follow-up in Östergötland and Kopparberg counties was published (Tabar et al. 2003). The deaths from breast cancer in the screening period 1978-1997 were compared to those in the pre-screening period 1958-1977. After adjustment for age, self-selection, and changes in incidence, the reduction in breast cancer mortality among 40-69 year old screened women was 44%, and among screened and non-screened women combined 41%.

The latest papers on breast cancer mortality within routine screening in Sweden were published in 2006 and 2007 (SOSSEG 2006a, SOSSEG 2006b, Jonsson et al. 2007). The two studies of the

Swedish Organised Service Screening Evaluation Group (SOSSEG) used incidence-based deaths from breast cancer and all breast cancer deaths as outcome measures, and compared mortality in the screening period to that in the pre-screening period in 13 counties. The estimates were adjusted for self-selection, contemporaneous changes in incidence, and changes in mortality independent of screening. A 43% decrease in incidence-based breast cancer mortality among 40-69 year old screening participants was reported. Among the invitees, the corresponding reduction was 27% (SOSSEG 2006b). In the companion article (SOSSEG 2006a), a reduction of 39% in total breast cancer mortality was observed among participants of screening. In the study of Jonsson and colleagues (2007), the follow-up of women from the previous study of Jonsson and Lenner (1997) was extended by six years. The estimated reduction in breast cancer mortality among invitees compared to those not yet invited was 30%.

In Denmark, a Poisson regression with a study group, historical control group, national control group, and historical national control group was used to estimate the effect of organised screening on incidence-based breast cancer mortality in the Copenhagen area (Olsen et al. 2005a, Olsen et al. 2005b) (Table 5). The study group included 50-69 year old women invited to screening during the first five invitation rounds from 1 April 1991 to 31 March 2001. Pseudo-invitation rounds of the same length as the invitation rounds of the study group were constructed for the three control groups and pseudo-invitation dates were allocated. The historical control group consisted of women living in Copenhagen before the start of the programme. The five pseudo-invitation rounds for the historical control group started on 31 December 1979 and ended on 30 December 1989. The national control group consisted of women living outside the areas offering screening. Their pseudo-invitation rounds were constructed for the same period as the invitation rounds for the study group. The historical-national control groups consisted of women living outside the areas offering screening in the period preceding the programme. Their pseudo-invitation rounds were the same as

those of the historical control group. The incidence-based breast cancer mortality in the study group was compared to that in the three control groups, adjusting for age, period, and region. The effect among participants was adjusted for self-selection. Breast cancer mortality in the screening period was reduced by 25%. Among the participants, the adjusted reduction was 37%.

In the Netherlands, breast cancer mortality trends in the screening period were compared to those in the pre-screening period first in Nijmegen and later in the whole country (Broeders et al. 2001, Otto et al. 2003) (Table 5). The pilot study on screening was started in Nijmegen in 1975 with 30 000 invitations to women aged 35 or older. Women aged 65-69 years were invited after the second round. The breast cancer mortality rates in Nijmegen were compared to those in Arnheim (where screening was started in 1989, reaching full coverage in 1990) and to the Netherlands as a whole (where almost full coverage of screening was achieved in 1997) in 1975-1997. The age-standardised breast cancer mortality rates and ratios showed inconclusive patterns with regard to the expected impact of screening. The period-cohort group analysis indicated a non-significant 6-16% reduction of breast cancer mortality in Nijmegen compared to mortality in Arnheim and in the country as a whole. The authors mentioned that the attendance rates in Nijmegen were relatively low (60% after 9 rounds), and the annual number of breast cancer deaths small with considerable fluctuation. They concluded that the earlier detection of possibly less aggressive tumours, and more effective treatment might have reduced the mortality gap. In addition, the small number of screening invitees in Nijmegen made it difficult to obtain stable estimates of breast cancer mortality reduction resulting from early detection by mammography.

Otto and colleagues (2003) compared the observed mortality rates in women aged 55-74 years with the pre-screening rates in the Netherlands as a whole. The researchers also compared by the MIcrosimulation SCreening ANalysis (MISCAN) observed rates from 1986 onwards in the age

groups 45-54, 55-64, 65-74, and 75-84 years with the breast cancer mortality rates predicted for theoretical areas in which the screening programme did and did not exist. The annual percentage increase in breast cancer mortality among women aged 55-74 in the pre-screening period was 0.32% (0%-0.65%), and the corresponding annual decrease in the screening period 1.67% (2.39%, 0.96%). Compared to the pre-screening era of 1986-1988, the overall reduction in breast cancer mortality among women aged 55-74 reached 20% in 2001.

The change in breast cancer mortality due to organised screening in the United Kingdom has been reported by three papers (Blanks et al. 2000a, McCann et al. 2001, Fielder et al. 2004) (Table 5). Blanks and colleagues estimated the effect of population-based screening on breast cancer mortality in England and Wales. They fitted a log-linear age-cohort model to the annual number of breast cancer deaths in five-year age groups in 1971-1989 and extrapolated the fit to the years 1990-1999. The total reduction in breast cancer mortality in 1998 in women aged 55-69 was 21.3%, but direct effect of screening on breast cancer mortality was estimated as 6.4% (5.4, 11.8%). The combination of changing incidence rates, down-staging unrelated to screening *per se*, and better treatment was estimated to reduce the breast cancer mortality by 14.9%. The researchers concluded that many deaths from breast cancer in the 1990s could be attributed to cancers that had been diagnosed before invitation to screening.

McCann and colleagues (2001) estimated the achievable mortality reduction of the East-Anglian mammography programme by comparing cancer prognosis among screening invitees and those not yet invited. The combined predicted reduction in breast cancer mortality by 2004 among the invited women compared with the non-invited women was 15%. In Wales, a case-control study with 1:2 matching was established to estimate the impact of screening on breast cancer mortality (Fielder et al. 2004). The cases were deaths from breast cancer in women aged 50-75 years at diagnosis who

were diagnosed after the start of screening in 1991, and who died after 1998. The women alive served as controls. Based on 419 cases, the odds ratio for the risk of death from breast cancer for women who had attended at least once compared to those never screened was 0.62 (0.47-0.82). After excluding cases diagnosed prior to 1995 (when prevalence screen cases dominated), and adjusting for self-selection bias, the estimated, non-significant, reduction was 25%.

In Italy, women in Florence aged 50-69 with breast cancers diagnosed in 1990-1996 were divided into invitees and non-invitees to mammography screening prior to the date of diagnosis. The incidence-based breast cancer mortality rates for pre-screening (1985-1986) and screening (1990-1999) periods were used to compare the breast cancer mortality between invitees and non-invitees in the screening and pre-screening era (Paci et al. 2002a) (Table 5). The rates of incidence-based breast cancer mortality were 25% lower for invited compared to non-invited women. The authors estimated, however, that only one third of the reduction was directly due to screening. In another study, Paci and colleagues studied incidence-based mortality among invitees in Florence in 1990-1999 by using breast cancer-specific case-fatality rates to formulate the expected number of breast cancer deaths in the screening period (Paci et al. 2002b) (Table 5). The cumulative number of expected deaths from incident breast cancer cases was then compared with the observed deaths. The ratio between the observed and expected breast cancer deaths was 0.81. Due to low invitational coverage (35%), the mortality reduction in the whole female population was estimated at 3.2%.

The effectiveness of service mammography in Finland was studied in the implementation phase of the programme (Hakama et al. 1997), and later with two designs in three cities (Table 5). The results from the implementation phase are reported elsewhere (see Table 2). The second Finnish study concentrated on the city of Helsinki and compared breast cancer mortality between the invited (born 1935-9) and the non-invited (born 1930-4) birth cohorts aged 50-59 at entry (Anttila et al.

2002). The screening effect was controlled for treatment or other background changes by adjusting the difference in the incidence-based breast cancer mortality between invitees and non-invitees in the screening era for the corresponding difference in the pre-screening era at ages 40-49, and for age at death. There was a 19%, non-significant decrease in the incidence-based mortality in the invited birth cohorts compared to the mortality in the non-invited cohorts.

The third study from Finland assessed breast cancer mortality in three cities, Helsinki, Tampere and Turku, which all employed different screening policies (Parvinen et al. 2006). Incidence-based breast cancer deaths in the screening period 1987-1997 among women aged 55-69 in 1987 (born 1928-32) were compared to incidence-based breast cancer deaths in the pre-screening period 1976-86 among women born 1918-22. In Turku, where all the birth cohorts studied were invited to screening in 1987-97, a significant 36% reduction in breast cancer mortality was observed. In Tampere, where only some of the birth cohorts (1928-32) were invited, the reduction in breast cancer mortality was non-significant, 14%. In Helsinki with no invitations to screening for the birth cohorts studied, no reduction in breast cancer mortality was observed. The breast cancer mortality in Turku was more than 40% lower than that in Helsinki (0.58, 0.41-0.83).

Table 5 summarises the area, design, outcome, study periods, age range and effect and efficacy estimates (with 95% CI) for the screening invitees and attendees within organised breast cancer screening in Sweden, Denmark, the Netherlands, the United Kingdom, Italy and Finland.

Table 5. Effect estimates of organised breast cancer screening in Europe by selected publications.

COUNTRY, REFERENCE	AREA	DESIGN	OUTCOME	COMPARISON
SWEDEN				
Törnberg (1994)	Whole country	Geographic	Bcm	Bcm in regions with and without screening
Lenner, Jonsson (1997)	Northern region	Geographic	Excess bcm	Bcm in counties with and without screening
Jonsson (2007)	Northern region	Geographic	Excess bcm	Bcm in counties with and without screening
Jonsson (2000)	14 areas in 14 counties	Geographic	Incidence-based bcm	Bcm in counties starting at the age of 40 to bcm in counties starting at the age of 50
Jonsson (2001)	7 areas in 7 counties	Geographic	Incidence-based bcm	Bcm in counties starting 1986-87 to bcm in counties starting 1993 or later
Jonsson (2003a)	16 areas in 16 counties	Geographic	Incidence-based bcm	Bcm in counties screening until age 74 to bcm in counties screening until age 69
Jonsson (2003b)	Gävleborg county	Geographic	Incidence-based bcm	Bcm in pilot programme to bcm in counties starting later and to Sweden as a whole
Duffy (2002)	7 counties	Before-after	Incidence-based bcm	Bcm in screening and pre-screening era
Tabar (2003)	2 counties	Before-after	Incidence-based bcm	Bcm in screening and pre-screening era
SOSSEG (2006b)	13 areas in 8 counties	Before-after	Incidence-based bcm	Bcm in screening and pre-screening era
SOSSEG (2006a)	13 areas in 8 counties	Before-after	Bcm	Bcm in screening and pre-screening era
DENMARK				
Olsen (2006)	City of Copenhagen	Individual cohort with three control groups[1]	Incidence-based bcm	Bcd among invitees and participants to bcd in three control groups
NETHERLANDS				
Broeders (2001)	City of Nijmegen	Geographic	Bcm	Bcm in Nijmegen to bcm in Arnheim and in whole country
Otto (2003)	Whole country	Before-after, model	Bcm	Trends in bcm between screening and pre-screening era
UNITED KINGDOM				
Blanks (2000a)	England and Wales	Model	Bcm	Bcm from pre-screening era extrapolated to screening era
McCann (2001)	East-Anglia	Model	Bcm	Bcm among invitees to bcm among non-invitees, 2 yrs lead-time
Fielder (2004)	Wales	Case-control 1:2	Bcm	Bcm among ever screened to bcm among never screened
ITALY				
Paci (2002a)	City of Florence	Individual cohort	Incidence-based bcm	Bcm between invitees and non-invitees, screened and non-screened
Paci (2002b)	City of Florence	Model	Incidence-based bcm	Bcd among invitees to modelled bcd in the absence of screening
FINLAND				
Anttila (2001)	City of Helsinki	Model	Incidence-based bcm	Modelled bcd between invitees and non-invitees
Parvinen (2006)	City of Turku	Before-after	Incidence-based bcm	Bcd among invited cohorts in pre-screening and screening era

Bcm= breast cancer mortality; Bcd= breast cancer death; CI= confidence interval
[1]Historical, national and historical national control groups

Table 5 (contd)

COUNTRY, REFERENCE	STUDY PERIOD, REFERENCE PERIOD	AGE AT ENTRY, AGE AT DEATH	EFFECT ESTIMATE[1] (95% CI)	EFFICACY ESTIMATE[2] (95% CI)
SWEDEN				
Törnberg (1994)	1971-90, -	50-74, 50-74	0.81 (0.63-1.01)	-
Lenner, Jonsson (1997)	1990-95, -	40-74, not specified	0.72 (0.53-0.99)	-
Jonsson (2007)	1988-2002, -	40-74, not specified	0.70 (0.56-0.87)	0.65 (0.51-0.84)
Jonsson (2000)	1986-96, 1976-86	40-49, not specified	0.91 (0.72-1.15)	-
Jonsson (2001)	1986-97, 1979-97	50-69, 50-79	0.84 (0.67-1.05)	-
Jonsson (2003a)	1986-98, 1976-88	65-74, 70-84	0.84 (0.54-1.19)	-
Jonsson (2003b)	1977-98, -	40-64, not specified	0.84 (0.71-1.00)	-
Duffy (2002)	5-20 yrs, 5-20 yrs	40-69, not specified	0.68 (0.60-0.77)[a] 0.82 (0.72-0.94)[b]	0.61 (0.55-0.68)
Tabar (2003)	1978-97, 1958-77	40-69, not specified	0.59 (0.53-0.66)	0.56 (0.49-0.64)
SOSSEG (2006b)	11-22 yrs, 11-22 yrs	40-69, not specified	0.73 (0.69-0.77)	0.57 (0.53-0.62)
SOSSEG (2006a)	11-22 yrs, 11-22 yrs	40-69, not specified	-	0.61 (0.55-0.68)
DENMARK				
Olsen (2006)	1991-2001, 1979-89	50-69, 50-79	0.75 (0.63-0.89)	0.63
NETHERLANDS				
Broeders (2001)	1971-95, -	50-79, 50-79	0.84 (0.61-1.17)	-
Otto (2003)	1989-2001, 1980-88	55-74, 55-74	0.80 in 2001	-
UNITED KINGDOM				
Blanks (2000a)	1990-99, 1971-89	55-69, 55-69	0.79	-
McCann (2001)	1989-96, -	50-64, 50-64	0.85 (0.78-0.93)	-
Fielder (2004)	1991-2000, -	50-75, not specified	-	0.75 (0.49-1.14)
ITALY				
Paci (2002a)	1990-99, 1985-92	50-69, 50-69	0.75 (0.54-1.04)	0.63 (0.42-0.94)
Paci (2002b)	1990-99, 1988-89+1990-92	50-69, 50-74	0.81 (0.64-1.01)	-
FINLAND				
Anttila (2001)	1986-97, 1976-86	50-59, 50-71	0.81 (0.62-1.05)	-
Parvinen (2006)	1987-97, 1976-86	55-69 in 1987, 55-80	0.64 (0.47-0.88)	-

[1] Risk ratio among invitees
[2] Risk ratio among attendees
[a] Counties with more than ten years of screening; [b] Counties with ten years or less than ten years of screening

3. AIMS OF THE STUDY

The aim of this study was to evaluate performance and effectiveness of population-based breast cancer screening using individual data from organised screening in Finland.

The specific aims were to estimate:

1. Process and validity indicators of screening
2. Breast cancer mortality in relation to invitation, participation and target age of screening
3. Associations between process, validity and outcome of screening

4. MATERIAL AND METHODS

4.1 Finnish breast cancer screening programme

Establishment and data registration

The Finnish population-based mammography programme was started gradually by a group-randomised design (Hakama et al. 1997). The first women screened in 1987 and 1988 were those born in 1928, 1930, 1932, 1934, and 1938. Women born in 1927, 1929, 1933, and 1939 were controls. The randomised implementation period ended at the beginning of 1992.

The Bylaw on Public Health from 1992 required Finnish municipalities to offer mammography screening every second year for women aged 50-59 years. Screenings for women aged 60-69 years remained optional. The organised mammography was carried out in several municipal or private screening centres.

The Law and Decree on personal records kept under the health care system from 1989 required the municipalities to store their data on screening process and outcome. The Mass Screening Registry of Finland established in 1968 was named as the central registration authority. The only providers sending information on invitations and visits from the beginning of service screening to the Mass Screening Registry were the centres of the Cancer Society of Finland (CSF). The invitations in the data covered 50-60% of all screening invitations in the 1990's and early 2000's in Finland.

Invitees of the Cancer Society of Finland

The ten CSF centres invited women aged 40-74 years to free-of-charge mammography. The coverage of invitations was more than 95% among women aged 50-59 years, 20-40% among women aged 60-69 years, and less than 10% among 40-49 and 70-74 year old women. Some centres also invited women for screening with a fee, i.e. at their own expense. These invitations were mostly targeted at women aged 60-69 years, and were not registered in the Mass Screening Registry.

Women received a personal letter of invitation. The test was a two-view screen-film mammography (craniocaudal and mediolateral oblique), which two radiologists interpreted independently. If either interpretation was positive, simultaneous reading was performed. The recall examinations included one or several of the following: breast palpation, repeat mammography, breast ultrasound, galactography, pneumocystography, fine-needle biopsy, and core-needle biopsy. If these examinations could not exclude breast cancer, surgery was performed.

In the Mass Screening Registry, the individual information on screening invitations and visits were augmented with data on incidence and mortality from cancer from the Finnish Cancer Registry, and data on migration and lifespan from the National Population Registry using a personal identifier (social security number) as a key.

4.2 Formulation of data and study parameters

Screening process

The subjects were 40-74 year old women invited to the centres of the CSF in 1991-2000 (Studies I and II). The invitation and visit (attendance) files from 1987-1990 were used to define the correct screening rounds and number of screens for each woman. The data were divided into three groups: first round-first screen, subsequent round-first screen and subsequent round-subsequent screen. Since the data on screening invitations in the CSF centres were available from 1987, but the data on screening visits before 1991 were incomplete, the group 'subsequent round-first screen' (n=18 825) included women who had been screened before their first registered visit. This group was excluded from the analysis.

Tumours were classified by stage. In situ cancer represented stage 0. Tumours less than 2 cm in diameter without regional or distant lymph node metastasis represented stage I. Tumours more than 2 cm in diameter with or without regional lymph node and/or distant metastasis represented stage II+. Approximately 93% of carcinomas had a pathological classification for primary tumour, regional lymph nodes, and distant metastasis (pTNM).

The process indicators studied were *coverage* of and *attendance* at screening, *rate of recall*, *rate of histological confirmation*, *rate of screen-detected cancers* (detection rate), *ratio of detection rate to background incidence*, *positive predictive value of mammography*, *benign-to-malignant biopsy ratio* and *stage-distribution of screen-detected cancers*. The indicators were calculated by five-year age groups, invitational year and screening centres separately for the first and for the subsequent screens.

Screening validity

The subjects were 50-64 year old women, who attended screening on invitation from the CSF centres in 1991-1999 (n=153 452) (Study III). The invitation and visit files from 1987-1999 were used to define the correct screening rounds and number of screens for each woman. The group subsequent round-first screen (n=18 825) was excluded from the analysis.

Our primary validity indicator was *episode sensitivity*. Since screening is essentially a multiple-step process with an initial screening mammogram leading, if positive, to a series of more detailed investigations, episode sensitivity provided the means to demonstrate the entire chain of events connected to screening. The episode sensitivity was estimated by the detection and the incidence methods. The interval cancers were primary breast cancers diagnosed after a negative screening episode. They arose either before the next invitation to screening or within a period equal to the screening interval after a woman had reached the upper age limit for screening. The median screening interval was 24 months (Q1=23, Q3=25). The maximum time after the last negative screen was set at 27 months (24+3). Hence, the interval cancers were breast cancers that had been identified between two successive screens or within the maximum of 27 months after the last screen. To exclude the possible findings of screening with a fee, the interval cancer follow-up was stopped after 23 months among women, who were older than 58 years and had their last screening test in the data.

Based on the protocol of the Finnish Cancer Registry, breast cancers were divided into four categories: carcinoma in situ, invasive localised (tumours without lymph node or distant metastasis), invasive non-localised (tumours with lymph node or distant metastasis) and unknown.

The background trend was based on incidence of breast cancers in 1980-1986. An average 2.7% annual linear increase was assumed for all CSF centres.

The overall episode sensitivity in 1991-2001, and the sensitivity within the first and the second year following the screening visit were estimated. The rate of screen-detected and interval cancers and the overall episode sensitivity were calculated by three-year screening periods, five-year age groups, and three centre categories separately for the first and the subsequent screens. The categorisation of screening centres to low, intermediate and high recall rates reflected screening specificity, and was based on recall rates and visits per each centre at the subsequent screens. In the category 'low recall rate' (three centres) the range in recall rates was 0.9-1.9%, in the category 'intermediate recall rate' (four centres) 2.3-2.7%, and in the category 'high recall rate' (three centres) 2.8-3.5%.

Screening effectiveness

The overall effectiveness (of invitation) and *efficacy* (of participation) were analysed by comparing observed breast cancer deaths with expected breast cancer deaths without screening in 1992-2003 (Study IV). The main outcome was incidence-based mortality from breast cancer in ages 50-69. The mortality estimates were calculated by age at death and screening centre categories. The categorisation of centres was done similarly as in the validity study.

The observed deaths were obtained from a cohort of women aged 50 years or more, who had been invited at least once to any of the ten centres of the CSF in the study period (n=361 848). To achieve comparability between the individual and the population data, we set the individual cohort entry date at January 1st in the first invitation year for 1992-2003. The follow-up was finished at

death, emigration or December 31st 2003. Women responding positively to their first invitation in 1992-2003 were classified as participants, the other invitees were non-participants.

Since no non-invited controls were available, the expected breast cancer mortality rates without screening were estimated by modelling, using population data from 1974-1985 and 1992-2003. The population data consisted of age-specific numbers of women and breast cancer deaths by calendar year and by municipality, and were derived from the same municipalities (n=260) that were incorporated into the cohort.

The analysis on *effectiveness* and *efficacy* was extended to ages 60-79 at death by comparing observed breast cancer deaths with expected breast cancer deaths without screening *in varying invitational policies* in 1992-2003 (Study V). The data from 260 municipalities were divided into three categories: regular invitations at ages 50-59 only, regular invitations at ages 50-69, and regular invitations at ages 50-59 with irregular invitations at ages 60-69. Observed deaths from breast cancer were compared with expected breast cancer deaths without screening among all women and among screened and non-screened women. The group 'screened' included the participants of screening, whereas the group 'non-screened' contained the non-participants and the non-invitees of screening.

For all women, the person-years and the observed deaths were derived from population level data. For the screened women, they were obtained from an individual cohort. The cohort follow-up was organised similarly as in the study on overall effectiveness. The person-years and the observed deaths for the non-screened women were obtained by subtracting the figures of the screened women from those of the whole female population. The expected mortality rates without screening were estimated by modelling, using population data from 1974-1985 and 1992-2003.

4.3 Statistical analyses

Screening process and validity

Logistic regression analysis was performed for the detection rates of screen-detected cancers, Poisson regression for the positive predictive values. The estimates were adjusted by five-year age group (40-74), birth cohort, calendar year (1991-2000), number of screens (first, subsequent), and screening centre.

Poisson regression analysis was performed for the validity estimates of screening. The rate of screen-detected and interval cancers and the episode sensitivity were adjusted by five-year age group (50-64), three-year calendar period (1991-93, 1994-96, 1997-99) and centre category (high recall rate 2.8-3.5%, intermediate recall rate 2.3-2.7%, low recall rate 0.9-1.9%).

Screening effectiveness

The expected breast cancer mortality rates without screening were modelled by Poisson regression. For the overall effectiveness, five-year age groups at death (50-54, 55-59, 60-64, 65-69), three centre categories (low, intermediate, high recall rates), calendar year within the two periods (1, …, 12), and period before (1974-1985) and with screening (1992-2003) by five-year age groups at death were used as explanatory variables. Model 1 is described in detail in the Appendix.

For the comparison of different invitational policies within the whole female population and between the screened and non-screened women, two models were formulated. Five-year age

groups at death (60-64, 65-69, 70-74, 75-79), policy categories (regular 50-59, regular 50-69, irregular 50-69), calendar year within the two periods (1, …, 12) were used as explanatory variables in both models. In the analysis of whole female population (model 2), the period before (1974-1985) and with screening (1992-2003) by ten-year age groups at death was included. For the screened and non-screened women (model 3), the period before (1974-1985) and with screening (1992-2003) by five-year age groups at death and screening indicator (screened, not screened) were included. Models 2 and 3 are described in detail in the Appendix.

Average difference in incidence-based breast cancer mortality between the two twelve-year periods (1974-1985, 1992-2003) among the cohort members (model 1), among the whole female population (model 2) and between the screened and non-screened women (model 3) represented the screening effect. Further details are presented in the Appendix.

The effect estimates were formulated by dividing the number of observed deaths from breast cancer by the number of expected breast cancer deaths without screening. For the overall effectiveness and for the screened women in the policy study, the observed deaths were derived from the cohort follow-up. For the whole female population in the policy study, they were obtained from the population data. The expected deaths were calculated by multiplying the expected mortality rates derived from the model by the person-years derived from the cohort (overall effectiveness and screened women in the policy study) or from the population data (all females in the policy study). The confidence intervals of the effect estimates were corrected with over-dispersion constants produced by the models (1.36 for model 1; 1.39 for model 2; 1.25 for model 3). The bias due to self-selection was adjusted using a method described by Cuzick and co-workers (Cuzick et al. 1997).

5. RESULTS

5.1 Screening process

In 1991-2000, the CSF centres sent 1 015 100 invitations and completed 926 500 visits (Table 6) (Studies I and II). Participation was highest among women aged 50-59, and lowest among those aged 40-49. The adjusted breast cancer detection rates ($p< 0.001$) and the positive predictive value of mammography ($p<0.001$) increased significantly with increasing age.

Among the 50-64 year old invitees, compliance was 90-93%. Recall rate was 4.6% at the first, and 2.3% at the subsequent screens. The rate of histological confirmation was 0.89% and 0.52%, and the breast cancer detection rate 0.44% and 0.36% respectively. The adjusted recall and histological confirmation rates were stable or increased slightly in 1991-2000. Per five calendar years, the adjusted breast cancer detection rates increased by 18%, and the adjusted positive predictive value of mammography by 11% ($p=0.001$). The adjusted rate of DCIS was stable, while there was an increase in the adjusted rates of stage I ($p<0.001$) and stage II+ ($p=0.0139$) cancers.

The annual number of centre-specific invitations for women aged 50-64 varied between 1000 and 3500 at the first screen, and between 3100-14 200 at the subsequent screens. Clear centre-specific differences were found in the recall rate, in the ratio of detection rate to background incidence, and in the proportion of DCIS, stage II+ cancers, and cancers that were node negative, as well as in the benign to malignant biopsy ratio (Table 7).

Table 6. Number of invitations, proportion of attendees, proportion of breast cancers among attendees, and positive predictive value of screening mammography by five-year age groups in 1991-2000.

	40-44	45-49	50-54	55-59	60-64	65-69	70-74	All
First screen								
Invitations	14229	6317	178267	1216	9046	-	-	209075
Visits	11847	4681	160846	1091	7975	-	-	186440
Attendance (%)	83.3	74.1	90.2	89.7	88.2	-	-	89.2
Breast cancers (%)	0.17	0.28	0.43	0.55	0.70	-	-	0.42
PPV mammography (%)	4.0	7.1	9.2	12.0	20.1	-	-	9.2
Subsequent screens								
Invitations	23604	46992	263530	304197	118938	43418	5313	805992
Visits	21199	40872	247517	283534	103959	38135	4787	740003
Attendance (%)	89.8	87.0	93.9	93.2	87.4	87.8	90.1	91.8
Breast cancers (%)	0.10	0.22	0.32	0.39	0.40	0.58	0.75	0.36
PPV mammography (%)	3.3	7.1	13.2	17.2	19.2	26.1	30.3	15.3

PPV= Positive predictive value of screening mammography

Table 7. Process indicators and their centre-specific variation (in parentheses) among 50-64 years old women in 1991-2000 with the desirable levels of the European Union (EU) for the screening process at the subsequent screens.

	First screen		*Subsequent screens*	
	Present study	**EU level**	**Present study**	**EU level**
Attendance	90% (85-95%)	>75%	93% (90-95%)	>75%
Recall rate	4.6% (1.9-7.0%)	<5%	2.3% (1.0-3.6%)	<3%
Detection rate / background	2.7xIR (2.3-4.2xIR)	>3xIR	2.1xIR (1.6-3.1xIR)	>1.5xIR
DCIS	10% (5-18%)	10-20%	10% (4-18%)	10-20%
Stage II+	33% (25-46%)	<30%	26% (23-33%)	<25%
Node negative	76% (69-79%)	>70%	79% (75-81%)	>75%
B:M ratio	4:4 (2.0-7.2:4)	<1:4	1.6:4 (1.2-3.2:4)	<1:4

DCIS= Proportion of screen-detected breast cancers that are ductal carcinoma in situ
Stage II+= Proportion of screen-detected breast cancers that are stage II+
Node negative= Proportion of screen-detected invasive breast cancers that are node negative;
B:M ratio= benign to malignant biopsy ratio
IR= background incidence rate

5.2 Screening validity

In 1991-1999, there were altogether 785 800 invitations and 720 900 visits among women aged 50-64 in the CSF centres (Study II). The number of screen-detected breast cancers was 670 at the first and 2045 at the subsequent screens. The number of interval breast cancers was 289 and 1101, and the number of incident breast cancers assuming no screening 588 and 2415. The breast cancer detection rates were 4.4 and 3.6 per 1000 women screened, and the rate of interval breast cancers 0.91 and 0.93 per 1000 women years. The proportion of invasive, non-localised cancers was higher among the screen-detected cancers than among the interval cancers.

The overall episode sensitivity determined by the detection method was 70% for the first, and 65% for the subsequent screens in 1991-2001. With the incidence method, the corresponding estimates were 51% and 54%. The sensitivity estimates 0-11 and 12-23 months after the screening visit were 64% and 32% at the first, and 70% and 38% at the subsequent screens by the incidence method (Table 8). The overall episode sensitivity decreased towards the end of the study period and increased with age (Table 9).

The episode sensitivity determined by the incidence method was significantly higher in the category 'high recall rate' compared to the sensitivity in the category 'low recall rate' (p=0.006) (Table 9). Overall, the episode sensitivity increased 13% per 1% absolute increase in the recall rate (p=0.008).

Table 8. Episode sensitivity within first and second year after screen in 1991-2001 with desirable levels of the European Union (EU) for sensitivity estimates at subsequent screens.

	First screen		***Subsequent screens***	
	Present study	**EU level**	**Present study**	**EU level**
Episode sensitivity				
(0-11 months)	64%	>70%	70%	>70%
(12-23 months)	32%	>50%	38%	>50%

Table 9. Risk ratios (RR adjusted) and Poisson 95 % confidence intervals (CI adjusted), and adjusted p-values for episode sensitivity by period, age, centre categories, centres and recall rate at subsequent screens. Episode sensitivity was determined by the incidence method.

	RR Adjusted	CI Adjusted	p-value Adjusted
Period[a]			
1991-1993	1		
1994-1996	0.82	0.71-0.94	
1997-1999	0.73	0.64-0.84	
Per 3-year category	0.86	0.80-0.92	p<0.001
Age[b]			
50-54	1		
55-59	1.08	0.95-1.23	
60-64	1.17	1.00-1.37	
Per 5-year category	1.08	1.00-1.17	p=0.055
Centre categories[c]			
Low recall rate (0.9-1.9)	1		
Intermediate recall rate (2.3-2.7)	1.09	0.96-1.25	
High recall rate (2.8-3.5)	1.26	1.07-1.48	
Per centre category	1.12	1.03-1.22	p=0.006
Centres (recall %)[d]			
1 (0.9)	1		
2 (1.3)	0.76	0.56-1.02	
3 (1.9)	1.22	0.93-1.60	
4 (2.3)	1.28	0.95-1.70	
5 (2.5)	1.09	0.84-1.40	
6 (2.6)	0.98	0.73-1.31	
7 (2.7)	0.86	0.62-1.20	
8 (2.8)	1.21	0.90-1.63	
9 (2.9)	1.00	0.69-1.45	
10 (3.5)	1.35	1.01-1.80	
Per recall %	1.13	1.03-1.23	p=0.008

[a] Adjusted by five-year age groups and centre categories, p-value for trend
[b] Adjusted by three-year periods and centre categories, p-value for trend
[c] Adjusted by three-year periods and five-year age groups, p-value for trend, centre-specific variation in recall rates in parentheses
[d] Adjusted by three-year periods and five-year age groups, organised by increasing recall rate

5.3 Screening effectiveness

In 1992-2003, altogether 2 731 268 person-years were accumulated into the cohort in ages 50-69 at death (Table 10) (Study IV). The reduction in the incidence-based breast cancer mortality among invitees was 22% (relative risk 0.78, 95% confidence interval 0.70-0.87). The highest mortality reduction, 29%, was analysed among women aged 60-64 at death. The ratio between the observed and expected breast cancer deaths among the participants was 0.66 (0.58-0.75). The corresponding relative risk among the non-participants was 1.56 (1.25-1.91). After adjusting for the self-selection in attendance, the relative risk among the participants became 0.72 (0.56-0.88) (Table 11).

In the three centre categories grouped by ascending order of recall rates (low, intermediate, high), the risk ratios among the screening invitees at ages 50-69 at death were 0.83 (0.67-1.01), 0.71 (0.58-0.85), and 0.79 (0.60-1.02). After adjusting for self-selection in attendance, no clear association between the screening efficacy and the recall rates could be found (Figure 4).

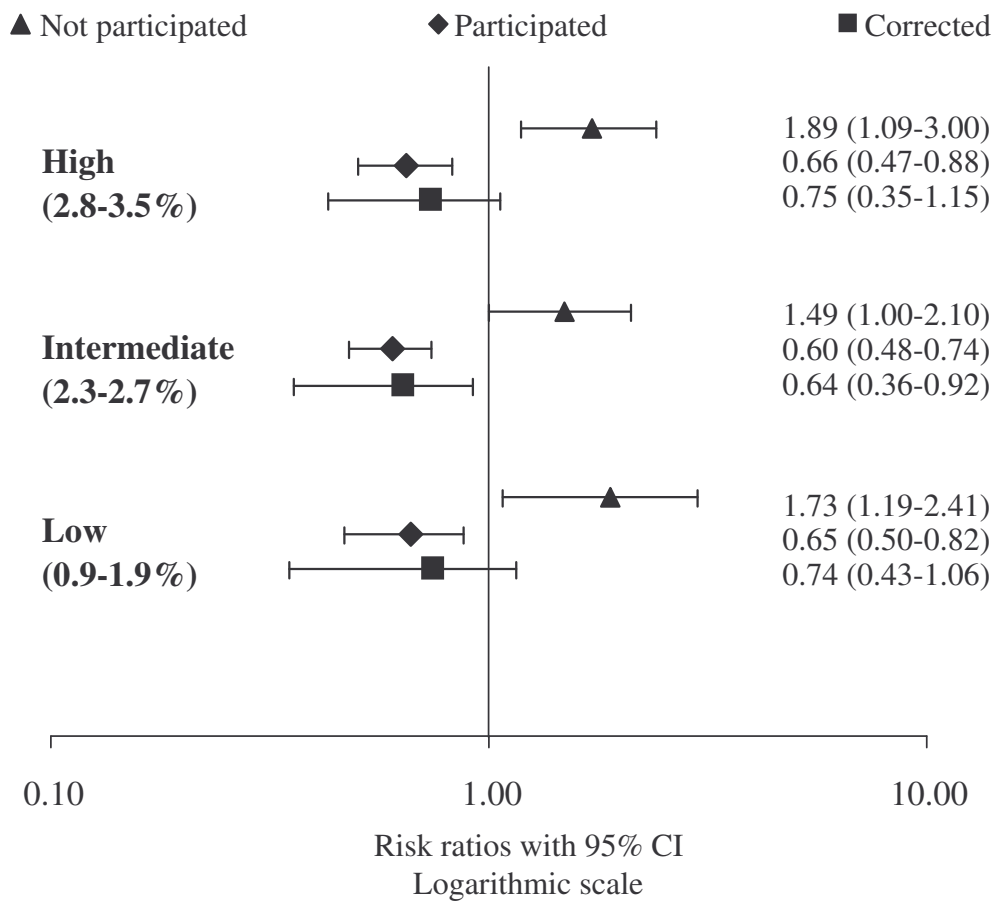
Table 10. Person-years, observed and expected numbers of breast cancer deaths, and effect estimates of invitation to screening (observed deaths/expected deaths) with 95% confidence intervals in 1992-2003 in ages 50-69 at death.

	50-54	**55-59**	**60-64**	**65-69**	**50-69**
Person-years	903692.6	795537.7	630446.3	401591.7	2731268.3
Observed deaths	97	203	174	143	617
Expected deaths	131.8	241.9	244.4	170.9	789.0
Observed/Expected (95% CI)	**0.74 (0.55-0.96)**	**0.84 (0.69-1.01)**	**0.71 (0.58-0.87)**	**0.84 (0.66-1.04)**	**0.78 (0.70-0.87)**

Table 11. Person-years, observed and expected numbers of breast cancer deaths, and effect estimates of participation to screening (observed deaths/expected deaths) with 95% confidence intervals in 1992-2003 in ages 50-69 at death. Corrected refers to effect estimates with adjustment for selection bias.

	50-54	**55-59**	**60-64**	**65-69**	**50-69**
Non-participants					
Person-years	98202.8	95143.0	90956.2	74562.8	358864.8
Observed deaths	29	58	45	33	165
Expected deaths	14.2	28.0	33.8	30.1	111.8
Observed/Expected	**2.04**	**2.07**	**1.33**	**1.10**	**1.56**
(95% CI)	**(1.15-3.26)**	**(1.41-2.91)**	**(0.86-1.95)**	**(0.65-1.71)**	**(1.25-1.91)**
Participants					
Person-years	805489.8	700394.7	539490.2	327028.8	2372403.4
Observed deaths	68	145	129	110	452
Expected deaths	117.6	213.9	210.6	140.8	682.9
Observed/Expected	0.58	0.68	0.61	0.78	0.66
	(0.41-0.79)	(0.54-0.84)	(0.48-0.77)	(0.60-1.00)	(0.58-0.75)
Corrected	**0.66**	**0.79**	**0.65**	**0.80**	**0.72**
(95% CI)	**(0.28-1.04)**	**(0.49-1.09)**	**(0.38-0.92)**	**(0.46-1.14)**	**(0.56-0.88)**

Figure 4. Effect estimates with 95% confidence intervals in three centre categories (recall rates, %, in parentheses). Corrected refers to estimates with adjustment for self-selection.



There were 2 388 775 person-years in ages 60-79 at death in the population data (Study V). Almost 80% of these accumulated from municipalities that had invited 50-59 year old women regularly and 60-69 year old women irregularly to organised mammography in 1992-2003 (Table 12). In this invitational category, the overall reduction in breast cancer mortality in ages 60-79 at death was 16% (0.84, 0.75-0.92). The greatest overall reduction, 28% (0.72, 0.51-0.97), was observed in municipalities that had invited 50-69 year old women regularly to mammography screening in the study period. No reduction in breast cancer mortality in ages 60-79 at death was seen in municipalities that had invited only 50-59 year old women to organised screening in 1992-2003.

Table 12. Person-years, observed and expected numbers of breast cancer deaths, and effect estimates (observed deaths/expected deaths) with 95% confidence intervals in 1992-2003 in ages 60-79 at death among all women by three policy categories.

	Regular 50-59	**Regular 50-69**	**Irregular 50-69**
Person-years	256548	228527	1903700
Observed deaths	131	75	728
Expected deaths	125.8	104.5	871
Observed/Expected (95% CI)	**1.04 (0.81-1.31)**	**0.72 (0.51-0.97)**	**0.84 (0.75-0.92)**

Regular 50-59= regular invitations for women aged 50-59
Regular 50-69= regular invitations for women aged 50-69
Irregular 50-69= regular invitations for women aged 50-59 with irregular invitations for women aged 60-69

In the cohort of screened women, there were altogether 1 023 598 person-years during the follow-up. A reduction in breast cancer mortality was observed in all invitational categories, but it was concentrated on different ages. In the municipalities inviting 50-59 year old women regularly, a non-significant reduction of 35% was observed in ages 60-69 at death. In the other two invitational categories, the reduction in ages 60-79 was 37-38%. The relative risk among the non-screened was elevated in municipalities inviting 50-59 or 50-69 years old women regularly. Owing to widespread non-registered screening in municipalities offering irregular screening to ages 60-69, the relative risk among the non-screened in this invitational category was slightly below 1.

6. DISCUSSION

This study evaluated the performance and effectiveness of breast cancer screening in Finland by using individual data on screening invitees and participants from the initiation of service screening period until the year 2003. The study elaborated screen-specific information on process and validity indicators by age and by period as well as variation in these parameters between the screening providers. The study was the first to evaluate the episode sensitivity and the breast cancer mortality among the invitees and attendees of mammography service screening in Finland, and also the first to examine associations between the screening performance and the screening outcome.

The data were gathered from ten centres of the Cancer Society of Finland. These centres covered 50-60% of biennial invitations within the first ten years of service screening in Finland, and sent over one million invitations to women aged 40-74. The coverage of invitations was almost 100% in the age group 50-59. About one quarter of women aged 60-69, and less than 10% of women aged 40-49 and 70-74 received invitations to biannual, free-of-charge mammography. Overall attendance was 90%.

Per one subsequent screen, less than 3% of attendees were inferred for further examinations. Among these, one sixth was sent for histological confirmation. The breast cancer detection rate among the attendees was 0.4%. The positive predictive value of mammography increased towards the end of the study period and with age. At the subsequent screens, the overall episode sensitivity by the detection method was 65% and by the incidence method 54%, and the sensitivity estimates during the first and second years after the screening visit 70% and 38%. The overall sensitivity decreased towards the end of the study period and increased with age.

The reduction in incidence-based breast cancer mortality among 50-69 year old screening invitees was 22% in ages 50-69 at death. Among participants of the same age, the adjusted reduction was 28%. When the invitational policies were compared, the strongest reduction in breast cancer mortality among all women aged 60-79 at death, 28%, was estimated in municipalities inviting 50-69 year old women regularly. No overall reduction in breast cancer mortality in ages 60-79 at death was observed in municipalities restricting invitations to 50-59 year old women.

Centre-specific differences in process parameters, such as recall rate, breast cancer detection rate, and positive predictive value of mammography, were wide. Variability in the centre-specific episode sensitivity was associated with variability in recall. In further analyses, no association between the screening outcome, the screening sensitivity, and the level of recall could be found.

6.1 Data and methods

Data on screening performance were derived from the Mass Screening Registry, the diagnostic data on episode sensitivity from the Finnish Cancer Registry. The data on screening effectiveness were obtained from the Finnish Cancer Registry, the National Population Registry and the Mass Screening Registry. These three nationwide registries covered all the municipalities studied, and provided reliable data on diagnostics and mortality, and complete information on invitations and visits to the organised mammography within the ten centres of Cancer Society of Finland. Lack of information on the invitations to mammography screening with a fee as well as on the delivery of opportunistic mammography hampered the investigation of effect estimates, however, and together with the widespread use of HRT (Hemminki 2004, Rutanen and Ylikorkala 2004) probably also had a role in the increase of interval cancer rates from 1990's to 2000's.

In the studies on process indicators (Studies I and II), the diagnostic data on screen-detected breast cancers were derived from the Mass Screening Registry, and could be classified by stage. In the validity study (Study III), the data were obtained from the Finnish Cancer Registry, which included diagnoses on both the screen-detected and interval breast cancers. Since the availability of pTNM stage or size of a tumour was limited in the Finnish Cancer Registry, the screen-detected and the interval breast cancers were classified into four categories: carcinoma in situ, invasive localised, invasive non-localised, and unknown. The determination of background incidence in the validity study was also rather crude: municipality level information from the pre-screening era was first generalized to centres covering these municipalities, and thereafter a similar increase in incidence was assumed for each centre.

We used both the detection and the incidence methods to achieve comparability with studies reporting their episode sensitivity estimate by the detection method only (Study III). The preferred incidence method is not biased by definition due to lead time and over-diagnosis. The incidence method, however, is sensitive to correct estimation of background incidence, to changes in diagnostic activities outside screening and to selection among screening participants (Fletcher et al. 1993, Hakama et al. 1997, Vainio and Bianchini 2002).

To obtain breast cancer deaths in the study period 1992-2003, we performed a cohort follow-up of individual screening invitees. Women invited at least once to any of the CSF centres remained members of the cohort until emigration, death or end of the follow-up irrespective of their possible migration to municipalities outside the study area. This fixed accumulation of person-years differed from the dynamic accumulation of person-years in the population data, but was not likely to affect the comparability between the individual and the population level data. Breast cancer mortality in the absence of screening was derived by modelling population level data.

An increase in breast cancer incidence and a decrease in breast cancer mortality (1995+) were reported among women aged 50-74 between the 1970's and the 2000's (Botha et al. 2003, FCR 2005). Our models for the incidence-based breast cancer mortality were constructed to capture both of these phenomena. Due to opposing directions in incidence and mortality, absolute differences between observed mortality rates in the pre-screening era and expected mortality rates in the screening era remained small. The use of observed rates from 1974-1985 instead of expected rates in the absence of screening in 1992-2003 would thus have given almost similar effect estimates.

Effects of screening, other diagnostic services and treatment on breast cancer mortality are difficult to differentiate without comparable controls or control areas. In the United States, the effects of screening and adjuvant therapy were estimated to be similar, but variation in the modelling approaches reflected considerable uncertainty. In East Anglia, 60% of improved survival was estimated to be due to earlier diagnosis by screening (Berry et al. 2005, Duffy et al. 2006). In our study with individual participants and non-participants, the corrected estimates among the participants were well below one, referring that self-selection did not explain the result. An earlier study from Finland, where a significantly greater mortality reduction was observed in a city that had provided screening to the studied birth cohorts compared to a city where no screening had been offered, also supported the impact of screening on breast cancer mortality (Parvinen et al. 2006).

To study further the role of screening in mortality reduction, we formulated a design in which the breast cancer mortality was compared between differential invitational coverage in ages 60-69. As no systematic variation in treatment or in other diagnostic services between the Finnish municipalities had been reported, and evidence-based guidelines for the current care had been provided from the beginning of 1990's, the varying invitational policies in women aged 60-69 enabled us to investigate the screening impact without substantial influence from other health care

activities on the effect estimates. Wide variation in the reduction of breast cancer mortality between the invitational categories confirmed the contribution of screening to breast cancer mortality in Finland.

6.2 Comparison to other studies

Screening delivery

Organised mammography programmes in Europe were launched after the first published results on randomised trials. Sweden was the pioneer country of randomisation, and achieved full coverage of screening in 1997. Nationwide coverage in the Netherlands was achieved the same year, in the UK two years earlier. The Norwegian programme covered the whole country in 2004. In Denmark, the population coverage of breast cancer screening is so far 20%, while in Italy the programme covers approximately 50% of the female population. In Finland, pilot projects on breast cancer screening were started in 1982-84 in the city of Tampere and in thirteen municipalities in Kymenlaakso district (Lauslahti et al. 1986, Kärkkäinen et al. 1987, Hakama et al. 1995). The national programme was launched in 1987, and achieved full coverage in 1992 (Hakama et al. 1997).

Most European programmes offer mammography screening for women aged 50 years or more. In ages 40-49, evidence for screening effectiveness has remained limited (Nystrom et al. 2002, Moss et al. 2006), and excess risk of radiation-induced breast cancer has been reported (e.g. Berrington de Gonzalez and Reeves 2005). Our study from Finland showed a significant increase in the positive predictive value of mammography by age. This concurred with the results of the randomised trial in the United Kingdom, where 23% of 40-49 year old women experienced at least one false positive result while the corresponding figure in women older than 50 years was 12% (Moss et al. 2006).

Consequently, instead of lowering the invitational age below 50 years, most European programmes invite women up to 69 years of age, and some have extended invitations to ages 70-74. Due to potential over-treatment, co-morbidity and limited number of additional life-years gained, no further increase in the upper age limit has been suggested (Fracheboud et al. 2006).

Screening performance

In 2001, the European Union (EU) published the first guidelines on the quality assurance of mammography screening. The updated guidelines from 2006 contain 39 'key performance indicators' (Perry et al. 2006). In Tables 4, 7 and 8, the most widely reported eight indicators from the organised mammography programmes in Europe are compared to the EU reference values. For the country-specific target age, all the eight indicators have been reported from only four countries.

The quality indicators of EU have given rise to criticism e.g. for inadequate calculation of participation over time, or for lack of consistence between the stage II+ and the node-negative cancers (corrected partly for the 2006 version) (Njor et al. 2003, von Euler-Chelpin et al. 2006). Moreover, presentation of the desirable values as proportions instead of rates may overestimate the success of screening e.g. by ignoring over-diagnosis. The variability in the target age, in the definitions of interval cancers (early recall, control examinations due to suspicious findings, determination of follow-up after the last screen), or in the determination of background incidence hampers comparison of quality parameters between countries and programmes.

The criteria for the quality indicators are generally originated to Two-County trial from Sweden. Publications on these trials have not, however, reported all of the eight quality indicators presented in Table 4, and some of the indicators seem to be based on non-specified screens, first or

subsequent screens only, or on the results from the early years of the trial. These suggest that the desirable levels of the EU quality indicators refer to expert opinion rather than evidence from randomised (or non-randomised) studies. Nevertheless, the EU indicators represent the best (and only) available reference for the quality assurance of organised mammography screening in Europe, and are therefore used in this study as a baseline in the comparison of the process and validity estimates of the current study with the estimates from the other European screening programmes.

Screening attendance indicates the potential of the screening programme to reduce mortality from breast cancer. Most European programmes achieved the EU desirable level (75%) (Table 4). Italy formed an exception with the range of attendance from 38% to 62%. The participation rate in Finland was among the highest (Tables 4 and 7).

Ratio of detection rate to background incidence characterises the capability of screening to detect breast cancers among the target population. The estimates from Finland, Denmark, the Netherlands and Norway differed from each other, and only the Danish and the Norwegian estimates met the EU desirable levels both at the first and at the subsequent screens (Tables 4 and 7). The EU criteria do not have upper limits for this indicator. At the subsequent screens, however, an observed ratio greater than two may indicate over-diagnosis (Study II).

Stage distribution of breast cancers at screening represents the success of screening in enabling earlier diagnosis and reducing mortality from advanced disease (Day et al. 1989). The proportion of node negative invasive cancers out of all screen-detected cancers fulfilled the EU criteria in Finland, Denmark, Norway, and the Netherlands (Fracheboud et al. 1998, Vejborg et al. 2002, Njor et al. 2003, Hofvind et al. 2007a) (Tables 4 and 7). The criterion for the proportion of stage II+ cancers was not met in Finland or in Denmark, however. In the Netherlands, the stage II+ level was

met (Fracheboud et al. 1998), and a later report showed a clear decline in the rates of advanced cancers after the start of screening (Fracheboud et al. 2004). Recent studies from Finland, too, have reported a decrease in the post screening incidence of non-localised breast cancers (Seppanen et al. 2006, Anttila et al. 2008). Reduction in the rates of node positive tumours, tumours more than two centimetres in diameter, and stage II+ tumours in the screening era compared to those in the pre-screening era have been reported in Sweden (SOSSEG 2007). An earlier study from Sweden showed a substantial increase in the incidence of invasive breast cancers more than six years after the first screening round, however (Jonsson et al. 2005). In Norway, no decline in advanced cancers was seen during the first years after screening (Jonsson et al. 2005, Hofvind et al. 2007b).

The proportion of DCIS out of all cancers fulfilled the EU criteria in all European countries. The UK and Norway were nearby the upper limit at the first and subsequent screens. High proportion of DCIS evokes concerns about over-diagnosis due to screening. No threshold value for over-diagnosis within organised screening exists, however, and the data on untreated DCIS is limited. The progression of DCIS to invasive cancer has been studied by following low-grade DCIS among women who have undergone diagnostic biopsy alone, by examining screen-detected tumours, and by analysing autopsy data. Among women with a previous diagnostic biopsy, 40-86% of low-grade DCIS remained non-progressive after ten years of follow-up (Burstein et al. 2004). By modelling the screen-detected DCIS, 37 % of first-screen, and only 4% of subsequent screen cases were non-progressive (Yen et al. 2003). The downward stage-shifting of screen-detected tumours from the Swedish Two-County trial suggested, however, that the detection of DCIS cases that would progress to invasive cancers saved considerably fewer lives than were saved by early detection of invasive disease (Duffy et al. 2003). The evidence from autopsy studies with median prevalence of 9% of DCIS among women without previous diagnosis of breast cancer also reveals that the detection of DCIS may not be very beneficial (Welch and Black 1997).

Recall rate and *benign to malignant biopsy ratio* reflect the screening specificity. The level of recall fulfilled the EU criteria in most European countries. The country-specific variation in Finland and in Sweden was wide, however (Tables 4 and 7). In many countries, the criteria for benign to malignant biopsy ratio were not met (Tables 4 and 7). The low invitational age in Finland probably affected our estimates. In the national statistics of mass screening from 2005, the overall benign to malignant biopsy ratio in ages 50-64 in Finland was 1.0:4, however (www.cancerregistry.fi/eng/statistics). Introduction of core needle biopsy has been suggested to have a role behind the improved estimate.

Episode sensitivity within the first follow-up year in Finland fulfilled the EU criteria (Tables 4 and 8). In the second follow-up year, the estimate remained below the desirable level. Similar results were reported from all the other four countries calculating their episode sensitivity by the incidence method (Woodman et al. 1995, Fracheboud et al. 1999, Njor et al. 2003, Hofvind et al. 2007a). The level of the EU criterion is based on the results of randomised trials on women aged 50-69 from Kopparberg and Östergötland (the counties of Two-County trial). The main target age in Finland was 50-59, and the sensitivity increased with age. Moreover, the reported service screening estimate in Östergötland was somewhat lower than our estimate (57% vs. 65% by the detection method for women aged 50-59) suggesting that service screening settings may differ from those of the rigorously controlled trial (Vitak et al. 1997, Broeders et al. 2001). Tumour distribution of interval cancers was less favourable than that of screen-detected cancers in Finland, and the proportion of non-localised interval cancers in the first follow-up year after a screen was higher than in the second follow-up year. This indicates that many of the screen-detected cancers were slow-growing, and cancers missed at screening were likely to have a shorter sojourn time. Similar results were reported e.g. in the Netherlands (Fracheboud et al. 1999).

We used our estimates for episode sensitivity with estimates for lead-time bias from a modelling study to examine the over-diagnosis of screening (Paci et al. 2004) (Study III). In the study, an excess of some 11% in the cumulative breast cancer incidence due to lead-time was ascertained during the first ten years of screening. With this value, our episode sensitivity by the detection method would have become 60%. The difference between this sensitivity estimate and the estimate by the incidence method (54%) could be attributed to over-diagnosis and self-selection bias. In other European countries, the estimates for over-diagnosis have varied 5-50% (Olsen et al. 2003, Zahl et al. 2004, Jonsson et al. 2005, Olsen et al. 2006). A randomised study from Sweden reported a 10% increase in the cumulative rate of breast cancers fifteen years after the last screen (Zackrisson et al. 2006).

Screening effectiveness

Our study on screening effectiveness with an individual cohort-follow up resulted in a 22% overall reduction in the incidence-based mortality from breast cancer in ages 50-69 at death (Table 10, Study IV). The estimate was at a similar level to those from Denmark, Italy and the latest estimates from the Netherlands and Sweden (Duffy et al. 2002, Otto et al. 2003, Olsen et al. 2005b, SOSSEG 2006b). The studies from the United Kingdom have provided lower figures (Blanks et al. 2000a, McCann et al. 2001). In ages 50-59 at death, the reduction in Finland was greater than in Denmark (the only country providing age-specific results). The accumulation of post-screening follow-up in ages 60-69 in Finland limited the mortality reduction in this age group, however. In our following article (Study V), we studied incidence-based mortality from breast cancer in three invitational categories in ages 60-79 at death. Among all females, the greatest mortality reduction (28%), was observed in municipalities inviting women aged 50-69 regularly to mammography screening (Table 12). The reduction of 32% (0.68, 0.42-1.03) in ages 70-79 at death in this invitational category was

close to the estimates from Denmark for ages 70-74 at death (42%) and 75-79 at death (31%) (Olsen et al. 2005b).

The overall efficacy of screening in ages 50-69 at death in Finland, 28%, was at a lower level compared to the estimates from Denmark and Sweden (Tables 5 and 11, Study IV). This was probably due to the high participation rate among the Finnish invitees. The only criterion for participation in the current study was a positive response after the first invitation to CSF centres in 1992-2003. In our earlier study on screening performance we had found, that over 90% of first-screen attendants re-attended at least once after their subsequent invitation. Furthermore, staying alive is a prerequisite for a subsequent invitation. The classification of participation on the basis of first invitation thus gave equal opportunity for all of the invitees to attend screening.

In light of earlier results from Helsinki and Turku (Parvinen et al. 2006), and our study comparing reduction in breast cancer mortality between various invitational policies, we attributed most of the overall reduction in the breast cancer mortality in Finland to screening. In the UK, screening was estimated to account for one third of the overall reduction (Blanks et al. 2000a). This may have been due to the large number of breast cancer deaths that had been diagnosed before the start of screening. The authors expected a 'substantial reduction' in breast cancer mortality by the year 2010, particularly in ages 55-69. In Sweden, the reduction in breast cancer mortality in the early years was studied by comparing areas with screening to areas without screening using various designs (Lenner and Jonsson 1997, Jonsson et al. 2000, Jonsson et al. 2001, Jonsson et al. 2003a, Jonsson et al. 2003b). In later studies, a comparison of incidence-based breast cancer mortality between the screening and the pre-screening era was performed (Duffy et al. 2002, Tabar et al. 2003, SOSSEG 2006a, SOSSEG 2006b).

The second article of the Service Screening Evaluation Group assessed the risk of dying from breast cancer among women unexposed to screening and estimated that it was 10 % lower in 1995 than in 1985 (SOSSEG 2006a). Among women exposed to screening, the corresponding risk of dying had decreased by 45%. According to the authors, this suggested that most of the mortality reduction among participants was due to screening. Similar conclusions were also reported in earlier papers (Duffy et al. 2002, Tabar et al. 2003). In the Netherlands, the authors argued that systemic adjuvant therapy was unlikely to affect mortality reduction, because the turning point in the mortality curves was around the time of the introduction of screening, regardless of whether screening was started early or late (Otto et al. 2003). The Danish authors stated that the screening invitees had received treatment similar to that provided for women outside Copenhagen, and it was thus reasonable to assume that the reduction in breast cancer mortality was due solely to screening (Olsen et al. 2005b).

The first Cochrane Review of the results of screening trials started an ongoing controversy on the validity of the randomisation of these trials (Gotzsche and Olsen 2000, Olsen and Gotzsche 2001, Twombly 2007). Nevertheless, further review of the evidence from the trials confirmed their effectiveness (Vainio and Bianchini 2002, Freedman et al. 2004), and the authors of the latest Cochrane Review admitted that 'screening likely reduces breast cancer mortality' (Gotzsche and Nielsen 2006). The discussions on studies evaluating the effectiveness of service screening have been less dramatic, and have been conducted in academic journals only. The mortality reduction of 20% in the Netherlands has been considered surprisingly large, because the Dutch mammography programme started up gradually in 1989-1997, and the effect was measured by the year 2001 (Voogd and Coebergh 2003). The Danish results have been criticised for not following the mortality trend in the non-screened ages (40-49), and for not taking into consideration the fall in the use of hormone replacement therapy in the late 1990's (Grant 2005, Zahl and Maehlen 2005). The full

mortality effect has also been considered to appear too early (Gotzsche et al. 2005). The Danish authors replied that the mortality reduction in women aged 40-49 was statistically non-significant 6% and warnings against the use of hormone replacement therapy were issued mostly after the study period. The significance in the effect was reached six years after the start of the programme, which concurred with the overview of the Swedish randomised trials.

6.3 Interpretation of results

The ultimate goal of breast cancer screening is to reduce mortality from breast cancer. Overall results from the randomised studies showed a 25% reduction in breast cancer mortality due to screening (Vainio and Bianchini 2002). The organised mammography programme in Finland achieved an almost similar overall effect, 22%, and the episode sensitivity was at equal level to those European screening programmes showing a significant reduction in breast cancer mortality (Otto et al. 2003, Olsen et al. 2005b). The organised breast cancer screening in Finland thus achieved the most prominent goal of screening by detecting breast cancers that would eventually have led to death.

Low delivery of service screening in ages 60-69 limited the overall mortality reduction in this age group in Finland. In 2007, the Ministry of Health Care and Social Welfare announced a Decree on Screening according to which the invitational age of breast cancer screening is to be extended gradually from 50-59 to 50-69. Some Finnish municipalities had offered screening for women aged 50-69 throughout the 1990's, however, and many had practised screening at the individuals' own expense for ages 60-64. Our study on the differing invitational policies in Finland showed, that uniform extension of invitations to ages 50-69 among the invitees of the CSF centres in 1992-2003 would almost have doubled (172 to 312) the number of prevented breast cancer deaths in ages 60-

79. Gradual expansion of invitations to ages 50-69 during the forthcoming years in Finland will thus increase the number of prevented breast cancer deaths among the elderly, although the annual number of deaths prevented will remain rather small. The evaluation of cost-effectiveness of this expansion will be challenging, e.g. due to lack of baseline estimates on the cost-effectiveness of opportunistic screening at ages 60-69.

Only few studies have analysed pathways from screening delivery to screening outcome, even though variation in screening performance has been reported by several programmes (Table 4). The results on screening performance are difficult to interpret, and the current recommendations hard to justify if information on the relationships between the level of performance and the level of mortality outcome is scarce. Our studies on the screening process and validity found variation in process parameters between screening centres and an inverse correlation between level of recall and incidence of interval cancers. Further analysis of screening effectiveness discovered no association between screening efficacy, episode sensitivity, and recall rate. This may indicate differences in diagnostic and treatment activities within and/or outside organised screening. In the Netherlands, however, where recall rates have constantly been lower than 1%, studies have shown reduction in the rates of advanced cancers and similar overall impact of mammography screening on breast cancer mortality as in our study (Fracheboud et al. 2001, Otto et al. 2003, Fracheboud et al. 2004). Yet, one Dutch study reported increasing detection rates with increasing level of recall, and recommended higher recall rates (up to 4%) to avoid interval cancers or advanced-stage cancers at screening (Otten et al. 2005). The Finnish and the Dutch results suggest limited potential for the performance parameters in predicting the impact of screening on breast cancer mortality. Modelling approaches from the United States with similar overall conclusion warrant incidence of advanced cancers and programme sensitivity for further consideration as surrogates for the mortality outcome (Habbema et al. 2006).

7. CONCLUSION

There is growing evidence that population-based breast cancer screening is effective. Our results from Finland showed significant reduction in breast cancer mortality among the invitees and participants of organised screening. Extension of invitations to ages 60-69 in Finland will decrease the number of breast cancer deaths among the elderly.

The Finnish estimates for screening performance were at a similar level to those in the other European service screening programmes. There was, however, variation in the process and validity indicators, and ambiguity in the relationships between the process and outcome. These findings indicate that performance indicators should be used with caution as predictors of mortality outcome.

ACKNOWLEDGEMENTS

This study was carried out in the Mass Screening Registry of the Finnish Cancer Registry.

My deepest gratitude is due to my two supervisors, Docent Ahti Anttila, head of research of the Mass Screening Registry, and Professor Matti Hakama. Their vast experience in epidemiology was a prerequisite for the completion of the study. Our long discussions, which sometimes were rather demanding, helped me to understand the various aspects of cancer epidemiology and screening.

Professor Timo Hakulinen, director of the Finnish Cancer Registry, I thank for his valuable guidance in statistics and useful comments on the manuscripts. Profound thanks also to Dr. Sirpa Heinävaara, who was an indispensable help in modelling, and provided excellent support in various phases of mortality evaluation. Dr. Irma Saarenmaa provided essential guidance in radiology. Statistician Tapio Luostarinen contributed to the statistical analyses. IT analyst Håkan Forsman did long hours in data management.

I sincerely thank for the official reviewers, Professor Peter B. Dean and Docent Pia Verkasalo, for their careful evaluation of the thesis. I am also grateful to Virginia Mattila for the editing of the English language.

My warm thanks go to the staff of the Mass Screening Registry and the Finnish Cancer Registry, especially to my two colleagues, MD Laura Kotaniemi-Talonen and Statistician Johanna Seppänen, who tirelessly gave their support and help whenever needed. Docent Nea Malila, director of the Mass Screening Registry, shared her thoughts and ideas on epidemiological research, and created a warm atmosphere for our department. Kaija Halonen, Minna Heikkilä, Sanna Kuivalainen, Liisa

Määttänen, Anni Pehkonen, Päivi Styrman and Liisa Rita I thank for many good laughs during the coffee breaks, and help in practical matters.

I am grateful to my friends for the good times we have shared and to my husband for his patience.

The financial support provided by the Finnish Cancer Society and the Doctoral School of Public Health is gratefully acknowledged.

Helsinki, May 2008

Tytti Sarkeala

REFERENCES

Alexander FE, Anderson TJ, Brown HK, Forrest APM, Hepburn W, Kirkpatrick AE, Muir BB, Prescott RJ and Smith A (1999)**:** 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. Lancet 353**:** 1903-08.

Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, Ljungberg O, Ranstam J and Sigfusson B (1988)**:** Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. BMJ 297**:** 943-8.

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

Anttila A, Sarkeala T and Heinavaara S (2008)**:** Impacts of the Finnish service screening programme on breast cancer rates. In press

Autier P, Shannoun F, Scharpantgen A, Lux C, Back C, Severi G, Steil S and Hansen-Koenig D (2002)**:** A breast cancer screening programme operating in a liberal health care system: the Luxembourg Mammography Programme, 1992-1997. Int J Cancer 97**:** 828-32.

Bakken K, Alsaker E, Eggen AE and Lund E (2004)**:** Hormone replacement therapy and incidence of hormone-dependent cancers in the Norwegian Women and Cancer study. Int J Cancer 112**:** 130-4.

Ballard-Barbash R, Klabunde C, Paci E, Broeders M, Coleman EA, Fracheboud J, Bouchard F, Rennert G and Shapiro S (1999)**:** Breast cancer screening in 21 countries: delivery of services, notification of results and outcomes ascertainment. Eur J Cancer Prev 8**:** 417-26.

Bernstein L and Ross RK (1993)**:** Endogenous hormones and breast cancer risk. Epidemiol Rev 15**:** 48-65.

Berrington de Gonzalez A and Reeves G (2005)**:** Mammographic screening before age 50 years in the UK: comparison of the radiation risks with the mortality benefits. Br J Cancer 93**:** 590-6.

Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD and Feuer EJ (2005)**:** Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 353**:** 1784-92.

Bjurstam N, Bjorneld L, Duffy SW, Smith TC, Cahlin E, Eriksson O, Hafstrom LO, Lingaas H, Mattsson J, Persson S, Rudenstam CM and Save-Soderbergh J (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**:** 2091-9.

Bjurstam N, Bjorneld L, Warwick J, Sala E, Duffy SW, Nystrom L, Walker N, Cahlin E, Eriksson O, Hafstrom LO, Lingaas H, Mattsson J, Persson S, Rudenstam CM, Salander H, Save-Soderbergh J and Wahlin T (2003)**:** The Gothenburg Breast Screening Trial. Cancer 97**:** 2387-96.

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

Blanks RG, Moss SM, McGahan CE, Quinn MJ and Babb PJ (2000a): Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: comparison of observed with predicted mortality. BMJ 321: 665-9.

Blanks RG, Moss SM and Patnick J (2000b): Results from the UK NHS breast screening programme 1994-1999. J Med Screen 7: 195-8.

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

Boyd NF, Guo H, Martin LJ, Sun L, Stone J, Fishell E, Jong RA, Hislop G, Chiarelli A, Minkin S and Yaffe MJ (2007): Mammographic density and the risk and detection of breast cancer. N Engl J Med 356: 227-36.

Boyle P and Ferlay J (2005): Cancer incidence and mortality in Europe, 2004. Ann Oncol 16: 481-8.

Bremnes Y, Ursin G, Bjurstam N, Lund E and Gram IT (2007): Different types of postmenopausal hormone therapy and mammographic density in Norwegian women. Int J Cancer 120: 880-4.

Brinton LA and Schairer C (1993): Estrogen replacement therapy and breast cancer risk. Epidemiol Rev 15: 66-79.

Broeders MJ, Peer PG, Straatman H, Beex LV, Hendriks JH, Holland R and Verbeek AL (2001): Diverging breast cancer mortality rates in relation to screening? A comparison of Nijmegen to Arnhem and the Netherlands, 1969-1997. Int J Cancer 92: 303-8.

Broeders MJ, Scharpantgen A, Ascunce N, Gairard B, Olsen AH, Mantellini P, Mota TC, Van Limbergen E, Seradour B, Ponti A, Trejo LS and Nystrom L (2005): Comparison of early performance indicators for screening projects within the European Breast Cancer Network: 1989-2000. Eur J Cancer Prev 14: 107-16.

Buerger H, Otterbach F, Simon R, Poremba C, Diallo R, Decker T, Riethdorf L, Brinkschmidt C, Dockhorn-Dworniczak B and Boecker W (1999): Comparative genomic hybridization of ductal carcinoma in situ of the breast-evidence of multiple genetic pathways. J Pathol 187: 396-402.

Buerger H, Mommers EC, Littmann R, Simon R, Diallo R, Poremba C, Dockhorn-Dworniczak B, van Diest PJ and Boecker W (2001): Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution. J Pathol 194: 165-70.

Bulliard JL, Sasieni P, Klabunde C, De Landtsheer JP, Yankaskas BC and Fracheboud J (2006): Methodological issues in international comparison of interval breast cancers. Int J Cancer 119: 1158-63.

Burstein HJ, Polyak K, Wong JS, Lester SC and Kaelin CM (2004): Ductal carcinoma in situ of the breast. N Engl J Med 350: 1430-41.

Cade JE, Burley VJ and Greenwood DC (2007): Dietary fibre and risk of breast cancer in the UK Women's Cohort Study. Int J Epidemiol 36: 431-8.

Cadman BA, Ostrowski JL and Quinn CM (1997): Invasive ductal carcinoma accompanied by ductal carcinoma in situ (DCIS): comparison of DCIS grade with grade of invasive component. The Breast 6: 132-7.

Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM, Geller BM, Abraham LA, Taplin SH, Dignan M, Cutter G and Ballard-Barbash R (2003): Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. Ann Intern Med 138: 168-75.

CGHFBC Collaborative Group on Hormonal Factors in Breast Cancer (1997): Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet 350: 1047-59.

CGHFBC Collaborative Group on Hormonal Factors in Breast Cancer (2002): Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer 87: 1234-45.

Chen WY, Colditz GA, Rosner B, Hankinson SE, Hunter DJ, Manson JE, Stampfer MJ, Willett WC and Speizer FE (2002): Use of postmenopausal hormones, alcohol, and risk for invasive breast cancer. Ann Intern Med 137: 798-804.

Chiarelli AM, Kirsh VA, Klar NS, Shumak R, Jong R, Fishell E, Yaffe MJ and Boyd NF (2006): Influence of patterns of hormone replacement therapy use and mammographic density on breast cancer detection. Cancer Epidemiol Biomarkers Prev 15: 1856-62.

Chlebowski RT, Hendrix SL, Langer RD, Stefanick ML, Gass M, Lane D, Rodabough RJ, Gilligan MA, Cyr MG, Thomson CA, Khandekar J, Petrovitch H and McTiernan A (2003): Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. JAMA 289: 3243-53.

Ciatto S, Visioli C, Paci E and Zappa M (2004): Breast density as a determinant of interval cancer at mammographic screening. Br J Cancer 90: 393-6.

Clamp A, Danson S and Clemons M (2002): Hormonal risk factors for breast cancer: identification, chemoprevention, and other intervention strategies. Lancet Oncol 3: 611-9.

Clemons M and Goss P (2001): Estrogen and the risk of breast cancer. N Engl J Med 344: 276-85.

Cuzick J, Edwards R and Segnan N (1997): Adjusting for non-compliance and contamination in randomized clinical trials. Stat Med 16: 1017-29.

Cuzick J, Powles T, Veronesi U, Forbes J, Edwards R, Ashley S and Boyle P (2003): Overview of the main outcomes in breast-cancer prevention trials. Lancet 361: 296-300.

Day NE, Williams DR and Khaw KT (1989): Breast cancer screening programmes: the development of a monitoring and evaluation system. Br J Cancer 59: 954-8.

Day NE (2005): Overdiagnosis and breast cancer screening. Breast Cancer Res 7: 228-9.

de Koning HJ (2003): Why improvement in survival of screen-detected cases is not necessarily equivalent to benefit? The Breast 12: 299-301.

de Waard F and Trichopoulos D (1988): A unifying concept of the aetiology of breast cancer. Int J Cancer 41: 666-9.

Dean PB and Pamilo M (1999): Screening mammography in Finland-1.5 million examinations with 97 percent specificity. Mammography Working Group, Radiological Society of Finland. Acta Oncol 38 Suppl 13: 47-54.

del Moral Aldaz A, Aupee M, Batal-Steil S, Cecchini S, Chamberlain J, Ciatto S, Elizaga NA, Gairard B, Grazzini G, Guldenfels C, Herity PA, Kosmidis PA, Lynge E, Paci E, Renaud R, Robra PB, Rodrigues VL, Sancho-Garnier H, Schaffer DP, da Silva DP, Tubiana M, del Turco MR, van Veen WA and Vandenbroucke A (1994): Cancer screening in the European Union. Eur J Cancer 30A: 860-72.

dos Santos Silva I (1999): Cancer epidemiology: principles and methods, Second edn.: International Agency for Research on Cancer IARC

Douglas-Jones AG, Gupta SK, Attanoos RL, Morgan JM and Mansel RE (1996): A critical appraisal of six modern classifications of ductal carcinoma in situ of the breast (DCIS): correlation with grade of associated invasive carcinoma. Histopathology 29: 397-409.

Duffy SW, Tabar L, Fagerberg G, Gad A, Grontoft O, South MC and Day NE (1991): Breast screening, prognostic factors and survival-results from the Swedish two county study. Br J Cancer 64: 1133-8.

Duffy SW, Tabar L, Chen HH, Holmqvist M, Yen MF, Abdsalah S, Epstein B, Frodis E, Ljungberg E, Hedborg-Melander C, Sundbom A, Tholin M, Wiege M, Akerlund A, Wu HM, Tung TS, Chiu YH, Chiu CP, Huang CC, Smith RA, Rosen M, Stenbeck M and Holmberg L (2002): The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties. Cancer 95: 458-69.

Duffy SW, Tabar L, Vitak B, Day NE, Smith RA, Chen HH and Yen MF (2003): The relative contributions of screen-detected in situ and invasive breast carcinomas in reducing mortality from the disease. Eur J Cancer 39: 1755-60.

Duffy SW, McCann J, Godward S, Gabe R and Warwick J (2006): Some issues in screening for breast and other cancers. J Med Screen 13 Suppl 1: S28-34.

Dupont WD, Parl FF, Hartmann WH, Brinton LA, Winfield AC, Worrell JA, Schuyler PA and Plummer WD (1993): Breast cancer risk associated with proliferative breast disease and atypical hyperplasia. Cancer 71: 1258-65.

Dupont WD, Page DL, Parl FF, Vnencak-Jones CL, Plummer WD, Jr., Rados MS and Schuyler PA (1994): Long-term risk of breast cancer in women with fibroadenoma. N Engl J Med 331: 10-5.

Easton DF, Bishop DT, Ford D and Crockford GP (1993): Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. Am J Hum Genet 52: 678-701.

Elston CW, Ellis IO and Pinder SE (1999): Pathological prognostic factors in breast cancer. Crit Rev Oncol Hematol 31: 209-23.

Ewertz M, Duffy SW, Adami HO, Kvale G, Lund E, Meirik O, Mellemgaard A, Soini I and Tulinius H (1990): Age at first birth, parity and risk of breast cancer: a meta-analysis of 8 studies from the Nordic countries. Int J Cancer 46: 597-603.

FCR Finnish Cancer Registry (2005): Cancer in Finland 2002-2003. Cancer Society of Finland Publication No 66: 88 p.

FCR Finnish Cancer Registry (2007): Cancer in Finland 2004 and 2005. Cancer Society of Finland Publication No 72: 88 p.

Fentiman IS, Fourquet A and Hortobagyi GN (2006): Male breast cancer. Lancet 367: 595-604.

Fielder HM, Warwick J, Brook D, Gower-Thomas K, Cuzick J, Monypenny I and Duffy SW (2004): A case-control study to estimate the impact on breast cancer death of the breast screening programme in Wales. J Med Screen 11: 194-8.

Fletcher SW, Black W, Harris R, Rimer BK and Shapiro S (1993): Report of the International Workshop on Screening for Breast Cancer. J Natl Cancer Inst 85: 1644-56.

Fracheboud J, de Koning HJ, Beemsterboer PM, Boer R, Hendriks JH, Verbeek AL, van Ineveld BM, de Bruyn AE and van der Maas PJ (1998): Nation-wide breast cancer screening in The Netherlands: results of initial and subsequent screening 1990-1995. National Evaluation Team for Breast Cancer Screening. Int J Cancer 75: 694-8.

Fracheboud J, de Koning HJ, Beemsterboer PM, Boer R, Verbeek AL, Hendriks JH, van Ineveld BM, Broeders MJ, de Bruyn AE and van der Maas PJ (1999): Interval cancers in the Dutch breast cancer screening programme. Br J Cancer 81: 912-7.

Fracheboud J, de Koning HJ, Boer R, Groenewoud JH, Verbeek AL, Broeders MJ, van Ineveld BM, Hendriks JH, de Bruyn AE, Holland R and van der Maas PJ (2001): Nationwide breast cancer screening programme fully implemented in The Netherlands. Breast 10: 6-11.

Fracheboud J, Otto SJ, van Dijck JA, Broeders MJ, Verbeek AL and de Koning HJ (2004): Decreased rates of advanced breast cancer due to mammography screening in The Netherlands. Br J Cancer 91: 861-7.

Fracheboud J, Groenewoud JH, Boer R, Draisma G, de Bruijn AE, Verbeek AL and de Koning HJ (2006): Seventy-five years is an appropriate upper age limit for population-based mammography screening. Int J Cancer 118: 2020-5.

Freedman DA, Petitti DB and Robins JM (2004): On the efficacy of screening for breast cancer. Int J Epidemiol 33: 43-55.

Frisell J, Glas U, Hellstrom L and Somell A (1986): Randomized mammographic screening for breast cancer in Stockholm. Design, first round results and comparisons. Breast Cancer Res Treat 8: 45-54.

Galea MH, Blamey RW, Elston CE and Ellis IO (1992): The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res Treat 22: 207-19.

Gayther SA, Mangion J, Russell P, Seal S, Barfoot R, Ponder BA, Stratton MR and Easton D (1997): Variation of risks of breast and ovarian cancer associated with different germline mutations of the BRCA2 gene. Nat Genet 15: 103-5.

Giordano L, Giorgi D, Fasolo G, Segnan N and Del Turco MR (1996): Breast cancer screening: characteristics and results of the Italian programmes in the Italian group for planning and evaluating breast cancer screening programmes (GISMa). Tumori 82: 31-7.

Giordano L, Giorgi D, Piccini P, Ventura L, Stefanini V, Senore C, Paci E and Segnan N (2007): Time trends of process and impact indicators in Italian mammography screening programmes--1996-2004. Epidemiol Prev 31: 21-32.

Giorgi D, Giordano L, Ventura L, Frigerio A, Paci E and Zappa M (2007): Mammography screening in Italy: 2004 survey and 2005 preliminary data. Epidemiol Prev 31: 7-20.

Gotzsche PC and Olsen O (2000): Is screening for breast cancer with mammography justifiable? Lancet 355: 129-34.

Gotzsche PC, Thornton H and Jorgensen KJ (2005): Reduction in mortality from breast cancer: presentation of benefits and harms needs to be balanced. BMJ 330: 1024; author reply 1025.

Gotzsche PC and Nielsen M (2006): Screening for breast cancer with mammography. Cochrane Database Syst Rev: CD001877.

Grant EC (2005): Reduction in mortality from breast cancer: fall in use of hormones could have reduced breast cancer mortality. BMJ 330: 1024; author reply 1025.

Habbema JD, Tan SY and Cronin KA (2006): Impact of mammography on U.S. breast cancer mortality, 1975-2000: are intermediate outcome measures informative? J Natl Cancer Inst Monogr: 105-11.

Hakama M, Pukkala E, Kallio M, Godenhjelm K and Svinhufvud U (1995): Effectiveness of screening for breast cancer in women under 50 years at entry: the Kotka pilot project in Finland. Int J Cancer 63: 55-7.

Hakama M, Pukkala E, Heikkila M and Kallio M (1997): Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. BMJ 314: 864-7.

Hakama M, Pukkala E, Soderman B and Day N (1999): Implementation of screening as a public health policy: issues in design and evaluation. J Med Screen 6: 209-16.

Hakama M, Auvinen A, Day NE and Miller AB (2007): Sensitivity in cancer screening. J Med Screen 14: 174-7.

Hemminki E (2004): Opposition to unpopular research results: Finnish professional reactions to the WHI findings. Health Policy 69: 283-91.

Hermon C and Beral V (1996): Breast cancer mortality rates are levelling off or beginning to decline in many western countries: analysis of time trends, age-cohort and age-period models of breast cancer mortality in 20 countries. Br J Cancer 73: 955-60.

Hofvind S, Thoresen S and Tretli S (2004a): The cumulative risk of a false-positive recall in the Norwegian Breast Cancer Screening Program. Cancer 101: 1501-7.

Hofvind S, Wang H and Thoresen S (2004b): Do the results of the process indicators in the Norwegian Breast Cancer Screening Program predict future mortality reduction from breast cancer? Acta Oncol 43: 467-73.

Hofvind S, Bjurstam N, Sorum R, Bjorndal H, Thoresen S and Skaane P (2006a): Number and characteristics of breast cancer cases diagnosed in four periods in the screening interval of a biennial population-based screening programme. J Med Screen 13: 192-6.

Hofvind S, Moller B, Thoresen S and Ursin G (2006b): Use of hormone therapy and risk of breast cancer detected at screening and between mammographic screens. Int J Cancer 118: 3112-7.

Hofvind S, Geller B, Vacek PM, Thoresen S and Skaane P (2007a): Using the European guidelines to evaluate the Norwegian Breast Cancer Screening Program. Eur J Epidemiol 22: 447-55.

Hofvind S, Sorum R and Thoresen S (2007b): Incidence and tumor characteristics of breast cancer diagnosed before and after implementation of a population-based screening-program. Acta Oncol: 1-7.

Hristova L and Hakama M (1997): Effect of screening for cancer in the Nordic countries on deaths, cost and quality of life up to the year 2017. Acta Oncol 36 Suppl 9: 1-60.

Jensen RA, Page DL, Dupont WD and Rogers LW (1989): Invasive breast cancer risk in women with sclerosing adenosis. Cancer 64: 1977-83.

Jonsson H, Tornberg S, Nystrom L and Lenner P (2000): Service screening with mammography in Sweden-evaluation of effects of screening on breast cancer mortality in age group 40-49 years. Acta Oncol 39: 617-23.

Jonsson H, Nystrom L, Tornberg S and Lenner P (2001): Service screening with mammography of women aged 50-69 years in Sweden: effects on mortality from breast cancer. J Med Screen 8: 152-60.

Jonsson H, Tornberg S, Nystrom L and Lenner P (2003a): Service screening with mammography of women aged 70-74 years in Sweden. Effects on breast cancer mortality. Cancer Detect Prev 27: 360-9.

Jonsson H, Nystrom L, Tornberg S, Lundgren B and Lenner P (2003b): Service screening with mammography. Long-term effects on breast cancer mortality in the county of Gavleborg, Sweden. Breast 12: 183-93.

Jonsson H, Johansson R and Lenner P (2005): Increased incidence of invasive breast cancer after the introduction of service screening with mammography in Sweden. Int J Cancer 117: 842-7.

Jonsson H, Bordas P, Wallin H, Nystrom L and Lenner P (2007): Service screening with mammography in Northern Sweden: effects on breast cancer mortality - an update. J Med Screen 14: 87-93.

Kavanagh AM, Cawson J, Byrnes GB, Giles GG, Marr G, Tong B, Gertig DM and Hopper JL (2005): Hormone replacement therapy, percent mammographic density, and sensitivity of mammography. Cancer Epidemiol Biomarkers Prev 14: 1060-4.

Kerlikowske K, Grady D, Barclay J, Sickles EA and Ernster V (1996): Effect of age, breast density, and family history on the sensitivity of first screening mammography. JAMA 276: 33-8.

Kerlikowske K, Barclay J, Grady D, Sickles EA and Ernster V (1997): Comparison of risk factors for ductal carcinoma in situ and invasive breast cancer. J Natl Cancer Inst 89: 76-82.

Key TJ, Verkasalo PK and Banks E (2001): Epidemiology of breast cancer. Lancet Oncol 2: 133-40.

Klemi P, Pylkkanen L, Toikkanen S, Immonen-Raiha P, Räsänen O and Parvinen I (1998): Turun mammografiaseulonnan kymmenen ensimmäisen vuoden tuloksia. Suomen Lääkäril 26: 2793-7.

Kärkkäinen A, Lauslahti K, Lehtimäki V, Schoultz E, Aine R and Hakama M (1987): Tampereen rintasyöpäprojektin II kierros: intervallisyöpien ongelma. Suomen Lääkäril 42: 3128-32.

Lauslahti K, Kärkkäinen A, Hartikainen R-L, Vekara L, Aine R, Auvinen O, Sulkava T, Schoultz E and Hakama M (1986): Tampereen rintasyöpäprojekti. Suomen Lääkäril 41: 1814-8.

Lenner P and Jonsson H (1997): Excess mortality from breast cancer in relation to mammography screening in northern Sweden. J Med Screen 4: 6-9.

Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A and Hemminki K (2000): Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343: 78-85.

Lidbrink EK, Tornberg SA, Azavedo EM, Frisell JO, Hjalmar ML, Leifland KS, Sahlstedt TB and Skoog L (1994): The general mammography screening program in Stockholm. Organisation and first-round results. Acta Oncol 33: 353-8.

Lynge E, Olsen AH, Fracheboud J and Patnick J (2003): Reporting of performance indicators of mammography screening in Europe. Eur J Cancer Prev 12: 213-22.

MacKenzie TA, Titus-Ernstoff L, Vacek PM, Geller B, Weiss JE, Goodrich ME and Carney PA (2007): Breast density in relation to risk of ductal carcinoma in situ of the breast in women undergoing screening mammography. Cancer Causes Control 18: 939-45.

Mandelson MT, Oestreicher N, Porter PL, White D, Finder CA, Taplin SH and White E (2000): Breast density as a predictor of mammographic detection: comparison of interval- and screen-detected cancers. J Natl Cancer Inst 92: 1081-7.

Marshall LM, Hunter DJ, Connolly JL, Schnitt SJ, Byrne C, London SJ and Colditz GA (1997)**:** Risk of breast cancer associated with atypical hyperplasia of lobular and ductal types. Cancer Epidemiol Biomarkers Prev 6**:** 297-301.

McCann J, Duffy S and Day N (2001)**:** Predicted long-term mortality reduction associated with the second round of breast screening in East Anglia. Br J Cancer 84**:** 423-8.

McCormack VA and dos Santos Silva I (2006)**:** Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 15**:** 1159-69.

McPherson K, Steel CM and Dixon JM (2000)**:** ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. BMJ 321**:** 624-8.

Miller AB, To T, Baines CJ and 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**:** 1490-9.

Miller AB, To T, Baines CJ and 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**:** 305-12.

Morrison A (1991)**:** Intermediate Determinants of Mortality in the Evaluation of Screening. Int J Epidemiol 20**:** 642-50.

Moss SM, Michel M, Patnick J, Johns L, Blanks R and Chamberlain J (1995)**:** Results from the NHS breast screening programme 1990-1993. J Med Screen 2**:** 186-90.

Moss SM, Cuckle H, Evans A, Johns L, Waller M and Bobrow L (2006)**:** Effect of mammographic screening from age 40 years on breast cancer mortality at 10 years' follow-up: a randomised controlled trial. Lancet 368**:** 2053-60.

Njor SH, Olsen AH, Bellstrom T, Dyreborg U, Bak M, Axelsson C, Graversen HP, Schwartz W and Lynge E (2003)**:** Mammography screening in the county of Fyn. November 1993-December 1999. APMIS Suppl**:** 1-33.

Nystrom L, Larsson LG, Rutqvist LE, Lindgren A, Lindqvist M, Ryden S, Andersson I, Bjurstam N, Fagerberg G, Frisell J and et al. (1995)**:** 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 34**:** 145-52.

Nystrom L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B and Rutqvist LE (2002)**:** Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 359**:** 909-19.

Olsen AH, Jensen A, Njor SH, Villadsen E, Schwartz W, Vejborg I and Lynge E (2003)**:** Breast cancer incidence after the start of mammography screening in Denmark. Br J Cancer 88**:** 362-5.

Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen MB, Tange UB, Blichert-Toft M, Rank F, Mouridsen H and Lynge E (2005a)**:** A model for determining the effect of mammography service screening. Acta Oncol 44**:** 120-8.

Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen MB, Tange UB, Blichert-Toft M, Rank F, Mouridsen H and Lynge E (2005b): Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study. BMJ 330: 220.

Olsen AH, Agbaje OF, Myles JP, Lynge E and Duffy SW (2006): Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. Breast J 12: 338-42.

Olsen O and Gotzsche PC (2001): Cochrane review on screening for breast cancer with mammography. Lancet 358: 1340-2.

Otten JD, Karssemeijer N, Hendriks JH, Groenewoud JH, Fracheboud J, Verbeek AL, de Koning HJ and Holland R (2005): Effect of recall rate on earlier screen detection of breast cancers based on the Dutch performance indicators. J Natl Cancer Inst 97: 748-54.

Otto SJ, Fracheboud J, Looman CW, Broeders MJ, Boer R, Hendriks JH, Verbeek AL and de Koning HJ (2003): Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. Lancet 361: 1411-7.

Paci E, Duffy SW, Giorgi D, Zappa M, Crocetti E, Vezzosi V, Bianchi S and del Turco MR (2002a): Quantification of the effect of mammographic screening on fatal breast cancers: The Florence Programme 1990-96. Br J Cancer 87: 65-9.

Paci E, Giorgi D, Bianchi S, Vezzosi V, Zappa M, Crocetti E and Rosselli del Turco M (2002b): Assessment of the early impact of the population-based breast cancer screening programme in Florence (Italy) using mortality and surrogate measures. Eur J Cancer 38: 568-73.

Paci E, Warwick J, Falini P and Duffy SW (2004): Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? J Med Screen 11: 23-7.

Parvinen I, Helenius H, Pylkkanen L, Anttila A, Immonen-Raiha P, Kauhava L, Rasanen O and Klemi PJ (2006): Service screening mammography reduces breast cancer mortality among elderly women in Turku. J Med Screen 13: 34-40.

Perry N, Broeders M, de Wolf C, Tornberg S, Holland R and von Karsa L (2006): European Commission. European Guidelines for Quality Assurance in Breast Cancer Screening and Diagnosis, 4th edition edn.: Luxembourg: Office for Official Publications of the European Communities

Rintala P, Pukkala E, Laara E and Vihko V (2003): Physical activity and breast cancer risk among female physical education and language teachers: a 34-year follow-up. Int J Cancer 107: 268-70.

Roberts MM, Alexander FE, Anderson TJ, Forrest AP, Hepburn W, Huggins A, Kirkpatrick AE, Lamb J, Lutz W and Muir BB (1984): The Edinburgh randomised trial of screening for breast cancer: description of method. Br J Cancer 50: 1-6.

Roberts MM, Alexander FE, Anderson TJ, Chetty U, Donnan PT, Forrest P, Hepburn W, Huggins A, Kirkpatrick AE, Lamb J and et al. (1990): Edinburgh trial of screening for breast cancer: mortality at seven years. Lancet 335: 241-6.

Rosen PP and Groshen S (1990): Factors influencing survival and prognosis in early breast carcinoma (T1N0M0-T1N1M0). Assessment of 644 patients with median follow-up of 18 years. Surg Clin North Am 70: 937-62.

Rutanen E-M and Ylikorkala O (2004) Vaihdevuosien hormonihoito. Kapseli 33. 62 p. Kansaneläkelaitos.

Sasieni P (2003): Evaluation of the UK breast screening programmes. Ann Oncol 14: 1206-8.

Seppanen J, Heinavaara S, Anttila A, Sarkeala T, Virkkunen H and Hakulinen T (2006): Effects of different phases of an invitational screening program on breast cancer incidence. Int J Cancer 119: 920-4.

Shapiro S, Strax P and Venet L (1971): Periodic breast cancer screening in reducing mortality from breast cancer. JAMA 215: 1777-85.

Shapiro S (1997): Periodic screening for breast cancer: the HIP Randomized Controlled Trial. Health Insurance Plan. J Natl Cancer Inst Monogr: 27-30.

Shapiro S, Coleman EA, Broeders M, Codd M, de Koning H, Fracheboud J, Moss S, Paci E, Stachenko S and Ballard-Barbash R (1998): Breast cancer screening programmes in 22 countries: current policies, administration and guidelines. International Breast Cancer Screening Network (IBSN) and the European Network of Pilot Projects for Breast Cancer Screening. Int J Epidemiol 27: 735-42.

Silverstein MJ, Poller DN, Waisman JR, Colburn WJ, Barth A, Gierson ED, Lewinsky B, Gamagami P and Slamon DJ (1995): Prognostic classification of breast ductal carcinoma-in-situ. Lancet 345: 1154-7.

Silverstein MJ, Lagios MD, Craig PH, Waisman JR, Lewinsky BS, Colburn WJ and Poller DN (1996): A prognostic index for ductal carcinoma in situ of the breast. Cancer 77: 2267-74.

Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM and Chen TH (2004): The randomized trials of breast cancer screening: what have we learned? Radiol Clin North Am 42: 793-806, v.

SOSSEG Swedish Organised Service Screening Evaluation Group (2006a): Reduction in breast cancer mortality from the organised service screening with mammography: 2. Validation with alternative analytic methods. Cancer Epidemiol Biomarkers Prev 15: 52-6.

SOSSEG Swedish Organised Service Screening Evaluation Group (2006b): Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. Cancer Epidemiol Biomarkers Prev 15: 45-51.

SOSSEG Swedish Organised Service Screening Evaluation Group (2007): Effect of mammographic service screening on stage at presentation of breast cancers in Sweden. Cancer 109: 2205-12.

Struewing JP, Hartge P, Wacholder S, Baker SM, Berlin M, McAdams M, Timmerman MM, Brody LC and Tucker MA (1997): The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. N Engl J Med 336: 1401-8.

Tabar L and Gad A (1981): Screening for breast cancer: the Swedish trial. Radiology 138: 219-22.

Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O, Ljungquist U, Lundstrom B, Manson JC, Eklund G, Day N and Pettersson F (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: 829-32.

Tabar L, Faberberg G, Day NE and Holmberg L (1987): What is the optimum interval between mammographic screening examinations? An analysis based on the latest results of the Swedish two-county breast cancer screening trial. Br J Cancer 55: 547-51.

Tabar L, Vitak B, Chen HH, Duffy SW, Yen MF, Chiang CF, Krusemo UB, Tot T and Smith RA (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: 625-51.

Tabar L, Yen MF, Vitak B, Chen HH, Smith RA and Duffy SW (2003): Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening. Lancet 361: 1405-10.

Threlfall AG, Collins S and Woodman CB (2003): Impact of NHS breast screening on advanced disease and mortality from breast cancer in the North West of England. Br J Cancer 89: 77-80.

Thurfjell EL and Lindgren JA (1994): Population-based mammography screening in Swedish clinical practice: prevalence and incidence screening in Uppsala County. Radiology 193: 351-7.

Tjonneland A, Christensen J, Olsen A, Stripp C, Thomsen BL, Overvad K, Peeters PH, van Gils CH, Bueno-de-Mesquita HB, Ocke MC, Thiebaut A, Fournier A, Clavel-Chapelon F, Berrino F, Palli D, Tumino R, Panico S, Vineis P, Agudo A, Ardanaz E, Martinez-Garcia C, Amiano P, Navarro C, Quiros JR, Key TJ, Reeves G, Khaw KT, Bingham S, Trichopoulou A, Trichopoulos D, Naska A, Nagel G, Chang-Claude J, Boeing H, Lahmann PH, Manjer J, Wirfalt E, Hallmans G, Johansson I, Lund E, Skeie G, Hjartaker A, Ferrari P, Slimani N, Kaaks R and Riboli E (2007): Alcohol intake and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition (EPIC). Cancer Causes Control 18: 361-73.

Tornberg S, Carstensen J, Hakulinen T, Lenner P, Hatschek T and Lundgren B (1994): Evaluation of the effect on breast cancer mortality of population based mammography screening programmes. J Med Screen 1: 184-7.

Tornberg S, Codd M, Rodrigues V, Segnan N and Ponti A (2005): Ascertainment and evaluation of interval cancers in population-based mammography screening programmes: a collaborative study in four European centres. J Med Screen 12: 43-9.

Twombly R (2007): American researchers question effect of Scandinavian mammography debate. J Natl Cancer Inst 99: 1216-7.

Vainio H and Bianchini F (2002): IARC Handbooks on Cancer Prevention. Breast cancer screening. Vol. 7. WHO/IARC Press.

Vainio H, Kaaks R and Bianchini F (2002): Weight control and physical activity in cancer prevention: international evaluation of the evidence. Eur J Cancer Prev 11 Suppl 2: S94-100.

Vejborg I, Olsen AH, Jensen MB, Rank F, Tange UB and Lynge E (2002): Early outcome of mammography screening in Copenhagen 1991-99. J Med Screen 9: 115-9.

Vitak B, Stal O, Manson JC, Thomas BA, Arnesson LG, Ekelund L, Mare K, Nordenskjold B, Kallstrom AC and Bang H (1997): Interval cancers and cancers in non-attenders in the Ostergotland Mammographic Screening Programme. Duration between screening and diagnosis, S-phase fraction and distant recurrence. Eur J Cancer 33: 1453-60.

von Euler-Chelpin M, Olsen AH, Njor S, Vejborg I, Schwartz W and Lynge E (2006): Women's patterns of participation in mammography screening in Denmark. Eur J Epidemiol 21: 203-9.

Voogd AC and Coebergh JW (2003): Mortality reduction by breast-cancer screening. Lancet 362: 245-6.

Wang H, Bjurstam N, Bjorndal H, Braaten A, Eriksen L, Skaane P, Vitak B, Hofvind S and Thoresen SO (2001a): Interval cancers in the Norwegian breast cancer screening program: frequency, characteristics and use of HRT. Int J Cancer 94: 594-8.

Wang H, Karesen R, Hervik A and Thoresen SO (2001b): Mammography screening in Norway: results from the first screening round in four counties and cost-effectiveness of a modeled nationwide screening. Cancer Causes Control 12: 39-45.

Weisburger JH (1997): Dietary fat and risk of chronic disease: mechanistic insights from experimental studies. J Am Diet Assoc 97: S16-23.

Welch HG and Black WC (1997): Using autopsy series to estimate the disease "reservoir" for ductal carcinoma in situ of the breast: how much more breast cancer can we find? Ann Intern Med 127: 1023-8.
Woodman CB, Threlfall AG, Boggis CR and Prior P (1995): Is the three year breast screening interval too long? Occurrence of interval cancers in NHS breast screening programme's north western region. BMJ 310: 224-6.

www-dep.iarc.fr/ (01.04.2008)

www.cancerregistry.fi/eng/statistics (01.04.2008)

www.cancerscreening.nhs.uk (01.04.2008)

Yankaskas BC, Klabunde CN, Ancelle-Park R, Renner G, Wang H, Fracheboud J, Pou G and Bulliard JL (2004): International comparison of performance measures for screening mammography: can it be done? J Med Screen 11: 187-93.

Yen MF, Tabar L, Vitak B, Smith RA, Chen HH and Duffy SW (2003): Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. Eur J Cancer 39: 1746-54.

Zackrisson S, Andersson I, Manjer J and Janzon L (2004): Non-attendance in breast cancer screening is associated with unfavourable socio-economic circumstances and advanced carcinoma. Int J Cancer 108: 754-60.

Zackrisson S, Andersson I, Janzon L, Manjer J and Garne JP (2006)**:** Rate of over-diagnosis of breast cancer 15 years after end of Malmo mammographic screening trial: follow-up study. BMJ 332**:** 689-92.

Zahl PH, Strand BH and Maehlen J (2004)**:** Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. BMJ 328**:** 921-4.

Zahl PH and Maehlen J (2005)**:** Reduction in mortality from breast cancer: decrease with screening was marked in younger age group. BMJ 330**:** 1024; author reply 1025.

Zelen M and Feinleib M (1969): On the theory of screening for chronic diseases. Biometrika 56: 601-614.

APPENDIX

The following definitions were made for the three models. Let i=1,…,6 denote the categorical variable five-year age group at death (50-54, 55-59, 60-64, 65-69, 70-74, 75-79), r=1,2,3 denote the categorical variable centre category (low recall rate, intermediate recall rate, high recall rate), p=1,2,3 denote the categorical variable screening policy (regular invitation at ages 50-59, regular invitation at ages 50-69, irregular invitation at ages 50-69), y denote the numerical variable calendar year of death, and n=1,…,12 denote the numerical variable calendar time in years since the beginning of the two periods (1974-1985, 1992-2003). In the period 1974-1985, the variable n=y-1973, and in the period 1992-2003, n=y-1991. Thus n is a function of y, i.e. n=n(y).

Let us define further the categorical screening effect variable o=1,…,6 for the screening invitees for model 1: o=1 in the calendar period 1974-1985, and o=2,…,6 for i=1,…,5 in the calendar period 1992-2003. Let us also define the categorical screening effect variable k=1,2,3 for the whole female population for model 2: k=1 in the calendar period 1974-1985, k=2 in the calendar period 1992-2003 for i=3,4, and k=3 in the calendar period 1992-2003 for i=5,6. Finally, let us define the categorical screening effect variable s=1,…,9 for the screened and non-screened women for model 3: s=1 in the calendar period 1974-1985, s=2,…,5 for i=3,..,6 in the calendar period 1992-2003 for the non-screened women, and s=6,…,9 for i=3,…,6 in the calendar period 1992-2003 for the screened women.

Among the invitees (i=1,…,5), the model for the incidence-based mortality m_{iry} in the age group i, in the centre category r, and the calendar year y can then be written as $log(m_{iry})=\alpha+\beta_i$i+$\gamma$n(y)+$\delta_i$in(y)+$\varepsilon_r$r+$\zeta_o$o (model 1). Among the whole female population (i=3,…,6), the model for the incidence-based mortality m_{ipy} in the age group i, in the policy

category p, and the calendar year y is expressed as

$log(m_{ipy})=\alpha+\beta_i i+\gamma n(y)+\delta_i in(y)+\varepsilon_p p+\zeta_k k+\eta_{pk} pk$ (model 2). Among the screened and the non-screened women (i=3,…,6), the model for the incidence-based mortality is

$log(m_{ipys})=\alpha+\beta_i i+\gamma n(y)+\delta_i in(y)+\varepsilon_p p+\zeta_s s+\varepsilon_p p+\zeta_s s+\eta_{ps} ps$ (model 3).

The expected breast cancer mortality rates without screening for the calendar period 1992-2003 were estimated by excluding the variables related to the screening effects, i.e. by using $exp(\alpha+\beta_i i+\gamma n(y)+\delta_i in(y)+\varepsilon_r r)$ (model 1), and $exp(\alpha+\beta_i i+\gamma n(y)+\delta_i in(y)+\varepsilon_p p)$ (models 2 and 3) with the estimates from the respective models.

ORIGINAL PAPER

Validity of process indicators of screening for breast cancer to predict mortality reduction

T Sarkeala, A Anttila, I Saarenmaa, M Hakama

J Med Screen 2005;**12**:33–37

See end of article for authors' affiliations

Correspondence to: T Sarkeala, Finnish Cancer Registry, Liisankatu 21 B, 00170 Helsinki, Finland; tytti.sarkeala @cancer.fi

Accepted for publication 18 October 2004

The aim of the study was to empirically assess the acceptable levels of process indicators as described in the European Community Guidelines using materials from the mammography service screening programmes. The Finnish programme was evaluated for effectiveness with a prior estimate of 0.74 for RR in Finland and 0.81 in Helsinki. Hence, the Finnish programme was likely to be somewhat less effective in terms of reduction in mortality than implied on the basis of early randomized trials, but probably approaching the same level of effectiveness. Finland therefore provides background data on the applicability of the process indicators that are indicators of performance and surrogates for effectiveness. The performance data on 10 Finnish screening centres at subsequent screens were used. These centres invited 687,000 women aged 50–64 years in 1991–2000. The mean compliance was 93% and the corresponding recall rate was 2.3%. The benign to malignant biopsy ratio was 0.43:1. The average breast cancer detection rate was 0.36%, 2.1 compared with the background incidence. The proportion of screen-detected stage II+ cancers was 26%. Most, but not all, of these process indicators met the desirable reference values of the European Community. The specific criteria of the European Community on stage distribution, rates of screen-detected cancers by stage and detection rate to background incidence may need reconsideration.

In this study we estimate the process indicators of the mammography screening that is run as a public health policy in Finland and compare them with the European Community criteria and with indicators derived from the Netherlands. The study is based on data at individual level from 10 screening centres of the Cancer Society of Finland (CSF) in 1991–2000. The aim is to have empirical verification on the acceptable levels of process indicators at subsequent screens, as described in the European Community Guidelines.[1]

MATERIALS AND METHODS

The onset of the Finnish programme was based on recommendations given by the Finnish National Board of Health in 1986. A Bylaw on Public Health in 1992 entitled municipalities to offer breast cancer screening every second year for women aged 50–59 years; screenings for women aged 60–69 years women remained optional. Individual municipalities were entitled to decide the organizational details of breast cancer screening activity.

Our study material is based on 10 CSF screening centres that covered two-thirds of approximately 460 Finnish municipalities during the first five years of the screening programme.[2,3] The material covered approximately half of the Finnish breast cancer screening target population in 1991–2000. All the screening centres invited women aged 50–64 years.[4]

Women to be screened in Finland are identified from the national population registry. All target women with permanent addresses in the municipality are invited independently, whether they have prevalent cases or symptoms of breast cancer or attended breast cancer screening earlier. Letters of invitation, usually with appointment times, are personally sent to each woman. The non-attenders are sent one reminder. All of the attenders are informed of their screening results by a letter. If there is a need for further examinations, women are given an appointment by phone or by letter.

The screening examinations are free of charge for all of the participants. The screening test is two-view screen film mammography (craniocaudal and mediolateral oblique), which two radiologists interpret independently. If either reading is positive, simultaneous reading will be performed. Women positive at the simultaneous reading (not read as negative by both of the radiologists) are recalled for further examinations. If a diagnosis of breast cancer cannot be excluded by recall examinations (including cytology), women are referred for surgery. The further histological confirmation and treatment are performed in local or central hospitals depending on the treatment facilities.

Registration of the nationwide breast cancer screening programme is centrally maintained at the Mass Screening Registry of Finland, established in 1968, which is part of the Finnish Cancer Registry. Information on screening invitations, screening attendance, screening results, recalls, histologically confirmed findings and treatment is recorded. Screening results are linked with the national population registry and cancer registration by the Mass Screening Registry using the unique personal identifier as a key.

The data on breast cancer screening parameters were collected at individual level. The invitation files from 1987 to 1990 were sufficiently complete and were used to define the number of screening invitational rounds. Only subsequent screens (i.e. screens in women with more than

one screen) were included. This criterion was chosen in order to avoid the problems of a large prevalence pool of positive findings with the first personal screen. Due to potential misclassification related to under-registration, the women with data on screening aged over 50 but data on previous screen(s) unknown ($n = 18,825$) were excluded from the analysis.[4]

Intermediate indicators (process parameters) are developed to measure the performance of screening and to be surrogates for the ultimate effect of screening on breast cancer mortality before the actual mortality evaluation is possible.[5] Coverage and attendance to screening indicate the potential effectiveness of the overall programme, the stage distribution of screen-detected cancers, the potential for reduction in the absolute rate of advanced cancers, and the rate of advanced cancers as an early surrogate of mortality reduction. Positive predictive value (PPV) represents the harm in terms of specificity. PPV of mammography and of histological confirmation (or benign to malignant biopsy ratio [B:M ratio]) are usually distinguished. Other frequently reported process indicators are the recall rate, the rate of histological confirmation, and the breast cancer detection rate.[6]

The European Community has created criteria or desirable levels for the values of the process indicators of screening, according to which the European breast cancer screening programmes can be evaluated.[1] In the Netherlands, screening for breast cancer is run as a public health policy and the protocol is similar to that in Finland.[7–9] We therefore compared process indicators of the Finnish programme with the European standards and with the indicators in the Netherlands, where there is good evidence for the effectiveness of the programme in terms of mortality reduction.[10]

The detection rates in this study were compared with the background incidence in 1980–1986, the period preceding the mammography programme in Finland. The increase in incidence from 1980–1986 (pre-screening) to 1991–2000 (screening) was estimated by average corrections of 35.1%. This estimated increase was based on the weighted estimate of age-specific increase in the incidence from 1980–1986, 18.9%, and the median year of pre-screening (1983.5) and screening (1996.5) periods. The background incidence was derived for each of the mammographic centres on the basis of target population of the centre, but the same 35.1% increase in risk was assumed for all of the centres.

Tumours were classified by stage. The *in situ* cancers represented stage 0. Tumours less than 2 cm in diameter and without regional or distant lymph node metastasis represented stage I. Tumours more than 2 cm in diameter with or without regional lymph node metastasis and/or distant metastasis represented stage II+. About 93% of carcinomas had pTNM code.

RESULTS

The screening centres in the study covered 687,000 invitations and 635,000 visits in total. The overall participation rate was 92.5% (Table 1). In addition, the centre-specific participation rates, 90–95%, were well above the level of 75% desired by the European Community. The

Table 1 Process indicators by five-year age groups at subsequent screens in 1991–2000 at CSF screening centres

	50–54 years	55–59 years	60–64 years	All
Invitations	263,530	304,197	118,938	686,665
Visits	247,517	283,534	103,959	635,010
Attendance (%)	93.9	93.2	87.4	92.5
Recalls	6090	6352	2180	14622
Rate (%)[a]	2.4	2.2	2.1	2.3
Histological confirmation	1231	1512	566	3309
Rate (%)[a]	0.50	0.53	0.54	0.52
Malignant	801	1093	419	2313
Rate (%)	0.32	0.39	0.40	0.36
In situ	0.03	0.04	0.03	0.04
Stage I	0.17	0.22	0.27	0.21
Stage II+	0.09	0.10	0.09	0.10
Unknown	0.03	0.03	0.01	0.02
PPV mammography (%)[b]	13.2	17.2	19.2	15.8
PPV histology (%)[c]	65.1	72.3	74.0	69.9
B:M ratio[d]	0.54	0.38	0.35	0.43

[a]Percentage among visits.
[b]Positive predictive value of mammography.
[c]Positive predictive value of histological confirmation.
[d]Benign to malignant biopsy ratio.

Table 2 Results from service screening programmes in the Netherlands[7–9,17] and Finland compared to the guidelines of the European Community[1] at subsequent screens

	Desirable level	Netherlands (50–69 years)	Finland (50–64 years[a])
Proportion of women invited that attend for screening	>75%	79%	93%
Proportion of participants recalled or referred for further examinations	<3%	0.7%	2.3%
Breast cancer detection rate in comparison with the background incidence rate	>1.5 × IR	1.44 × IR	2.1 × IR
Proportion of screen-detected cancers that are carcinoma *in situ*	10–20%	14%	10%
Proportion of screen-detected cancers that are stage II+	<20%	17%	26%
Proportion of screen-detected invasive cancers that are node negative	>75%	71%	79%
Benign to malignant biopsy ratio	<0.2:1	0.3:1	0.4:1

[a]Tumours without pTNM classification excluded.

average recall rate, 2.3%, was below the desirable 3% level but at a high level compared with the indicators reported from the Netherlands (Table 2).

The screen-specific breast cancer detection rate was 0.36%. The average ratio of detection rate to background incidence (D:I ratio) was 2.1 and varied from 1.6 to 3.1 by screening centre (Figure 1). The D:I ratio as well as the proportion of screen-detected invasive cancers that are node-negative satisfied the criteria of European Community: 1.5 times incidence rate and 75%.

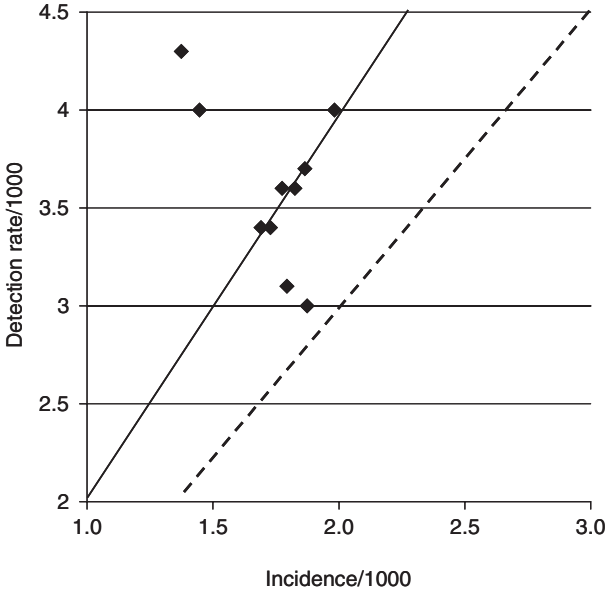


Figure 1 Detection rates compared with the background incidence at subsequent screens in the CSF screening centres in 1991–2000. The dashed line describes the reference value drawn from the European quality assurance guideline,[1] the solid line the ratio of 2.0

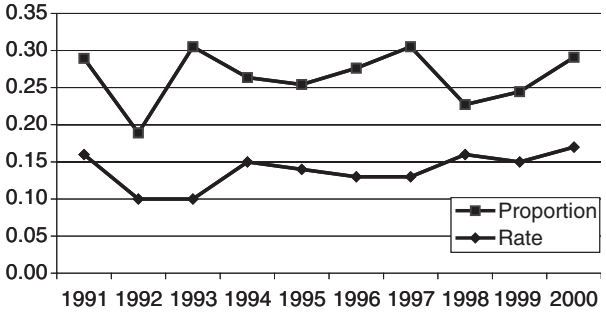


Figure 2 Annual proportion and rate of stage II+ cancers at subsequent screens in the CSF screening centres in 1991–2000

The screen-detected cases were of relatively unfavourable stage, and 26% were stage II+. Furthermore, there was no indication of a decreasing trend in the rate or in the proportion of these lesions (Figure 2).

Approximately 70% of the surgical biopsies revealed breast cancer. The benign to malignant biopsy ratio was high at 0.4:1, ranging from 0.3 to 0.8 by centre (Figure 3), compared with the desirable level of 0.2:1 by the European Community.

DISCUSSION

The ultimate goal of breast cancer screening is to reduce breast cancer mortality. There is evidence on the effectiveness of service screening programmes. At the national level, a decrease of 24% in breast cancer mortality among women invited to screening in the age group 50–64 years was reported in Finland.[3] A reduction of 19% in mortality was reported in Helsinki, where the programme targeted women aged 50–59 years.[11] There was, however, a substantial variation in the estimates and only the reduction in women aged under 56 years at entry in the national evaluation was significant. In the Netherlands, a 20% reduction was found in women aged 55–74 years.[10] Hence, there is a marked consistency in the Finnish and the Dutch data, indicating that reductions in mortality are somewhat lower or approach the average based on randomized trials.[6,12] Similar results were found for service screening programmes in Sweden,[13,14] where larger effects on mortality have also been reported.[15,16] Therefore, process indicators of the service screening programmes in these countries should be consistent with the desirable levels of the European Community.[1]

Consistently high participation rates over subsequent screens is considered to be one of the main intermediate indicators of success of a screening programme. In this study, the attendance rates among women aged 50–64 years were high and well above the desired levels of the European Community,[1] and were also very high when compared with the Netherlands (Tables 1 and 2). The detection rate in comparison with the background incidence rate satisfied the quality assurance criteria. The results from the Netherlands were at the level of acceptable value.[17] In the recommendations, there was no correction due to increase in the background risk.[4] Because of the increasing trend in the incidence, pre-screening rates become less appropriate the longer the programme is run. Therefore, the detection rates in the Netherlands were clearly lower than the limit of the European Community, 1.5 times the background incidence, estimated without correction for trend.[1]

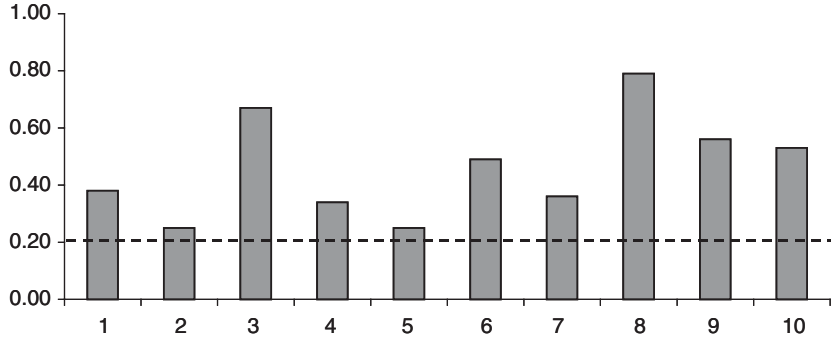


Figure 3 Centre-specific benign to malignant biopsy ratios at subsequent screens in 1991–2000. The level (0.2:1) desired by the European Community is marked by a dashed line

The ratio of detection rate to background incidence depends on the sensitivity of the test, the number of screens (first/subsequent), the sojourn time, the screening interval, and the availability of valid estimate of the reference value (i.e. the background incidence). The Finnish study was restricted to subsequent screens of a continuous programme with two-year screening intervals, which reduces the sojourn time to affect the values of D/I ratio larger than 2. It is, however, possible for the detection rate to exceed the incidence times the interval if the sojourn time is such that substantial numbers of screen-detected cancers would have surfaced after the upper age limit of screening. Assuming 14 years of programme coverage (50–64 years) with two years screening interval and an average sojourn time of 3.8 years,[18] the bias due to shift of the age-incidence curve could be approximately 13.6%. The sojourn time may also account for the results of high values of detection rates to background incidence if there is a difference in incidence between the attenders and non-attenders of screening. Due to the high participation rate in Finland, this kind of a bias is relatively small.

The D:I ratio more than 2 may also be due to false (too low) estimate of the growth rate in the background incidence. The variation in background incidence was large and consistent with differences in standard of living or reproduction in Finland. Therefore, we first estimated the background incidence by municipality and thereafter accumulated the incidences by screening centre and also took into account the increasing overall trend. Furthermore, the large differences in detection rates could not be accounted for by background incidence. In fact, there was indication of negative correlation between background incidences and detection rates (Figure 1).

If the test sensitivity is poor to start with and improves round by round, there may be a possibility for larger estimated values of D:I ratio than 2. In the Finnish screening programme, the average D/I ratio at first screen was high (2.74),[4] which does not imply a major improvement in test sensitivity by time.

In conclusion, the observed D:I ratio more than 2 in the present study probably implies overdiagnosis, at least in some studied areas. Hence, a very large D:I ratio indicates harm (overdiagnosis) rather than a benefit. Further confirmatory work (e.g. an analysis by more detailed age groups with a sufficiently long follow-up span) is indicated.

It is likely that the criteria of the European Community should have an upper limit of the D:I ratio as well, and not the lower limit only. On the basis of observations from Finland[4,11] and the Netherlands,[17] we propose that the acceptable limits were from 0.67 to 0.9 times the screening interval for the ratio of detection rate to background incidence.

In the Netherlands, the levels recommended by the European Community in the proportion of stage II+ cancers were clearly reached[8,9] (Table 2). In Finland they were not met. The proportion of non-localized disease is, however, a poor indicator and rates should be considered. If screening programmes are successful in allowing earlier diagnosis of cancer, an overall reduction in the rates of advanced cancers should be observed in the target population.[5] Data on only screen-detected cancers in the target population are insufficient for valid conclusions. Even data on all breast cancers (not only the screen-detected ones) may be insufficient; it has been shown that stage-based surrogates can be conservative compared with actual mortality.[19]

In our material, the rates of stage II+ cancers increased and the proportions did not meet the desirable levels. However, the programme is probably effective[3,9] and many other indicators were consistent with the effectiveness. In addition, there are also logical problems in the European Guidelines.[20] It seems that this recommendation of the European Commission also needs revision.

The sensitivity of cancer detection in breast cancer screening should be high, but it is also important to achieve high diagnostic specificity to avoid the high cost and morbidity associated with unnecessary examinations. The aim of screening should thus be both an accurate diagnosis with prompt referral and an infrequent and accurate diagnosis of benign changes or intraductal lesions.[5,21]

The recall rates in Finland were much higher than in the Netherlands,[7–9] where the referral rates are uniquely low. The Finnish rates were, however, below the limits of the European Community. Another indicator aiming at good screening specificity is the B:M ratio, the desirable (acceptable) level of which is 0.2:1 (1:1) in the European Community.[1] In our study the average figures (0.35:1 to 0.54:1 by age) reached the acceptable level but there were large differences between the centres (Figure 3). In the Netherlands, the average B:M ratio was 0.3:1.[7,8]

The desirable levels of process or intermediate indicators were settled mainly on the basis of the experience from the randomized controlled trials in Sweden.[5] Further application of these criteria for the European Community monitoring system may need revision with accumulation of data. We conclude that stage distribution in patients and rate of advanced cancers in screening need revision. The ratio of detection rate to background incidence should depend on the screening interval and an acceptable area instead of a lower threshold limit is to be agreed. We propose 0.67–0.9 times the screening interval as the acceptable area.

ACKNOWLEDGEMENTS

This study has been financially supported by the European Community Europe Against Cancer Action Programme. We thank Mr Håkan Forsman for programming and Mrs Päivi Styrman for careful handling and recording of screening data.

Authors' affiliations
T Sarkeala, *Researcher*, Finnish Cancer Registry, Liisankatu 21 B, FIN-00170 Helsinki, Finland
A Anttila, *Head of Research*, Finnish Cancer Registry, Liisankatu 21 B, FIN-00170 Helsinki, Finland
I Saarenmaa, *Chief Radiologist*, Pirkanmaa Cancer Society, Hämeenkatu 5A, FIN-33100 Tampere, Finland
M Hakama, *Director*, Finnish Cancer Registry, Liisankatu 21 B, FIN-00170 Helsinki, Finland; *Professor*, Department of Public Health, University of Tampere, FIN-Tampere, Finland

REFERENCES

1 The European Community. *European Guidelines for Quality Assurance in Mammography Screening*. Luxembourg: Europe Against Cancer Programme, Office for Official Publications of the European Communities, 2001
2 Hakama M, Elovainio L, Kajantie R, *et al*. Breast cancer screening as public health policy in Finland. *Br J Cancer* 1991;**64**:962–4
3 Hakama M, Pukkala E, Heikkilä M, *et al*. Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *Br Med J* 1997;**314**:864–7

4 Sarkeala T, Anttila A, Forsman H, *et al.* Process indicators of ten centres in the Finnish breast cancer screening programme in 1991–2000. *Eur J Cancer* 2004;**40**:2116–25
5 Day NE, Williams DRR, Khaw KT. Breast cancer screening programmes: the development of a monitoring and evaluation systems. *Br J Cancer* 1989;**59**:954–8
6 International Agency for Research on Cancer, IARC. *IARC Handbooks of Cancer Prevention. Vol 7: Breast Cancer Screening*. Oxford: Oxford University Press, 2002
7 Fracheboud J, de Koning HJ, Boer R, *et al.* Nationwide breast cancer screening programme in the Netherlands. Results of initial and subsequent screening 1990–1995. *Int J Cancer* 1998;**75**:694–8
8 Fracheboud J, de Koning HJ, Boer R, *et al.* Nationwide breast cancer screening programme fully implemented in the Netherlands. *Breast* 2001;**10**:6–10
9 Verbeek ALM, Broeders MJM. Evaluation of the Netherlands breast screening programme. *Ann Oncol* 2003;**14**:1203–5
10 Otto SJ, Fracheboud J, Looman CWN, *et al.* Initiation of population-based mammography screening in Dutch municipalities and effect of breast cancer mortality: a systematic review. *Lancet* 2003;**361**:1411–7
11 Anttila A, Koskela J, Hakama M. Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. *J Med Screen* 2002;**9**:153–8
12 Nyström L, Andersson I, Bjurstam N, *et al.* Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* 2002;**359**:909–19
13 Jonsson H, Nyström L, Törnberg S, *et al.* Service screening with mammography of women aged 50–69 years in Sweden: effect on mortality from breast cancer. *J Med Screen* 2001;**8**:152–60
14 Jonsson H, Nyström L, Törnberg S, *et al.* Service screening with mammography. Long-term effects on breast cancer mortality in the county of Gävleborg, Sweden. *Breast* 2003;**12**:183–93
15 Duffy SW, Tabar L, Chen HH, *et al.* The impact of organised mammography service screening on breast carcinoma mortality in seven Swedish counties. *Cancer* 2002;**3**:458–69
16 Tabar L, Yen MF, Vitak B, *et al.* Mammography service screening and mortality in breast cancer patients: 20 year follow-up before and after introduction of screening. *Lancet* 2003;**361**:1405–10
17 Fracheboud J, de Koning HJ, Beemsterboer PMM, *et al.* Interval cancers in the Dutch breast cancer screening programme. *Br J Cancer* 1999; **81**:912–7
18 Duffy SW, Day NE, Tabar L, *et al.* Markov models of breast tumour progression: some age specific results. *J Natl Cancer Inst Monogr* 1997; **22**:93–7
19 Morrison AS. Intermediate determinants of mortality in the evaluation of screening. *Int J Epidemiol* 1991;**20**:642–50
20 Njor SM, Olsen AH, Bellstrom T, *et al.* Mammography screening in the county of Fyn November 1993–December 1999. *APMIS* 2003; **111** (Suppl. 110):1–33
21 Duffy SW, Tabar L, Vitak B, *et al.* The relative contributions of screen-detected *in situ* and invasive breast carcinomas in reducing mortality from the disease. *Eur J Cancer* 2003;**39**:1755–60

Int. J. Cancer: **122,** 614–619 (2008)

Organised mammography screening reduces breast cancer mortality: A cohort study from Finland

Tytti Sarkeala[1]*, Sirpa Heinävaara[1,2] and Ahti Anttila[1]

[1]*Finnish Cancer Registry, Helsinki, Finland*
[2]*Department of Public Health, University of Helsinki, Helsinki, Finland*

We evaluated the effectiveness and the efficacy of population-based mammography programme in Finland, and explored associations between the screening performance and the screening efficacy. The main outcome, incidence-based mortality from breast cancer, was estimated by invitation, participation, age at death, and screening centres categorised by recall rates. The study was based on an individual followup of screening invitees and participants from 1992 to 2003. The coverage of screening invitations was 95% among 50–59 years old women, and 20–40% among women aged 60–69 years. We compared observed deaths from breast cancer to expected breast cancer deaths without screening in ages 50–69 at death. The observed deaths were obtained from a cohort of individual invitees (n = 361,848). The expected deaths were defined by modelling breast cancer mortality from 1974 to 1985 and 1992 to 2003 at population level. The population data were derived from the same municipalities (n = 260) that were incorporated into the cohort. The breast cancer mortality among the invited women was reduced by 22% (relative risk 0.78, 95% confidence interval 0.70–0.87). After adjusting for the self-selection, the efficacy among the participants was 28% (0.72, 0.56–0.88). No clear association between the recall rates and the screening efficacy was observed. The organised mammography screening in Finland is effective. The relationship between the estimates of process and outcome of mammography is not yet straightforward: effectiveness and efficacy remain the best estimates for evaluating the success of mammography screening.

Key words: effectiveness; mammography; screening; breast cancer; epidemiology

Several randomised trials have reported effectiveness of screening for breast cancer. The combined results showed 25% reduction in breast cancer mortality among women aged 50–69 years at randomisation.[1] The cohort studies on service screening from Sweden and Denmark have reported effects at similar level.[2–4]

In Finland, a pilot study on breast cancer screening began in 1982.[5] The nationwide mammography programme started in 1987 and it was implemented gradually. The effectiveness for the first 5 years of the programme was analysed using randomised birth cohorts.[6] The study demonstrated a nonsignificant, 24% reduction in breast cancer mortality associated with invitation to screening.

The adaptation of results from randomised trials to routine screening is not self-evident. The performance and validity of mammography screening within and between the European programmes have varied widely,[7–9] suggesting differences also in the effectiveness and adverse effects of screening. There has also been variation in the screening policies.[1]

The effectiveness of mammography screening has been debated during the last years,[10–12] and the impact of population-based screening on breast cancer mortality has been analysed in many European countries.[2–4,6,13–15] The main aim of the current study was to analyse the effectiveness of mammography service screening in Finland. The additional aims were: (*i*) to explore the efficacy of screening among the screening participants with appropriate adjustment for self-selection bias, and (*ii*) to report associations between the screening performance and the screening efficacy.

Material and methods

The randomised implementation period within the Finnish mammography programme ended in 1991. Since then, the actual coverage of screening invitations has been over 95% among women aged 50–59 years. Among 60–64 years old women the invitational coverage has been ~40%, and among women aged 65–69 years 20%.[7] The screening interval is 2 years.

We analysed the effect of mammography programme on breast cancer mortality in 1992–2003 by comparing observed breast cancer deaths among screening invitees, participants and nonparticipants with expected breast cancer deaths without screening. The observed deaths and the person-years at risk were obtained from a cohort of individual screening invitees. As all the women from the target age group (50–59) were invited to mammography screening in the study period, no noninvited controls were available. The expected mortality rates without screening were thus estimated by modelling, using population data at municipal level. The data were derived from the Mass Screening Registry, the Finnish Cancer Registry and the National Population Registry.

Defining observed breast cancer deaths and person-years at risk, cohort followup

The cohort included women aged 50 years or more, who in 1992–2003 had been invited at least once to any of the 10 screening centres of the Cancer Society of Finland (CSF). The CSF centres covered 50–60% of the organised mammography in Finland, and were the only screening providers sending service screening information to the Mass Screening Registry regularly throughout the study period.

As individual municipalities in Finland are entitled to conduct the mammography program, and may change their screening provider annually, we excluded 55 municipalities (96 001 women, 21.0%) from the study due to a limited duration of their service screening (<9 years) within the CSF centres. The final number of women in the study was 361,848 and the number of invitations in the study period is 1,166,331. The median number of invitations was 3 (ranging from 1 to 8), and the median followup time 9.8 years.

The individual followup of the invitees was initially started from the date of first invitation in 1992–2003 in ages 50–69. However, the expected breast cancer mortality rates were available at the population level only, where the followup was started at January 1st. To obtain comparability with the population data, we set the individual cohort entry date to January 1st in the first invitation year in 1992–2003. The followup was ended at death, emigration or December 31st 2003. Women responding positively to their first invitation in 1992–2003 were classified as participants, the other invitees were nonparticipants. In our previous study, we discovered that over 90% of first-screen participants reattended at least once after subsequent invitation.[7]

To study the performance of screening, we categorised the 10 CSF centres into 3 groups. The categorisation reflected variation in the screening specificity, and was based on information on centre-specific recall rates at the subsequent screens in

***Correspondence to:** Finnish Cancer Registry, Liisankatu 21 B, 00170 Helsinki, Finland. Fax: +358-9-673108. E-mail: tytti.sarkeala@cancer.fi
Received 15 February 2007; Accepted after revision 11 July 2007
DOI 10.1002/ijc.23070
Published online 10 September 2007 in Wiley InterScience (www.interscience.wiley.com).

TABLE I – PERSON-YEARS AND OBSERVED NUMBERS AND RATES OF REFINED DEATHS FROM BREAST CANCER IN 1974–1985; AND PERSON-YEARS AND OBSERVED AND EXPECTED NUMBERS AND EXPECTED RATES OF REFINED DEATHS FROM BREAST CANCER WITHOUT SCREENING IN 1992–2003 BY 5-YEAR AGE GROUPS AT DEATH AND BY CENTRE CATEGORIES, POPULATION DATA

	50–54	55–59	60–64	65–69	50–69
1974–1985					
Person-years	818,189	805,378	762,399	718,240	310,4206
Observed	115	232	275	269	891
Observed rate (/100,000)	14.06	28.81	36.07	37.44	28.70
1992–2003					
Person-years	960,169	804,801	743,934	734,340	324,3244
Observed	109	213	192	240	754
Expected	139.1	238.1	268.3	273.4	918.9
Expected rate (/100,000)	14.49	29.58	36.06	37.23	28.33
	Low (0.9–1.9%)	Intermediate (2.3–2.7%)		High (2.8–3.5%)	50–64 (0.9–3.5%)
1974–1985					
Person-years	687530	1138171		560265	2385966
Observed	202	285		135	622
Observed rate (/100 000)	29.4	25.0		25.0	26.1
1992–2003					
Person-years	765,784	1,179,557		563,563	2,508,904
Observed	187	221		106	514
Expected	220.3	292.3		132.8	645.5
Expected rate (/100,000)	28.8	24.8		23.6	25.7

1991–2001.[7,8] In the group "Low recall rate" (3 centres) the range in recall rates was 0.9–1.9%, in the group "Intermediate recall rate" (4 centres) 2.3–2.7%, and in the group "High recall rate" (3 centres) 2.8–3.5%. The corresponding attendance rates were 92.1, 91.5 and 93.7%, and the detection rates for the screen-detected cancers (/1000 screened) 3.44, 3.60 and 3.87. The recall rates correlated inversely with the incidence of interval cancers.[8]

We linked the individual screening data with the National Population Registry and with the Cancer Registry using the unique personal identifier as a key. Person-years at risk and incidence-based (refined) breast cancer deaths were calculated by participation, 5-year age groups at death, and 3 centre categories. Deaths from breast cancers diagnosed during the followup in 1992–2003 represented the observed, refined breast cancer deaths.

Estimating expected mortality rates without screening, population data

Population data consisted of age-specific numbers of women and refined breast cancer deaths by calendar year and by municipality from 1974 to 1985 and 1992 to 2003. The municipalities ($n = 260$) were the same as in the cohort followup. The period 1974–1985 represented the latest prescreening era of equal length to the study period 1992–2003: some municipalities engaged to the current study had started mammography screening in 1986, prior to the launch of the national mammography programme. The refined breast cancer deaths were defined as follows: breast cancers diagnosed in 1974–1985 or in 1992–2003 among 50–69 years old women formed the basis. Deaths from these breast cancers within the corresponding 2 periods represented the refined breast cancer deaths.

Population data were used to estimate the expected, refined breast cancer mortality rates without screening. The rates were modelled by Poisson regression with a logarithmic link function. In the model, 5-year age groups at death, centre categories (low, intermediate and high recall rates), period before (1974–1985) and with screening (1992–2003), calendar year within the 2 periods (1, 2, . . . ,12 years), and interaction between the calendar year and the age at death were used as explanatory variables. The first 3 variables were categorical, and the 4th variable was numerical. The average difference in the refined breast cancer mortality between the two 12-year periods represented the overall screening effect. Birth cohorts were also studied, but they were excluded from the model, because they were strongly correlated with the estimated screening effect. Likelihood ratio statistics and descriptive evaluation between the model-based and observed rates were used as decision criteria in formulation of the model.

The fitted mortality values of the model were calculated with the screening effect excluded, and they represent the expected, refined breast cancer mortality rates without screening in 1992–2003.

Table I contains the person-years and the observed numbers of refined deaths from breast cancer in the prescreening (1974–1985) and the screening (1992–2003) periods by age at death and by centre categories. The observed rates from 1974 to 1985 and the expected rates without screening from 1992 to 2003 are also shown. The accumulation of breast cancer deaths by 5-year age groups at death in 1974–1985 and 1992–2003 are illustrated in Figure 1.

Formulating effect estimates

We studied the ratio between the observed and expected refined breast cancer deaths within the cohort followup in ages 50–69 at death. The observed deaths were divided by the corresponding number of expected deaths. The expected breast cancer deaths without screening were calculated by multiplying the expected mortality rates derived from the population data with the corresponding person years derived from the cohort followup. The confidence intervals were corrected with an over-dispersion constant produced by the model (1.36). The bias due to self-selection was adjusted by method described by Cuzick *et al*.[16]

We further analysed, whether the variation in the refined breast cancer mortality by centre category was consistent with the previous association between rising recall rates and incidence of interval cancers.[8] We restricted this examination to deaths at 50–64 years of age, as this invitational age range was covered in all the 3 centre categories.

Results

In 1992–2003, altogether 2,731,268 person-years were accumulated to the cohort in ages 50–69. The number of refined breast cancer deaths was 617 (Table II). Among all invitees, the reduction in the refined breast cancer mortality in ages 50–69 at death was 22% (relative risk 0.78, 95% confidence interval 0.70–0.87). The mortality reduction was highest, 29%, in women aged 60–64 years at death (0.71, 0.58–0.87). In ages 50–54 and 55–59 years at death, the relative risks were 0.74 (0.55–0.96) and 0.84 (0.69–1.01).

The proportion of person-years among participants out of all person-years in the cohort was 87% in ages 50–69 (Table III). The

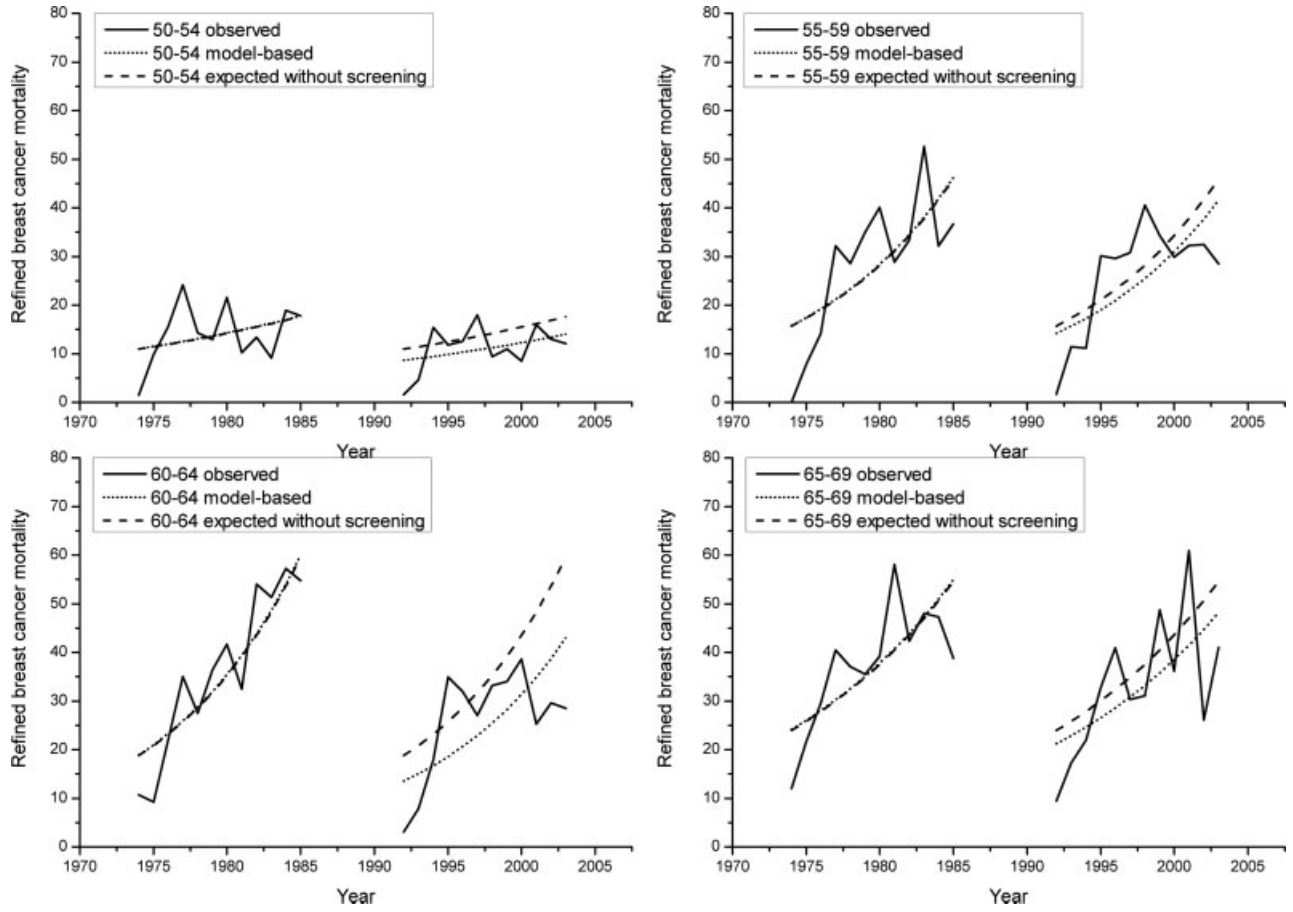


FIGURE 1 – Observed, model-based and expected breast cancer mortality rates by 5-year age groups at death in 1974–1985 and 1992–2003, population data.

TABLE II – PERSON-YEARS, OBSERVED AND EXPECTED NUMBERS OF BREAST CANCER DEATHS, AND EFFECT ESTIMATES WITH 95% CONFIDENCE INTERVALS IN 1992–2003 BY 5-YEAR AGE GROUPS AT DEATH, INDIVIDUAL FOLLOWUP

	50–54	55–59	60–64	65–69	50–69
Person-years	903692.6	795537.7	630446.3	401591.7	2731268.3
Observed	97	203	174	143	617
Expected	131.8	241.9	244.4	170.9	789.0
Observed/expected (95% CI)	0.74 (0.55–0.96)	0.84 (0.69–1.01)	0.71 (0.58–0.87)	0.84 (0.66–1.04)	0.78 (0.70–0.87)

ratio between the observed and expected breast cancer deaths among the participants was 0.66 (0.58–0.75). The corresponding relative risk among the nonparticipants was more than 2 times higher, 1.56 (1.25–1.91). After adjusting for the self-selection in attendance, the relative risk among the participants became 0.72 (0.56–0.88).

In the 3 centre categories grouped by ascending order of recall rates (low, intermediate, high), the risk ratios among the screening invitees were 0.83 (0.67–1.01), 0.71 (0.58–0.85) and 0.79 (0.60–1.02) (Table IV). Among participants, the person-years at risk were 626,171 (87.5% of invitees), 953,664 (87.7%), and 465,540 (88.3%). After adjusting for the self-selection, no clear association between the screening efficacy and the recall rates could be found (Fig. 2).

Discussion

We compared the observed deaths from breast cancer with screening to the expected breast cancer deaths without screening in 1992–2003. The observed deaths were obtained from a cohort of individual invitees, participants and nonparticipants. The expected deaths were derived by modelling mortality rates at population level. Concurrent, noninvited controls were not available.

There was a significant, 22% reduction in the incidence-based breast cancer mortality among the screening invitees. The reduction was seen only among the screening participants, while the relative risk among the nonparticipants was significantly elevated. The large difference in the relative risks between the participants and the nonparticipants remained significant after adjustment for selection-bias. No clear association between the performance and the efficacy of mammography screening was observed.

Data and methods

Women invited at least once to any of the 10 centres of CSF in 1992–2003 were members of the cohort until emigration, death or end of the followup irrespective of possible migration to municipalities outside the study area. The fixed accumulation of person-years in the cohort thus slightly differed from the dynamic accu-

TABLE III – PERSON-YEARS, OBSERVED AND EXPECTED NUMBERS OF BREAST CANCER DEATHS, AND EFFECT ESTIMATES WITH 95% CONFIDENCE INTERVALS IN 1992–2003 BY 5-YEAR AGE GROUPS AT DEATH AMONG PARTICIPANTS AND NONPARTICIPANTS

	50–54	55–59	60–64	65–69	50–69
Nonparticipants					
Person-years	98202.8	95143.0	90956.2	74562.8	358864.8
Observed	29	58	45	33	165
Expected	14.2	28.0	33.8	30.1	111.8
Observed/expected	2.04	2.07	1.33	1.10	1.56
(95% CI)	(1.15–3.26)	(1.41–2.91)	(0.86–1.95)	(0.65–1.71)	(1.25–1.91)
Participants					
Person-years	805489.8	700394.7	539490.2	327028.8	2372403.4
Observed	68	145	129	110	452
Expected	117.6	213.9	210.6	140.8	682.9
Observed/expected	0.58	0.68	0.61	0.78	0.66
	(0.41–0.79)	(0.54–0.84)	(0.48–0.77)	(0.60–1.00)	(0.58–0.75)
Corrected	0.66	0.79	0.65	0.80	0.72
(95% CI)	(0.28–1.04)	(0.49–1.09)	(0.38–0.92)	(0.46–1.14)	(0.56–0.88)

Corrected refers to effect estimates with adjustment for selection bias.[16]

TABLE IV – PERSON-YEARS, OBSERVED AND EXPECTED NUMBERS OF BREAST CANCER DEATHS, AND EFFECT ESTIMATES WITH 95% CONFIDENCE INTERVALS IN 1992–2003 BY CENTRE CATEGORIES (RECALL RATES, %, IN PARENTHESES), INDIVIDUAL FOLLOWUP

	Low (0.9–1.9%)	Intermediate (2.3–2.7%)	High (2.8–3.5%)	50–64 (0.9–3.5%)
Person-years	715643.8	1087013.5	527019.3	2329676.6
Observed	176	197	101	489
Expected	212.5	277.9	127.7	636.7
Observed/expected	0.83	0.71	0.79	0.77
(95% CI)	(0.67–1.01)	(0.58–0.85)	(0.60–1.02)	(0.68–0.86)

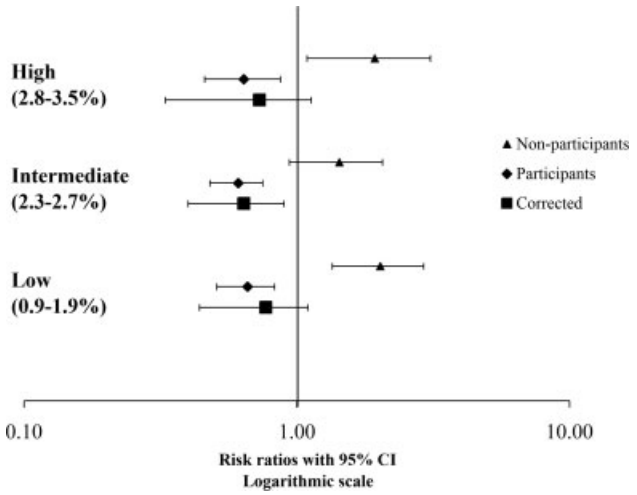


FIGURE 2 – Effect estimates with 95% confidence intervals in 3 centre categories by participants and nonparticipants (recall rates, %, in parentheses). Corrected refers to effect estimates with adjustment for selection bias.[16]

mulation of person-years in the population data. These differences were small, however, and are not likely to affect the comparability between the individual and the population level data.

During the study period, an increase in breast cancer incidence and a decrease in breast cancer mortality were reported in the Finnish female population aged 50–69. The decrease in mortality was consistent only since 1995, after a persistent increase since the early 1970s.[17,18] The increase in breast cancer incidence refers to increase in the background risk, and suggests similar pattern also for the background trends of the refined breast cancer mortality in the current study. The decrease in breast cancer mortality indicates improvements in breast cancer diagnostics and treatment.

Our model on the expected rates used information from 1974 to 1985 and 1992 to 2003, and was constructed to attain both the increase in the background risk and the decrease in the breast cancer mortality. Because of contradicting developments, the eventual difference between the observed rates in 1974–1985 and the expected rates in 1992–2003 remained small: the expected rates in ages 50–69 at death were only 1.3% lower than the corresponding observed rates in the prescreening period (see Table I).

When the reduction of breast cancer mortality within service screening is examined, the contributions of screening, other diagnostic services, and treatment are difficult to separate. In the United States, the effect of screening and adjuvant therapy on breast cancer mortality was estimated to be similar, but the variation in the modelling approaches reflected considerable uncertainty.[19] In East Anglia, an analysis of survival by tumour size was used to separate screening effects from other effects. The authors estimated that 60% of the improved survival was due to earlier diagnosis by screening.[20]

We studied the effectiveness of organised mammography screening in a cohort of individual participants and nonparticipants. The information on screening attendance elaborates the assessment, because health care practices independent of screening are assumed to be similar to all. The nonparticipants, however, may be persons with breast cancers at the time invitations are generated, or persons with unhealthy behaviour.[21] We therefore adjusted the effect estimates of the participants for self-selection. The corrected estimates were well below one referring that selection did not explain the current result. Variability in treatment or in the access of other diagnostic services between the participants and the nonparticipants may have modified the screening effect, however. This issue is beyond the scope of the current study.

The firm impact of screening on breast cancer mortality in Finland is supported by a recent study in 2 Finnish cities employing different screening policies.[15] Compared to a city with no screening (Helsinki), the breast cancer mortality among studied birth cohorts was over 40% lower in a city providing regular screening (Turku). The mortality rates from breast cancer increased between the prescreening and screening periods in Helsinki, while in Turku the rates reduced significantly by 36%.

TABLE V – OVERVIEW OF THE FINNISH STUDIES ON THE EFFECTIVENESS OF BREAST CANCER SCREENING

	Hakama *et al.*[6]	Anttila *et al.*[14]	Parvinen *et al.*[15a]	Current study[b]
Design	Randomisation of birth cohorts	Population-based cohort study	Population-based cohort study	Individual-based cohort study
Main outcome	Incidence-based breast cancer mortality	Incidence-based breast cancer mortality	Incidence-based breast cancer mortality	Incidence-based breast cancer mortality
Comparison	Breast cancer deaths among invited and noninvited individuals	Modelled breast cancer deaths among invited and noninvited birth cohorts	Breast cancer deaths among invited birth cohorts in screening and prescreening periods	Observed breast cancer deaths among individual invitees and participants with modelled, population-based breast cancer deaths in the absence of screening
Area	65% of Finnish municipalities	The city of Helsinki	The city of Turku	50% of Finnish municipalities
Period				
Study	1987–1991	1986–1997	1987–1997	1992–2003
Reference	1987–1991	1970–1997	1976–1986	1974–1985
Birth cohorts				
Study	1928,1930,1932, 1934,1938	1935–39	1918–32	1923–53
Reference	1927,1929,1933, 1935,1939	1930–34	1907–21	–
Age	49–60 at entry 49–64 at death	50–59 at entry 50–71 at death	55–69 at entry in 1987 55–80 at death	50–69 at entry 50–69 at death
RR (95% CI)	0.76 (0.53–1.09)	0.81 (0.62–1.05)	0.64 (0.47–0.88)	0.78 (0.70–0.87)

[a]The study reported also comparisons of effect estimates in Turku with estimates from the cities Helsinki and Tampere.
[b]The study did not include Helsinki and Tampere. Approximately 5% of women in Parvinen *et al.*[15] were cohort members of the study.

Comparison of results to other studies

A reduction of 24% in the refined breast cancer mortality was demonstrated among 49–64 years old invitees by the group-randomised design in 1987–1991 in Finland.[6] In 1986–1997, a decrease of 19% in the refined mortality was reported by a dynamic study on screening among Finnish women aged 50–59 years at entry.[14] In another dynamic cohort study, the breast cancer mortality in 1987–1997 was compared to mortality during the prescreening period of 1976–1986.[15] The organised mammography was associated with reduced breast cancer mortality, particularly among elderly. Design, geographical coverage and main results of the Finnish studies on the effectiveness of mammography screening are summarised in Table V.

The cohort studies on service screening from Sweden have reported similar or even higher overall effect estimates than the Swedish randomised controlled trials.[2,3,22] In 9 counties, representing 45% of the Swedish women, the reduction in breast cancer mortality was significant, 27%, among the 40–69 years old invitees.[2,3] After adjustment for a self-selection, contemporaneous changes in incidence and changes in mortality independent of screening, the reduction among the participants became 39%.

In the Netherlands, the screening programme started in the late 1980s, and by 1997 all the women aged 50–69 were covered. The breast cancer mortality rates in the screening period were compared to those in the prescreening period.[13] From 1997 onwards the reduction was significant. In 2001, the rates were 20% lower than in 1986–1988 among 55–74 years old women.

In Denmark, the effect of organised programme on breast cancer mortality was studied among 50–69 years old invitees in Copenhagen.[4] Historical, national and historical-national groups were used as controls. During the first 10 years of the programme (1991–2000), the refined breast cancer mortality at ages 50–79 years reduced significantly by 25%. After adjusting for the self-selection bias, the reduction among the participants became 37%. The relative risks were 0.57 (0.25–1.30) (9 breast cancer deaths) and 1.08 (0.55–1.20) (34 breast cancer deaths) in ages 50–54 and 55–59 at death in the Copenhagen study.

In the current study, the overall reduction in the refined breast cancer mortality in ages 50–69 was at similar level as reported from Denmark, the Netherlands, and Sweden.[2–4,13] The attendance rate in Finland was higher, and the efficacy estimates therefore lower than in Denmark and in Sweden, where the organised screening was continued until 69 years of age. In Finland, the considerable accumulation of postscreening followup in ages 60–69 reduced the mortality reduction in this age group. In ages 50–59, the reduction in Finland was greater than in Denmark.[4] The remarkable decrease in the refined breast cancer mortality in ages 50–54 was unexpected, but appeared consistent both in the individual and population level analyses.

No earlier reports on breast cancer mortality by screening performance have been published. In the current study, no difference in the screening efficacy between the centre categories (low, intermediate, high recall rate) was observed in ages 50–64 ($p = 0.4105$). Our previous study from 1991 to 2003 reported significant increase in the incidence of interval cancers by ascending recall rate category.[8] The detection rates of nonlocalised breast cancers (interval and screen-detected combined) were similar in all the 3 centre categories, however.[8] The current result may at least partially indicate differences in diagnostic and treatment practices within and/or outside the organised screening. Noteworthy, studies on organised mammography from the Netherlands have reported low recall rates (0.7–1.3%),[23,24] reduced rates of advanced cancers,[25] and similar impact on the breast cancer mortality as the current study.[13] Thus it is possible that indicators of screening performance may be of limited use in predicting the quantity of mortality reduction. Further studies are warranted.

Conclusions

The organised breast cancer screening in Finland has been effective. The number of prevented breast cancer deaths could probably be increased, if service screening were uniformly extended to 60–69 years old women. The relationship between the estimates of process and outcome of mammography is not yet straightforward: effectiveness and efficacy remain the best estimates for evaluating the success of mammography screening.

References

1. International Agency for Research on Cancer, IARC. IARC handbooks of cancer prevention, Vol. 7. Breast cancer screening. Oxford: Oxford University Press, 2002.229 p.
2. The Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from the organised service screening with mammography: Part 1. Further confirmation with extended data. Cancer Epidemiol Biomarkers Prev 2006;15:45–51.
3. The Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from the organised service screening with mammography: Part 2. Validation with alternative analytic methods. Cancer Epidemiol Biomarkers Prev 2006;15:52–6.
4. Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen M-B, Brix Tange U, Blichert-Toft M, Rank F, Mouridsen H, Lynge E. Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study. BMJ 2005;330:220.
5. Hakama M, Pukkala E, Kallio M, Godenhjelm K, Svinhufvud U. Effectiveness of screening for breast cancer in women under 50 years at entry: the Kotka pilot project in Finland. Int J Cancer 1995;63:55–7.
6. 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.
7. Sarkeala T, Anttila A, Forsman H, Luostarinen T, Saarenmaa I, Hakama M. Process indicators from ten centres in the Finnish breast cancer screening programme from 1991 to 2000. Eur J Cancer 2004; 40:2116–25.
8. Sarkeala T, Hakama M, Saarenmaa I, Hakulinen T, Forsman H, Anttila A. Episode sensitivity in association with process indicators in the Finnish breast cancer screening program. Int J Cancer 2006;118: 174–9.
9. Lynge E, Olsen AH, Fracheboud J, Patnick J. Reporting of performance indicators of mammography screening in Europe. Eur J Cancer Prev 2003;12:213–22.
10. Miller AB. Is mammography screening for breast cancer really not justifiable? Recent Results Cancer Res 2003;163:115–28.
11. Olsen O, Gotzsche P. Cochrane review on screening for breast cancer with mammography. Lancet 2001;358:1340–42.
12. Gotzche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Libr 2006;4:60.
13. Otto SJ, Fracheboud J, Looman CWN, Broeders MJM, Boer R, Hendriks JH, Verbeek AL, de Koning HJ. Initiation of population-based mammography screening in Dutch municipalities and effect of breast cancer mortality: a systemic review. Lancet 2003;361:1411–17.
14. Anttila A, Koskela J, Hakama M. Programme sensitivity and effectiveness of mammography service screening in Helsinki, Finland. J Med Screen 2002;9:153–8.
15. Parvinen I, Helenius H, Pylkkänen L, Anttila A, Immonen-Räihä P, Kauhava L, Räsänen O, Klemi PJ. Service screening mammography reduces breast cancer mortality among elderly women in Turku. J Med Screen 2006;13:34–40.
16. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomised clinical trials. Stat Med 1997;16:1017–29.
17. 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.
18. Finnish Cancer Registry. Cancer in Finland 2002 and 2003. Helsinki: Cancer Society of Finland, 2005.88 p. Publication No. 66.
19. Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD, Feuer EJ. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 2005;353:1784–92.
20. Duffy SW, McCann J, Godward S, Gabe R, Warwick J. Some issues in screening for breast and other cancers. J Med Screen 2006;13 (Suppl. 1):28–34.
21. Aro A, de Koning HJ, Absetz P, Schreck M. Psychosocial predictors of first attendance for organised mammography screening. J Med Screen 1999;6:82–8.
22. 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;359:1903–8.
23. Fracheboud J, de Koning HJ, Beemsterboer PM, Boer R, Hendriks JH, Verbeek AL, van Ineveld BM, de Bruyn AE, van der Maas PJ. Nationwide breast cancer screening programme in the Netherlands. Results of initial and subsequent screening 1990–1995. Int J Cancer 1998;75:694–8.
24. Fracheboud J, de Koning HJ, Boer R, et al. Nationwide breast cancer screening programme fully implemented in the Netherlands. Breast 2001;10:6–10.
25. Fracheboud J, Otto SJ, van Dijk JA, Broeders MJ, Verbeek AL, de Koning H. Decreased rates of advanced breast cancers due to mammography screening in the Netherlands. Br J Cancer 2004;91:861–7.