

ORIGINAL ARTICLE

Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet

EUROSCREEN Working Group

J Med Screen 2012;19 Suppl 1:5–13

DOI: 10.1258/jms.2012.012077

See end of article for authors' affiliations

Correspondence to:
Eugenio Paci, Director,
Clinical and Descriptive
Epidemiology Unit, ISPO,
Cancer Prevention and
Research Unit, 50144
Florence, Italy;
e.paci@ispo.toscana.it

Accepted for publication
13 June 2012

Objectives To construct a European 'balance sheet' of key outcomes of population-based mammographic breast cancer screening, to inform policy-makers, stakeholders and invited women.

Methods From the studies reviewed, the primary benefit of screening, breast cancer mortality reduction, was compared with the main harms, over-diagnosis and false-positive screening results (FPRs).

Results Pooled estimates of breast cancer mortality reduction among invited women were 25% in incidence-based mortality studies and 31% in case-control studies (38% and 48% among women actually screened). Estimates of over-diagnosis ranged from 1% to 10% of the expected incidence in the absence of screening. The combined estimate of over-diagnosis for screened women, from European studies correctly adjusted for lead time and underlying trend, was 6.5%. For women undergoing 10 biennial screening tests, the estimated cumulative risk of a FPR followed by non-invasive assessment was 17%, and 3% having an invasive assessment. For every 1000 women screened biennially from age 50–51 until age 68–69 and followed up to age 79, an estimated seven to nine lives are saved, four cases are over-diagnosed, 170 women have at least one recall followed by non-invasive assessment with a negative result and 30 women have at least one recall followed by invasive procedures yielding a negative result.

Conclusions The chance of saving a woman's life by population-based mammographic screening of appropriate quality is greater than that of over-diagnosis. Service screening in Europe achieves a mortality benefit at least as great as the randomized controlled trials. These outcomes should be communicated to women offered service screening in Europe.

INTRODUCTION

We aimed to present a 'balance sheet' based on estimates of breast cancer mortality reduction as the primary benefit, and over-diagnosis of breast cancer and false-positive screening tests as the most important harms. The balance sheet is derived from published results of the European, population-based, mammographic screening programmes that are systematically reviewed in this supplement of the *Journal of Medical Screening*.

At the beginning of the 1990s, meta-analysis of randomized controlled trials (RCTs) confirmed the efficacy of mammographic screening for reducing breast cancer mortality.¹ On that basis, service screening programmes were initiated in Europe and the implementation of pilot programmes was supported by the 'Europe Against Cancer' programme.

Population-based screening according to similar protocols has commenced in most European countries. The extension of screening programmes in the various countries in the European Union (EU) has been documented in a report on the implementation of the EU policy on cancer screening.² Population-based screening, as defined in the European

Report, means that in each round of screening the eligible women in the target population in the area served by a programme are individually identified and personally invited to attend screening. Population-based screening programmes generally require a high degree of organization in order to ensure that the invitation activities are performed reliably and effectively, and are adequately coordinated with the subsequent steps of the screening process.² The population-based approach to implementation of cancer screening is recommended in the EU, because it aims to give each eligible person an equal chance of benefiting from screening and because it provides an infrastructure for effective quality assurance.^{2,3}

The majority of European countries limit screening invitations to women of 50 or more years of age, with varying upper age limits.⁴ The more challenging task of achieving an appropriate balance between benefit and harm of mammographic screening in women of younger age has been widely acknowledged in Europe. However, some countries and regions invite women under 50 years of age, and lowering the minimum age from 50 to 47 is under trial in the United Kingdom (UK). Typically, the upper age limit is

69 years. In the UK the upper age limit of 64 years was raised to 70 in 2003, and there is now a trial in England evaluating continuing screening up to age 73. Several countries, based on recent evaluations of cost-effectiveness, plan to raise the upper age limit to 73–74 years. In the Netherlands the upper age limit is 75 years.²

After the publication of the 2001 Cochrane review⁵ that questioned most of the RCTs, there was wide debate on mammographic screening. In 2002, a working group of the International Agency for Research on Cancer re-examined the available evidence and confirmed the benefit of mammographic screening in women from 50 to 69 years of age.⁶ While population-based service screening programmes have continued to be implemented without substantial changes in screening policy, there is still discussion over its effectiveness.^{7–10}

The project published in this supplement of the *Journal of Medical Screening* for the evaluation of outcomes of service screening in Europe was launched in the context of the European Screening Network. EUROSCREEN is a cooperative group that includes experts involved in planning and evaluating most of the population-based screening programmes in Europe. Based on studies published in peer-reviewed scientific journals, the experts sought to review the accumulated evidence and develop the best current estimate of the impact of population-based service screening in Europe on breast cancer mortality, and to assess screening side-effects, in particular, the risks of over-diagnosis and false-positive screening test results.

The aim was to develop an evidence-based consensus on the estimates, and to promote accurate communication of the main outcomes of breast cancer screening to stakeholders, advocates, women and other interested parties.

Breast cancer mortality

In 1989, Day *et al.*¹¹ stressed the importance of monitoring and evaluating the breast cancer screening programmes that were starting at that time. They put forward mortality reduction as the main indicator of beneficial outcome, and recommended concentrating impact evaluation on 'breast cancer deaths occurring among breast cancer cases diagnosed after the start of the screening programme'. They also discussed the problematic issue of the choice of comparison group. On this basis, the evidence from observational studies of the impact of European service screening programmes is reviewed in three papers in this supplement.^{12–14} In the paper on mortality trends, Moss *et al.*¹² discuss the methodological limitations of time trend analysis as a method for evaluation of organized mammographic screening. Due to the inclusion in the screening period of breast cancer deaths occurring in women diagnosed before screening programmes were started, the authors conclude that the analysis of breast cancer mortality trends is not adequate for evaluating the impact of screening. Njor *et al.*¹³ discuss specific methodological issues of the incidence-based mortality (IBM) approach. These studies include only breast cancer deaths occurring in women diagnosed with breast cancer after service screening started. Broeders *et al.*¹⁴ summarize all European observational studies, i.e. trend studies, IBM studies and case-control

studies, including their methodological strengths and limitations, which are well known from previous use in the evaluation of screening.^{15–18} The combined estimated mortality reduction from IBM studies was 25% (pooled relative risk [RR] 0.75, 95% Confidence interval [CI] 0.69–0.81) with invitation to screening, and 38% (pooled RR 0.62, 95% CI 0.56–0.69) with participation in screening.¹⁴ In the case-control studies, the authors found that breast cancer mortality was reduced among invited women by 31% (pooled odds ratio [OR] 0.69; 95% CI 0.57–0.83) and among screened women, after adjustment for self-selection bias, by 48% (pooled OR 0.52, 95% CI 0.42–0.65).¹⁴

Over-diagnosis risk

Over-diagnosis is usually defined as diagnosis of a breast cancer through screening, that would not have been diagnosed in the woman's lifetime had she not been screened.¹⁹ The major methodological difficulty in estimating of the rate of over-diagnosis is disentangling the excess incidence due to lead time from that due to over-diagnosis. The excess incidence due to lead time is an expected and necessary effect of breast cancer screening, in order to achieve the desired outcome of detection at a more treatable stage. In the absence of over-diagnosis, the initial increase in breast cancer occurrence in the screened group would be fully compensated by a similar decrease in cancers among older age groups no longer offered screening – the so-called 'compensatory drop'.²⁰ The compensatory drop method requires that the screening programme has been running long enough to achieve a full adjustment for lead time. If there is a short or absent follow-up after the last screen, there will be a lead time bias that will need adjustment by statistical methods.

In the paper on over-diagnosis, Puliti *et al.*²¹ show that methodological approaches to estimate over-diagnosis vary between studies and until now there has been no consensus on the optimal analytic method. The authors classify the papers by adjustment for breast cancer risk and for lead time, and conclude that the most plausible estimates of over-diagnosis range from 1% to 10%, where over-diagnosis is expressed as a percentage of the expected incidence in the absence of screening. To obtain an overall estimate for the balance sheet, we take a simple average of the individual estimates, after first transforming to represent the same disease states (invasive plus *in situ*), screening exposure measure and age group, as described below.

False-positive results

Another harm of screening for some women is recall for further investigation due to suspicious findings on the mammograms that do not result in diagnosis of breast cancer. These cases are known as false-positives. European guidelines stress the importance of control of the rate of recalls in breast cancer screening in order to limit the rate of false-positives. Lower recall rates have been observed in Europe than in the USA.^{22,23}

The recall rate is clearly highly correlated with the false-positive rate, the latter being the former minus the screen

detection rate. Assessment is defined as non-invasive if the procedures are confined to additional imaging with or without clinical examination. Assessment is defined as invasive when a needle biopsy or (rarely in the case of quality-assured population-based screening programmes) a surgical biopsy are performed. In this supplement's paper,²⁴ the cumulative risk of a false-positive recall in women undergoing 10 biennial screening tests varied between 8% and 21%, with a pooled estimate of 20% (17% without invasive assessment and 3% with invasive assessment).

THE BALANCE SHEET

Communication of benefits and harms is central to screening, and should provide the invitee with the information needed to make an informed choice about participation. A balance sheet may help policy-makers, stakeholders and especially potential participants in the target population to weigh up the benefits and harms which accrue with participation in a screening programme.

There are several possible ways to summarize the benefits and harms of screening. There is a lack of consensus on the best communication strategy that would take into account the different audiences and conditions of various screening programmes. The usual measures are estimates of the absolute number of lives saved and the number of breast cancer cases over-diagnosed in a given decision-making scenario. These estimates are sometimes transformed into the number of women who must be screened for each life saved versus the number screened for each over-diagnosed cancer. In estimation and communication of the balance sheet parameters, no judgement is made as to the relative value of a breast cancer death avoided or a case over-diagnosed – this is a matter for individual judgement. The interpretation of the screening process is complex and sometimes counterintuitive. For example, at the start of a programme the incidence rates will increase in the population. The content of the information provided to women invited to screening should be agreed between competent scientists and professionals engaged in screening, policy-makers and public opinion leaders. A representative spectrum of women in the target group should be involved in this dialogue in order to take into account the information needs of women invited to and participating in screening.

Essential components of the decision-making scenario

It is necessary to define some essential components of the screening scenario in order to develop a balance sheet that is relevant to the decision-making process in various programmes. These include the target age range, the screening regimen, periods of screening and follow-up and population denominator to which the balance sheet pertains (Table 1). These are components from the screening protocol and the longitudinal time framework to which measures can be applied to estimate the major outcomes of breast cancer service screening. Some of the components may have different values in different European countries, for example, where the age limit varies or, as in the UK, the screening interval (maximum 3 years) deviates from the EU

Table 1 Essential components of the decision-making scenario

Components	Value	Comments and communicative implications
Number of women	1000	The average number of women aged 50–51 years in a small city
Age at the start of the risk period (years)	50	Recommended starting age for service screening in Europe ²
Status in regard to screening	Screened	The outcomes in terms of benefits and harms to screened women are informative to invited women who are making the decision whether or not to attend
Number of screening mammograms expected in the screening period (interval in years)	10 (every 2 years)	Recommended number for service screening in Europe ²
Age span for screening (years)	50–69	Recommended age range for service screening in Europe ²
Age at the end of follow-up (years)	79	The outcomes in terms of benefits and harms refer to the period from 50 to 79 years

recommendation (2 years). Because estimates of screening effects may be influenced by the protocol, performance and conditions of the local screening programme, or by the characteristics of the target population, each programme may vary the components of the balance sheet taking local conditions into account. However, we consider the present scenario a good approximation for most European programmes offering screening to women in their 50s and 60s. The outcomes in terms of benefits and harms refer to women actually screened, in order to inform the decision-making process of invited women.

Measures of individual cumulative risk in the absence of screening

Estimates of the absolute numbers of lives saved and over-diagnosed cases have been obtained from the estimates of screening effects noted above and the cumulative risks of breast cancer incidence and mortality in the absence of screening shown in Table 2. These risks are defined as the probability that a woman will be diagnosed with the disease (invasive and *in situ* breast cancer) or that a woman will die from the disease, respectively, during a certain age span. To obtain these cumulative risks, we used the age-specific breast cancer incidence and mortality in women aged between 50 and 79 in the period 1985–1986 (i.e. before the start of screening programmes) in the UK, the Nordic countries (Denmark, Finland, Norway and Sweden) and areas covered by the Italian Association of Cancer Registries. The cumulative breast cancer risk estimates were, respectively, 6.7%, 6.4% and 6.9% for incidence (based on the invasive incidence rates and assuming 1 carcinoma *in situ* for 20 invasive cancers) and 3.6%,

Table 2 Balance sheet estimates of screening effects and of rates of incidence and mortality in the absence of screening

Estimation	Parameter	Reference	Comment
Cumulative risk of BC (<i>in situ</i> + invasive) from 50 to 79 years in the absence of screening	6.7%	Cancer Registry 1985–86 (UK, NordCan, Italy)	Based on the age-specific invasive BC incidence rates in the prescreening period adjusted to include one <i>in situ</i> per 20 invasive cancers
Cumulative risk of BC death from 50 to 79 years in the absence of screening	3.0%	Cancer Registry 1985–86 (UK, NordCan, Italy)	Based on the age-specific BC mortality rates in the prescreening period
Reduction in BC mortality	38–48%	Review of IBM studies and case-control studies ¹⁴	Pooled estimates for screened versus unscreened (adjusted for self-selection bias in case-control studies)
Estimate of over-diagnosis (proportion of the incidence in the absence of screening)	1–10%	Review of over-diagnosis ²⁵	Range of the six estimates adjusted for BC risk and lead time bias
Cumulative risk of a false-positive result with or without invasive assessment	3% and 17%, respectively	Review of false-positive results ²⁴	Estimated for women who participate in all of the 10 expected biennial screening tests

BC, breast cancer; IBM, incidence-based mortality

Table 3 Population, age range and cancer referred to of the six most reliable estimates of over-diagnosis, with unadjusted and adjusted over-diagnosis estimates

Study	Population	Age range	Cancers	Estimated excess due to over-diagnosis	Adjusted estimates*
Olsen <i>et al.</i> ²⁶	Screened	Screening ages	Invasive + <i>in situ</i>	7.0%	4.4%
Paci <i>et al.</i> ²⁷	Invited	Screening ages	Invasive + <i>in situ</i>	4.6%	5.9%
Waller <i>et al.</i> ²⁸	Screened	Lifetime	Invasive	10.0%	17.0%
Puliti <i>et al.</i> ²⁵	Invited	Screening ages and older	Invasive + <i>in situ</i>	1.0%	1.0%
Duffy <i>et al.</i> ²⁹	Invited	Screening ages	Invasive	3.3%	4.3%
de Gelder <i>et al.</i> ³⁰	Invited	Lifetime	Invasive + <i>in situ</i>	2.8%	6.3%

*Adjusted to apply to screened women, to 50–79 ages and to include carcinoma *in situ*

2.5% and 2.8% for mortality. Average incidence was 6.7% and average mortality 3.0%.

Estimates of screening effects

The pooled estimates of the breast cancer mortality reduction in screened women were 38% from IBM studies and 48% from case-control studies. Using the cumulative risk estimates for breast cancer mortality in the absence of screening we estimated 30 breast cancer deaths from 50 to 79 years of age (17 deaths in ages 50–69 and 13 deaths in ages 70–79). Based on East of England Cancer Registry data, 26% of breast cancer deaths occurring in the 50–69 age group were diagnosed before 50 years and 49% of breast cancer deaths occurring in women aged 70–79 were diagnosed after 70 years. Therefore, breast cancer deaths occurring at 50–79 in which screening may have had a protective effect are 19 out of the 30 expected:

$$17 \times 0.74 + 13 \times 0.51 = 19$$

Applying the estimates of mortality reduction (38% and 48%) to 19 breast cancer deaths, we obtained seven to nine lives saved. The total breast cancer mortality reduction from 50 to 79 years for a screened population varied from 23% (7/30) to 30% (9/30).

In the review on over-diagnosis,²¹ the authors conclude that the most plausible estimates of over-diagnosis (those

with adequate adjustment for incidence trends and lead time) range from 1% to 10%. Table 3 shows the estimates of over-diagnosis of the six studies,^{25–30} classified according to whether they pertain to screened or invited women, whether they refer to excess in the screening ages only or other age ranges, and whether the figures include ductal carcinoma *in situ* (DCIS). There were three major sources of variability among the estimates:

- (1) *Application to screening or invitation.* Because attendance averaged 70% in the studies used for estimation of over-diagnosis, we divided the estimates of Paci *et al.*,²⁷ Puliti *et al.*,²⁵ Duffy *et al.*²⁹ and de Gelder *et al.*³⁰ which were all based on invited rather than screened women, by 0.7.
- (2) *Application to ages 50–79 years.* Some studies estimated lifetime over-diagnosis, some over-diagnosis as a proportion of cancers detected in the screening age range and some in the screening ages and older. Around 64% of all breast cancer cases were diagnosed in ages 50–79 during the prescreening period. We therefore divided the Waller *et al.*²⁸ and de Gelder *et al.*³⁰ figures by 0.64. Age-specific incidence in the prescreening period also gives us figures with which to correct the other four estimates (details available from authors DP and SWD).
- (3) *Inclusion of all cancer, invasive + in situ, or invasive only.* The papers by Paci *et al.*²⁷ and Puliti *et al.*²⁵ estimate

over-diagnosis both including and excluding *in situ* disease. In the former, inclusion of carcinoma *in situ* conferred an increase of 1.4% in the estimate (from 3.2 to 4.6%), and in the later an increase of 1% (from 0 to 1%). We therefore added 1.4% to the estimates of Waller *et al.*²⁸ and Duffy *et al.*²⁹

The average estimate from the adjusted figures in Table 3 was 6.5% of the incidence in the absence of screening. This is a measure for over-diagnosis in screened women between 50 and 79 years, including carcinoma *in situ*, based on the studies which adequately adjusted for underlying risk and lead time.

The occurrence of DCIS is usually detected at screening and one possible reason of over-diagnosis (6, 19). Based on the age-specific breast cancer invasive incidence rates in the absence of screening (1985–1986), and assuming one carcinoma *in situ* for 20 invasive cancers, the cumulative expected number of breast cancers per 1000 women between 50 and 79 years was calculated as 67. The estimated number of over-diagnosed cases was $67 \times 0.065 = 4$.

In the review by Hofvind *et al.*,²⁴ estimates of the cumulative risk of false-positive results over 10 screens were 17% without invasive assessment and 3% with invasive assessment. These estimates pertain to women aged 50–51 at the outset who participate in 10 biennial screening tests.

Weighing up the benefits and harms

The balance sheet, based on breast cancer incidence and mortality in the absence of screening and on estimates of over-diagnosis, mortality reduction and cumulative false-positive rates derived from the results of European studies, is shown in Table 4. On average, for every 1000 women screened biennially from age 50–51, and followed until age 79, 7–9 lives are estimated to be saved and four cases are over-diagnosed. Screening between age 50 and 69 will also lead to further procedures among women not diagnosed with breast cancer (false-positives): 170 with a non-invasive assessment, and 30 with invasive assessment of which two will have a surgical biopsy. The ratio of lives saved to over-diagnosed cases is 1 to 0.4–0.6.

DISCUSSION

The breast cancer mortality reduction demonstrated in RCTs provided the evidence for initiating population-based mammographic screening programmes in Europe. There is consensus among professionals and scientists, however, that the impact of a public health programme outside of the trial environment should be evaluated in terms of its benefits and harms. Furthermore, continuous monitoring and evaluation of screening programmes to assure an appropriate balance between harm and benefit is a cornerstone of the EU policy on cancer screening.^{2,3} Harris *et al.*³¹ reconsider the criteria for evaluating screening programmes and propose a ‘balance sheet’ approach. While admitting that ‘judgment about the certainty of evidence and about the magnitude and trade-offs of benefits and harms is still

Table 4 Balance sheet for 1000 women aged 50–51 years, screened biennially until 69 years (according to the EU policy on cancer screening³) and followed until 79 years

Outcome	For every 1000 women screened for 20 years:	The number of women that need to be screened:
Number of breast cancer cases diagnosed	71	14 women: to diagnose 1 case
BC mortality reduction	7–9 women’s lives are saved (out of 30 BC deaths expected)*	111–143 women: to save 1 life
Over-diagnosis	4 cases are over-diagnosed (in addition to 67 BC expected)	250 women: to over-diagnose 1 case
False-positive test results among women without breast cancer	200 women recalled for further assessment procedures: 170 women with non-invasive assessment only 30 women with invasive assessment	6 women: to have 1 with at least one who has non-invasive assessment only 33 women: to have 1 with at least one invasive assessment

BC, breast cancer; EU, European Union

*19 out of the 30 expected BC death were diagnosed in ages 50–69

required’, they argue that a balance sheet would provide a ‘standard approach’.

In this paper, we present a balance sheet based on evidence, published in peer-reviewed journals, from European service screening programmes. The estimates of mortality reduction, over-diagnosis and false-positive screening results are based on the analysis of empirical data from areas where screening programmes have been active for several years. More comprehensive balance sheets have been proposed,^{32,33} that take into account a larger range of benefits and harms, such as social and economic effects. The focus of the present balance sheet on the chief benefit (reduction of breast cancer mortality) and harms (over-diagnosis and false-positives) does not aim to divert attention from other issues that may also be important to women invited to screening. The intention is to ensure that women are fully aware of the chief benefit and harms when they decide whether or not they wish to attend screening.

Population-based mammographic screening programmes started in some areas at the end of the 1980s (for example, the UK). Numerous additional programmes were established in Western Europe in the 1990s. Breast cancer service screening is a long-term investment. Implementation takes several years and, due to good survival rates in many European countries, final outcomes can only be fully evaluated after a long period of follow-up (at least 20 years). For this reason, the outcome evidence presented in this supplement is still evolving. The results are from selected areas and countries where screening was implemented several years ago. Given the similar organizational context, the present results are also relevant to more recently established programmes and those starting up in the future. But they clearly reflect the experience of population-based

breast cancer service screening in Europe over the last 20 years.

The analyses here do not take into account more recent innovations in breast cancer screening, such as the use of digital mammography. Particularly, in countries with ample diagnostic and therapeutic services and experience in breast cancer screening, the proportion of breast cancer clearly detected by symptoms is lower, and the distinction between screening and clinical detection of early breast cancer may be subtle. Conventional mammography, and to a greater extent digital mammography, are now more sensitive to early breast cancers or borderline lesions such as DCIS, causing both a greater potential for effectiveness in decreasing breast cancer mortality and a greater potential for over-diagnosis. Nevertheless, readers of mammograms and radiographers are likely to have received better training, and programmes are subject to more intensive quality assurance compared with the early trials,³⁴ causing a favourable shift in the balance between benefit and harm. The use of risk factors and genetic testing³⁵ has been proposed to develop tailored screening according to individual risk, although such a strategy has not been fully evaluated. More and more women are referred for assessment and treatment to multidisciplinary teams and specialized breast units, many of which came about as a by-product of, or in symbiosis with, the advent of screening programmes.

The above suggests that service screening effectiveness may be greater today³⁴ than at the time when the studies on which the present European balance sheet is based were conducted. Given the long time frame required to assess the level of over-diagnosis in a cohort of women, it is more difficult to reflect on whether rates of over-diagnosis may have changed recently.

There is increasing interest in the literature in evaluation of outcomes of service screening, but methodology across the studies is not uniform.^{29,34,36–41} This is especially true in the assessment of over-diagnosis in breast cancer service screening. A certain amount of over-diagnosis is unavoidable from early detection of lesions in an asymptomatic population. Future efforts should therefore focus on reducing the potential negative impact of over-diagnosis through better understanding of its potential determinants, including age at screening, number and frequency of tests, procedures in the assessment of positive mammograms, and the management of early or less aggressive lesions including DCIS.

There is particular interest in the potential for over-diagnosis of DCIS.⁴² Because of its good prognosis and its increased incidence in the epoch of mammographic screening, it may be intuitive to expect it to make a disproportionate contribution towards over-diagnosis. The proportion of carcinoma *in situ* varies considerably among countries and the natural history of these lesions still needs to be better understood. The magnitude of the contribution of DCIS to over-diagnosis therefore remains uncertain. However, there is some evidence that screen-detected DCIS is a progressive disease with a high likelihood of progression to invasive cancer in a large proportion of cases. In two of the randomized trials of screening, an excess of DCIS was balanced by a reduction in invasive disease.⁴³ Over a 10-year period, in one randomized trial, more than 30%

of DCIS cases treated with complete local excision alone had a recurrence or other breast cancer event.⁴⁴ There is also evidence that screen-detected DCIS is biologically more aggressive than symptomatic DCIS.⁴⁵

There is a clear need to develop appropriate protocols for patient communication, treatment and follow-up of cancers that have low life-threatening potential. Biological and clinical studies are urgently needed to improve the knowledge base for more appropriate management of these lesions, such as research on biomarkers of tumour aggressiveness. Studies are also needed to improve the specificity of the screening process and thereby reduce the other main harm of mammographic screening, resulting from false-positive tests. Research in the technology of breast imaging is therefore also recommended. The improvement of screening in terms of more selective and tailored approaches and reduction of screening harms is important from the user's perspective, but it has also important implications in cost-benefit ratio and the impact of service screening on public health costs.

Recently published balance sheets differ considerably with regard to the sources of information used for the estimates, the modality of presentation of the outcomes and the communication implications for decision-making.^{29,38–40} Furthermore, essential components of balance sheets are not always explicitly reported. For this reason, it is timely that an international expert working group assesses the impact of population-based service screening in Europe and presents a summary of the chief benefits and harms applicable to a wide range of European programmes. It is important to note that quoted benefits and harms are dependent on the intervention delivered, the data sources used and the denominator referred to. For example, our estimated numbers of lives saved exceed those estimated by Beral *et al.*⁴¹ because ours is based on 20 years of screening and on the service screening results in Europe. Those of Beral *et al.*⁴¹ pertained to 10 years of screening and the randomized trial results. Communication methods should also be improved to raise women's awareness of benefits and harms, and to make information more accessible, relevant and comprehensible.⁴⁶

To enable fully informed choices, further follow-up of service screening outcomes with well designed, coordinated, regional and national studies should be conducted, to permit more precise estimates and possibly allow variation of the expected outcomes to reflect the parameters and populations of individual programmes. Until applicable evidence for individual programmes is available, the presently reported outcomes should be taken into account when informing women offered mammographic screening in Europe.

Available cumulative evidence from population-based service screening in Europe shows that the chance of a woman's life being saved by mammographic screening is greater than that of being over-diagnosed by screening. The key outcomes and estimates of the balance of benefit and harm are presented in a narrative form in Box 1. It should be noted that the cumulative risks of breast cancer incidence and mortality in the absence of screening were calculated from the age-specific rates observed in 1985–1986. Breast cancer incidence and mortality rates have changed over time throughout Europe, due to respective

Box 1 Screening balance sheet – a narrative of benefit and harm related to screening

Consider a small city in Europe where there are 1000 female residents aged 50–51 years. If these women are followed up for 30 years until they reach the age of 79, and if a breast cancer service screening programme has not been previously established in the city, 67 women are expected to be diagnosed with breast cancer and 30 expected to die from the disease. Now consider another city with 1000 women aged 50–51 years in which a service breast screening programme has been implemented in accordance with the European guidelines. A screening invitation scheme offers all women regular mammography at intervals of two years over a period of 20 years commencing at 50–51 years of age – and all eligible women accept the invitation to screening. Among the 1000 women, 21–23 breast cancer deaths are expected over a 30-year period compared with the 30 breast cancer deaths that would have occurred in the absence of screening (i.e. 7–9 fewer deaths).

Most of the women participating in screening will have only negative mammograms and, therefore, will have no benefits other than a reassurance about their health status, and only short-term harms from service screening (discomfort, anxiety). Women with a positive mammogram can have both benefits and harms.

Over-diagnosis is arguably a significant harm. It implies diagnosis of breast cancer without benefit for a woman, because if she had not participated in service screening the cancer would not have been diagnosed during her lifetime, nor received treatments for it. In our estimation, four women are harmed in this way in the city that offers screening; this is in addition to the 67 women who would be diagnosed with breast cancer in the absence of service screening. Cases of over-diagnosis usually arise from the detection in service screening of less aggressive tumours early in their development, for example, *in situ* and very early invasive tumours, but at the present time it is not possible to distinguish between those tumours that will become aggressive and those that pose no threat to life. For this reason, research into prognostic indicators and evidence-based protocols that will reduce the impact of breast cancer treatment are current research priorities.

Another potential harm for screened women is a ‘false-positive’ result, where the screening mammogram shows signs of suspicious abnormalities, but the assessment during her recall visit shows that she has no cancer – her initial result was a false alarm. Although strictly speaking, a suspicious finding of screening mammogram is not a positive diagnosis, the mammogram is referred to as ‘false-positive’ because it prompted further investigation.

Assessment at recall often involves additional imaging, such as a second mammogram or an ultrasound scan, and may also involve clinical examination of the breast, but an invasive procedure is sometimes

appropriate before a final diagnosis can be made. The invasive procedure is usually needle biopsy to obtain tissue for histopathological examination, but occasionally there is surgical intervention. The psychological harm of this process – recall and assessment – has been investigated but there is only evidence of short-term anxiety. The cumulative proportion of women having at least one false-positive result over 20 years of screening has been estimated in some European service screening programmes as 17% for assessment involving only non-invasive procedures, and 3% for assessment also involving invasive procedures. Because such false alarms are inevitable and inherent in any screening process, the European guidelines set standards for the proportion of screened women who are recalled for assessment (recall rate) and recommend that service screening performance indicators are readily available to the invited population. Long-term effects such as radiation risk are not considered here, as they tend to be based on extrapolation and are subject to considerable uncertainty, but the risk is estimated to be much lower than the benefit of screening.

changes in the prevalence of risk factors and therapeutic regimens. Consequently, the estimates of lives saved and over-diagnosed cases are approximations. They clearly indicate, however, that the relationship between benefit and harm of mammographic screening is much more favourable than some recent publications suggest.^{8,9,36–37}

CONCLUSIONS

A comprehensive review of European studies of service screening shows that population-based mammographic screening is contributing to the reduction in breast cancer mortality in the EU, achieving a mortality benefit at least as great as that observed in the RCTs. Furthermore, the estimated number of breast cancer deaths avoided by screening clearly exceeds the estimated number of cases of over-diagnosis. These results are intended to help a woman who is invited to screening to make an informed personal choice about the possible outcomes and the implications of participating in screening.

EUROSCREEN WORKING GROUP

Coordinators

E Paci, M Broeders, S Hofvind and S W Duffy

Members

Ancelle-Park, R (F)¹, Armaroli P (I)², Ascunce N (E)³, Bisanti, L (I)⁴, Bellisario C (I)², Broeders M (NL)⁵, Cogo C (I)⁶, de Koning H (NL)⁷, Duffy S W (UK)⁸, Frigerio A (I)², Giordano L (I)², Hofvind S (N)⁹, Jonsson H (S)¹⁰, Lyngø E (DK)¹¹, Massat N (UK)⁸, Miccinesi G (I)¹², Moss S (UK)⁸, Naldoni C (I)¹³, Njor S (DK)¹¹, Nyström L (S)¹⁴, Paap E

(NL)⁵, Paci E (I)¹², Patnick J (UK)¹⁵, Ponti A (I)², Puliti D (I)¹², Segnan N (I)², Von Karsa L (D)¹⁶, Törnberg S (S)¹⁷, Zappa M (I)¹², Zorzi M (I)⁶

Affiliations

¹Ministère du travail de l'emploi et de la santé, Paris, France

²CPO-Piedmont, Turin, Italy

³Navarra Breast Cancer Screening Programme, Pamplona, Spain

⁴S.C. Epidemiologia, ASL di Milano, Italy

⁵Radboud University Nijmegen Medical Centre & National Expert and Training Centre for Breast Cancer Screening, Nijmegen, The Netherlands

⁶Veneto Tumor Registry, Padua, Italy

⁷Erasmus MC, Dept. of Public Health, Rotterdam, The Netherlands

⁸Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

⁹Cancer Registry of Norway, Research Department and Oslo and Akershus University College of Applied Science, Oslo, Norway

¹⁰Department of Radiation Sciences, Oncology, Umeå University, Umeå, Sweden

¹¹Centre for Epidemiology and Screening, University of Copenhagen, Copenhagen, Denmark

¹²ISPO Cancer Research and Prevention Institute, Florence, Italy

¹³Regional Cancer Screening Center, Emilia-Romagna Region, Bologna, Italy

¹⁴Department of Public Health and Clinical Medicine, Division of Epidemiology and Global Health, Umeå University, Umeå, Sweden

¹⁵NHS Cancer Screening Programmes and Oxford University, UK

¹⁶International Agency for Research on Cancer, Lyon, France

¹⁷Stockholm Cancer Screening, Stockholm, Sweden

ACKNOWLEDGEMENTS

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

REFERENCES

- Wald N, Frost C, Cuckle H. Breast cancer screening: The current position. *BMJ* 1991;**302**:845–6
- von Karsa L, Anttila A, Ronco G, et al. Cancer screening in the European Union. Report on the Implementation of the Council Recommendation on Cancer Screening – First Report. European Commission, Luxembourg, 2008
- Council of the European Union. Council Recommendation of 2 December 2003 on cancer screening (2003/878/EC). *Off J Eur Union*. 2003;**L327**:34–8
- Giordano L, von Karsa L, Tomatis M, et al. Mammographic screening programmes in Europe: Organisation, coverage and participation. *J Med Screen* 2012 (in press)
- Olsen O, Gøtzsche PC. Screening for breast cancer with mammography. *Cochrane Database Syst Rev* 2001;(4):CD001877. Review
- IARC Handbooks of Cancer Prevention. *Breast cancer screening*. Vol 7. IARC: Lyon, 2002
- McPherson K. Screening for breast cancer – balancing the debate. *BMJ* 2010;**340**:c3106
- Jørgensen KJ, Zahl PH, Gotzsche PC. Breast cancer mortality in organized mammography screening in Denmark: Comparative study. *BMJ* 2010;**340**:c241
- Suhrke P, Mæhlen J, Schlichting E, Jørgensen KJ, Gøtzsche PC, Zahl PH. Effect of mammography screening on surgical treatment for breast cancer in Norway: Comparative analysis of cancer registry data. *BMJ* 2011;**343**:d4692
- Kopans DB, Smith RA, Duffy SW. Mammography screening and 'overdiagnosis'. *Radiology* 2011;**260**:616–20
- Day NE, Williams DR, Khaw KT. Breast cancer screening programmes: The development of a monitoring and evaluation system. *Br J Cancer* 1989;**59**:954–8
- Moss S, Nyström L, Jonsson H, et al. The EUROSCREEN Working Group. A review of trend studies to evaluate the impact of mammographic screening on breast cancer mortality in Europe. *J Med Screen* 2012 (in press)
- Njor S, Nyström L, Moss S, et al. The EUROSCREEN Working Group. Breast cancer mortality after start of service mammography screening in Europe: A review of incidence based mortality studies. *J Med Screen* 2012 (in press)
- Broeders M, Moss S, Nyström L, et al. The EUROSCREEN Working Group. Review of the impact of population-based screening with mammography on breast cancer mortality in Europe. *J Med Screen* 2012 (in press)
- Paap E, Verbeek ALM, Puliti D, Paci E, Broeders MJM. Breast cancer screening case-control study design: Impact on breast cancer mortality. *Ann Oncol* 2011;**22**:863–9. Epub 5 October 2010
- Demissie K, Mills OF, Rhoads GG. Empirical comparison of the results of randomized controlled trials and case-control studies in evaluating the effectiveness of screening mammography. *J Clin Epidemiol* 1998;**51**:81–91
- Connor RJ, Boer R, Prorok PC, Weed DL. Investigation of design and bias issues in case-control studies of cancer screening using microsimulation. *Am J Epidemiol* 2000;**151**:991–8
- Verbeek AL, Hendriks JH, Holland R, Mravunac M, Sturmans F, Day NE. Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975–81. *Lancet* 1984;**1**:1222–4
- Paci E, Duffy S. Overdiagnosis and overtreatment of breast cancer: Overdiagnosis and overtreatment in service screening. *Breast Cancer Res* 2005;**7**:266–70
- Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdiagnosis with mammography screening: A systematic review. *Lancet Oncol* 2007;**8**:1129–38
- Puliti D, Duffy S, Miccinesi G, et al. The EUROSCREEN Working Group. Methodology and estimate of overdiagnosis in breast cancer service screening: A review of the European studies. *J Med Screen* 2012 (in press)
- Smith-Bindman R, Ballard-Barbash R, Miglioretti DL, Patnick J, Kerlikowske K. Comparing the performance of mammography screening in the USA and the UK. *J Med Screen* 2005;**12**:50–4
- Hofvind S, Vacek PM, Skelly J, Weaver DL, Geller BM. Comparing screening mammography for early breast cancer detection in Vermont and Norway. *J Natl Cancer Inst* 2008;**100**:1082–91
- Hofvind S, Ponti A, Patnick J, et al. The EUNICE/EUROSCREEN Working Group. False positive results in mammographic screening for breast cancer. *J Med Screen* 2012 (in press)
- Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* 2009;**45**:3166–71
- Olsen AH, Agbaje OF, Myles JP, Lyng E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast* 2006;**12**:338–42
- Paci E, Miccinesi G, Puliti D, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. *Breast Cancer Res* 2006;**8**:R68
- Waller M, Moss S, Watson J, Moller H. The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. *Cancer Epidemiol Biomarkers Prev* 2007;**16**:2257–61
- Duffy SW, Tabar L, Olsen AH, et al. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. *J Med Screen* 2010;**17**:25–30
- de Gelder R, Heijnsdijk EA, van Ravestein NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* 2011;**33**:111–21
- Harris R, Sawaya GF, Moyer VA, Calonge N. Reconsidering the criteria for evaluating proposed screening programs: Reflections from 4 current and

- former members of the U.S. Preventive services task force. *Epidemiol Rev* 2011;**33**:20–35
- 32 Austoker J. Cancer prevention: Setting the scene. *BMJ* 1994;**308**:1415–20
- 33 van der Maas PJ, de Koning HJ, van Ineveld BM, *et al.* The cost-effectiveness of breast cancer screening. *Int J Cancer* 1989;**43**:1055–60
- 34 van Schoor G, Moss SM, Otten JD, *et al.* Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer* 2011;**104**:910–4
- 35 Pashayan N, Duffy SW, Chowdhury S, *et al.* Polygenic susceptibility to prostate and breast cancer: Implications for personalised screening. *Br J Cancer* 2011;**104**:1656–63
- 36 Autier P, Boniol M, La Vecchia C, *et al.* Disparities in breast cancer mortality trends between 30 European countries: Retrospective trend analysis of WHO mortality database. *BMJ* 2010;**341**:c3620
- 37 Autier P, Boniol M, Gavin A, Vatten LJ. Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: Trend analysis of WHO mortality database. *BMJ* 2011;**343**:d4411
- 38 Welch HG. Screening mammography – a long run for a short slide? *N Engl J Med* 2010;**363**:1276–8
- 39 Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;**102**:605–13. Review
- 40 Gotzsche PC, Hartling OJ, Nielsen M, Brodersen J, Jorgensen KJ. Breast screening: The facts – or maybe not. *BMJ* 2009;**338**:b86
- 41 Beral V, Alexander M, Duffy S, *et al.* The number of women who would need to be screened regularly by mammography to prevent one death from breast cancer. *J Med Screen* 2011;**18**:210–2
- 42 Jones JL. Overdiagnosis and overtreatment of breast cancer: Progression of ductal carcinoma in situ: The pathological perspective. *Breast Cancer Res* 2006;**8**:204
- 43 Duffy SW, Agbaje O, Tabar L, *et al.* Estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res* 2005;**7**:258–5
- 44 Cuzick J, Sestak I, Pinder SE, *et al.* Effect of tamoxifen and radiotherapy in women with locally excised ductal carcinoma in situ: Long-term results from the UK/ANZ DCIS trial. *Lancet Oncol* 2011;**12**:21–9
- 45 de Roos MA, van der Vegt B, de Vries J, Wesseling J, de Bock GH. Pathological and biological differences between screen-detected and interval ductal carcinoma in situ of the breast. *Ann Surg Oncol* 2007;**14**:2097–104
- 46 Giordano L, Cogo C, Patnick J, Paci E, The Euroscreen Working Group. Communicating the balance sheet. *J Med Screen* 2012 (in press)