BJC

The benefits and harms of breast cancer screening: an independent review

A report jointly commissioned by Cancer Research UK and the Department of Health (England) October 2012.

M G Marmot*,[1], D G Altman[2], D A Cameron[3], J A Dewar[4], S G Thompson[5], M Wilcox[6] – The Independent UK Panel on Breast Cancer Screening

[1]UCL Department of Epidemiology and Public Health, UCL Institute of Health Equity, 1-19 Torrington Place, London WC1E 7HB, UK; [2]Centre for Statistics in Medicine, University of Oxford, Botnar Research Centre, Windmill Road, Oxford, OX3 7LD, UK; [3]University of Edinburgh Cancer Research Centre and NHS Lothian, Western General Hospital, Edinburgh, EH4 2XR, UK; [4]Department of Surgery and Molecular Oncology, Medical School, Ninewells Hospital, Dundee DD1 9SY, UK; [5]Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 8RN, UK and [6]Independent Cancer Patient's Voice, 17 Woodbridge Street, London EC1R 0LL, UK

1. SUMMARY

1.1 Introduction

The breast cancer screening programmes in the United Kingdom currently invite women aged 50–70 years for screening mammography every 3 years. Since the time the screening programmes were established, there has been debate, at times sharply polarised, over the magnitude of their benefit and harm, and the balance between them. The expected major benefit is reduction in mortality from breast cancer. The major harm is overdiagnosis and its consequences; overdiagnosis refers to the detection of cancers on screening, which would not have become clinically apparent in the woman's lifetime in the absence of screening.

Professor Sir Mike Richards, National Cancer Director, England, and Dr Harpal Kumar, Chief Executive Officer of Cancer Research UK, asked Professor Sir Michael Marmot to convene and chair an independent panel to review the evidence on benefits and harms of breast screening in the context of the UK breast screening programmes. The panel, authors of this report, reviewed the extensive literature and heard testimony from experts in the field who were the main contributors to the debate.

The nature of information communicated to the public, which too has sparked debate, was not part of the terms of reference of the panel, which are listed in Appendix 1.

1.2 Relative mortality benefit

The purpose of screening is to advance the time of diagnosis so that prognosis can be improved by earlier intervention. A consequence of earlier diagnosis is that it increases the apparent incidence of breast cancer in a screened population and extends the average time from diagnosis to death, even if screening were to confer no benefit. The appropriate measure of benefit, therefore, is reduction in mortality from breast cancer in women offered screening compared with women not offered screening.

In the panel's judgement, the best evidence for the relative benefit of screening on mortality reduction comes from 11 randomised controlled trials (RCTs) of breast screening. Meta-analysis of these trials with 13 years of follow-up estimated a 20% reduction in breast cancer mortality in women invited for screening. The relative reduction in mortality will be higher for women actually attending screening, but by how much is difficult to say because women who do not attend are likely to have a different background risk. Three types of uncertainties surround this estimate of 20% reduction in breast cancer mortality. The first is statistical: the 95% confidence interval (CI) around the relative risk (RR) reduction of 20% was 11–27%. The second is bias: there are a number of potential sources of distortion in the trials that have been widely discussed in the literature ranging from suboptimal randomisation to problems in adjudicating cause of

*Correspondence: Professor M Marmot; E-mail: m.marmot@ucl.ac.uk

death. The third is the relevance of these old trials to the current screening programmes. The panel acknowledged these uncertainties, but concluded that a 20% reduction is still the most reasonable estimate of the effect of the current UK screening programmes on breast cancer mortality. Most other reviews of the RCTs have yielded similar estimates of relative benefit.

The RCTs were all conducted at least 20–30 years ago. More contemporary estimates of the benefit of breast cancer screening come from observational studies. The panel reviewed three types of observational studies. The first were ecological studies comparing areas, or time periods, when screening programmes were and were not in place. These have generated diverse findings, partly because of the major advances in treatment of breast cancer, which have a demonstrably larger influence on mortality trends than does screening, and partly because of the difficulty of excluding imbalances in other factors that could affect breast cancer mortality. The panel did not consider these studies helpful in estimating the effect of screening on mortality. The other two types of studies, case–control studies and incidence-based mortality studies, showed breast screening to confer a greater benefit than did the trials. Although these studies, in general, attempted to control for non-comparability of screened and unscreened women, the panel was concerned that residual bias could inflate the estimate of benefit. However, the panel notes that these studies' findings are in the same direction as the trials.

1.3 Absolute mortality benefit

Estimates of absolute benefit of screening have varied from one breast cancer death avoided for 2000 women invited to screening to 1 avoided for about 100 women screened, about a 20-fold difference. Major determinants of that large variation are the age of women screened, and the durations of screening and follow-up. The age of the women invited is important, as mortality from breast cancer increases markedly with age. The panel therefore applied the relative mortality reduction of 20% to achieve the observed cumulative absolute risk of breast cancer mortality over the ages 55–79 years for women in the United Kingdom, assuming that women who began screening at 50 years would gain no benefit in the first 5 years, but that the mortality reduction would continue for 10 years after screening ended. This yielded the estimate that for every 235 women invited to screening, one breast cancer death would be prevented; correspondingly 180 women would need to be screened to prevent one breast cancer death. Uncertainties in the figure of a 20% RR reduction would carry through to these estimates of absolute mortality benefit. Nonetheless, the panel's estimate of benefit is in the range of one breast cancer death prevented for $\sim$250 women invited, rather than the range of 1 in 2000.

1.4 Overdiagnosis

The major harm of screening considered by the panel was that of overdiagnosis. Given the definition of an overdiagnosed cancer, either invasive or non-invasive, as one diagnosed by screening, which would not otherwise have come to attention in the woman's lifetime, there is need for a long follow-up to assess the frequency of overdiagnosis. In the view of the panel, some cancers detected by screening will be overdiagnosed, but the uncertainty surrounding the extent of overdiagnosis is greater than that for the estimate of mortality benefit because there are few sources of reliable data. The issue for the UK screening programmes is the magnitude of overdiagnosis in women who have been in a screening programme from age 50 to 70, then followed for the rest of their lives. There are no data to answer this question directly. Any estimate will therefore be, at best, provisional.

Although the definition of an overdiagnosed case, and thus the numerator in a ratio, is clear, the choice of denominator has been the source of further variability in published estimates. Different studies have used: only the cancers found by screening; cancers found during the whole screening period, both screen-detected and interval; cancers

diagnosed during the screening period and for the remainder of the women's lifetime. The panel focused on two estimates: the first from a population perspective using as the denominator the number of breast cancers, both invasive and ductal carcinoma *in situ* (DCIS), diagnosed throughout the rest of a woman's lifetime after the age that screening begins, and the second from the perspective of a woman invited to screening using the total number of breast cancers diagnosed during the screening period as the denominator.

The panel thought that the best evidence came from three RCTs that did not systematically screen the control group at the end of the screening period and followed these women for several more years. The frequency of overdiagnosis was of the order of 11% from a population perspective, and about 19% from the perspective of a woman invited to screening. Trials that included systematic screening of the control group at the end of the active part of the trial were not considered to provide informative estimates of the frequency of overdiagnosis.

Information from observational studies was also considered. One method that has been used is investigation of time trends in incidence rates of breast cancer for different age groups over the period that population screening was introduced. The published results of these studies varied greatly and have been interpreted as providing either reassurance or cause for alarm. So great was the variation in results that the panel conducted an exercise by varying the assumptions and statistical methods underlying these studies, using the same data sets; estimates of overdiagnosis rates were found to vary across the range of 0–36% of invasive breast cancers diagnosed during the screening period. The panel had no reason to favour one set of estimates over another, and concluded that this method could give no reliable estimate of the extent of overdiagnosis.

Were it possible to distinguish at screening those cancers that would not otherwise have come to attention from those that, untreated, would lead to death, the overdiagnosis problem could be much reduced, at least in terms of unnecessary worry and treatment. Currently this is not possible, so neither the woman nor her doctor can know whether a screen-detected cancer is an 'overdiagnosed' case or not. In particular, DCIS, most often diagnosed at screening, does not inevitably equate to overdiagnosis – screen-detected DCIS, after wide local excision (WLE) only, is associated with subsequent development of invasive breast cancer in 10% of women within 10 years.

The consequences of overdiagnosis matter, women are turned into patients unnecessarily, surgery and other forms of cancer treatment are undertaken, and quality of life and psychological well being are adversely affected.

1.5 The balance of benefit and harm

The panel estimates that an invitation to breast screening delivers about a 20% reduction in breast cancer mortality. For the UK screening programmes, this currently corresponds to about 1300 deaths from breast cancer being prevented each year, or equivalently about 22 000 years of life being saved. However, this benefit must be balanced against the harms of screening, especially the risk of overdiagnosis. In the panel's view, overdiagnosed cancers certainly occur, but the frequency in a screening programme of 20 years duration is unknown. Estimates from trials of shorter duration suggest overdiagnosis of about 11% as a proportion of breast cancer incidence during the screening period and for the remainder of the woman's lifetime, or equivalently about 19% as a proportion of cancers diagnosed during the screening period. Any excess mortality stemming from the investigation and treatment of breast cancer is considered by the panel to be small and considerably outweighed by the benefits of treatment. Some other harms, including increased anxiety and discomfort caused by screening, are also acknowledged.

Notionally, for 10 000 women invited to screening, from age 50 for 20 years, it is estimated that 681 cancers (invasive and DCIS)

will be diagnosed, of which 129 will represent overdiagnosis (using the 19% estimate of overdiagnosis) and 43 deaths from breast cancer will be prevented.

Given that the treatment for breast cancer has improved, is screening no longer relevant? The panel's view is that the benefits of screening and those of better treatments are reasonably considered independent. Uncertainty about possible interaction between the benefits of screening and of contemporary treatments is not a reason for stopping breast screening.

The panel was not asked to comment on costs, both of interventions and the consequences of overdiagnosis. With accurate figures an estimate of cost-benefit could be made and compared with other interventions, but would be a significant piece of work in its own right.

An individual woman cannot know whether she is one of the numbers who will benefit or be harmed from screening. If she chooses to be screened, it should be in the knowledge that she is accepting the chance of benefit, having her life extended, knowing that there is also a risk of overdiagnosis and unnecessary treatment. Similarly, a woman who declines the invitation to screening needs to recognise that she runs a slightly higher risk of dying from breast cancer.

1.6 Conclusions and recommendations

Breast screening extends lives. The panel's review of the evidence on benefit – the older RCTs, and those more recent observational studies – points to a 20% reduction in mortality in women invited to screening. A great deal of uncertainty surrounds this estimate, but it represents the panel's overview of the evidence. This corresponds to one breast cancer death averted for every 235 women invited to screening for 20 years, and one death averted for every 180 women who attend screening.

The panel's best estimate is that the breast screening programmes in the United Kingdom, inviting women aged 50–70 every 3 years, prevent about 1300 breast cancer deaths a year, a most welcome benefit to women and to the public health.

However, there is a cost to women's well being. In addition to extending some lives by early detection and treatment, mammographic screening detects cancers, proven to be cancers by pathological testing, that would not have come to clinical attention in the woman's life, were it not for screening - called overdiagnosis. The consequence of overdiagnosis is that women have their cancer treated by surgery, radiotherapy and medication, but neither the woman nor her doctor can know whether this particular cancer would be one that could possibly lead to death, or one that would have remained undetected for the rest of the woman's life.

The panel sought to estimate the level of overdiagnosis in women screened for 20 years and followed to the end of their lives. Estimates of overdiagnosis abound, from near to zero to 50%, but there is a paucity of reliable data to answer this question. There has not even been agreement on how to measure overdiagnosis. On the basis of follow-up of three RCTs, the panel estimated that in women invited to screening, about 11% of the cancers diagnosed in their lifetime constitute overdiagnosis, and about 19% of the cancers diagnosed during the period that women are actually in the screening programme; but the panel emphasises these figures are the best estimates from a paucity of reliable data.

Putting together benefit and overdiagnosis from the above figures, the panel estimates that for 10 000 UK women invited to screening from age 50 for 20 years, about 681 cancers will be found of which 129 will represent overdiagnosis, and 43 deaths from breast cancer will be prevented. In round terms, therefore, for each breast cancer death prevented, about three overdiagnosed cases will be identified and treated. Of the ~307 000 women aged 50–52 who are invited to screening each year, just >1% would have an overdiagnosed cancer during the next 20 years. Given the uncertainties around the estimates, the figures quoted give a spurious impression of accuracy.

The panel concludes that the UK breast screening programmes confer significant benefit and should continue. The greater the proportion of women who accept the invitation to be screened, the greater is the benefit to the public health in terms of reduction in mortality from breast cancer. However for each woman the choice is clear: on the plus side screening confers a likely reduction in mortality from breast cancer because of early detection and treatment. On the negative side, is the knowledge that she has perhaps a 1% chance of having a cancer diagnosed, and treated with surgery and other modalities, which would never have caused problems had she not been screened.

Evidence from a focus group conducted by Cancer Research UK and attended by two panel members, and in line with previous similar studies, was that this was an offer many women will feel is worth accepting: the treatment of overdiagnosed cancer may cause suffering and anxiety, but that suffering is worth the gain from the potential reduction in breast cancer mortality. Clear communication of these harms and benefits to women is of utmost importance and goes to the heart of how a modern health system should function. There is a body of knowledge on how women want information presented, and this should inform the design of information to the public.

2. INTRODUCTION

2.1 The UK NHS breast screening programmes

The NHS breast cancer screening programme in England began inviting women to be screened in 1988. This followed the recommendations made by Professor Sir Patrick Forrest in his report on breast screening in 1986 (Forrest, 1986). The breast screening programmes in the United Kingdom currently invite women aged 50–70 years for a screening mammography every 3 years. The mammography is designed to detect changes in the breast tissue that may indicate the presence of cancer. The screening programme in England is currently conducting a randomised trial to ascertain whether there would be benefit in extending the age at which women are invited to 47–73 years.

2.2 Principles of screening

Screening is concerned with the detection of disease at an early stage, with the expectation that treatment will be more effective if begun earlier in the disease process. Screening is therefore based on the principle of there being an effective treatment. It is well recognised that an apparent benefit of increased survival time could be illusory because of simply bringing forward the time of diagnosis without changing the course of the disease. Therefore, the appropriate way to assess benefit is to look at breast cancer mortality of screened and unscreened cohorts rather than just survival time from diagnosis (see section 3).

As the principle of screening is to diagnose cases earlier, at any particular time point during the period of successive screenings, there will be more cases of breast cancer in a group of screened women compared with a similar group of unscreened women. However, it is possible that some of these additional cases may be cancers that would not otherwise have been diagnosed or caused the woman any problem during her lifetime. These cancers are referred to as overdiagnosis (see section 4).

2.3 The debate over benefits and harms of breast screening

Since the screening programmes were established, there has been debate over the potential benefits and harms. Recently, the debate has focussed on the reduction in mortality attributable to screening, the numbers of women overdiagnosed, and the way that the risks and benefits are communicated to women invited for screening. The arguments have become quite polarised between those who believe that the benefit of decreased breast cancer mortality outweighs the harms and those who believe the harms outweigh the benefit. These differing views of the evidence have arisen, in part, from disagreements over the validity and applicability of the available RCTs of breast screening, and from questions about the usefulness and interpretation of observational data on breast cancer incidence and mortality.

The debate over the benefits and harms of breast screening is not unique to the UK and the NHS breast screening programmes. In 2002, the International Agency for Research on Cancer at the World Health Organisation reviewed the evidence on breast screening, and put forward recommendations on further research and on implementing screening programmes (IARC, 2002). The US Preventive Services Task Force in 2009 re-examined the efficacy of different screening modalities. They recommended that women under the age of 50 not be routinely screened, and that women aged 50–74 have biennial rather than annual screens (Woolf, 2010). The Canadian Taskforce on Preventative Health Care updated their guidelines on breast screening in 2011, and concluded that the reduction in mortality associated with screening mammography is small for women aged 40–74 years at average risk of breast cancer. They also found a greater reduction in mortality for women aged ⩾50 compared with those <50, and that harms of overdiagnosis and unnecessary biopsy may be greater for younger women than for older women. They recommended that women aged 50–74 be routinely screened but state that appreciable uncertainty exists around the evidence for this (Canadian Task Force on Preventive Health Care, 2011). Published reports from the Nordic Cochrane Centre concluded that, despite their substantial methodological limitations, the trials of screening showed that screening saved lives, but at the cost of considerable harm from overdiagnosis (Gøtzsche and Nielsen, 2011).

2.4 Breast cancer in the UK

Incidence and mortality

In the United Kingdom, breast cancer remains the most commonly diagnosed cancer in women (48 417 cases in 2009) and is the second most common cause of death from cancer in women (11 556 deaths in 2010). UK breast cancer incidence rates have been rising in all age groups since the late 1970s (Figure 1A). The causes of these increasing rates are thought to include: increased use of hormone replacement therapy; later age at child birth; lower parity; and increasing obesity and alcohol intake in women. Also, there is believed to be better ascertainment, especially in older women. In common with most countries, the introduction of the screening programme for women aged 50–64 in 1988 and those aged 65–70 in 2001 led to additional increases in incidence (Figure 1A).

By contrast with incidence rates, since the early 1990s, the mortality rates for breast cancer have been decreasing – shown both as annual mortality rates and 35-year cumulative risk of dying from breast cancer (Figure 1B). It is believed that the causes of these decreases may include: improvement in treatment, in particular adjuvant therapies; specialisation and better organisation of cancer care; screening; and increased breast awareness (Appendix 2).

Contribution of screening to decreased breast cancer mortality

It is widely agreed that screening alone cannot be the major factor responsible for the decrease in breast cancer mortality over the last
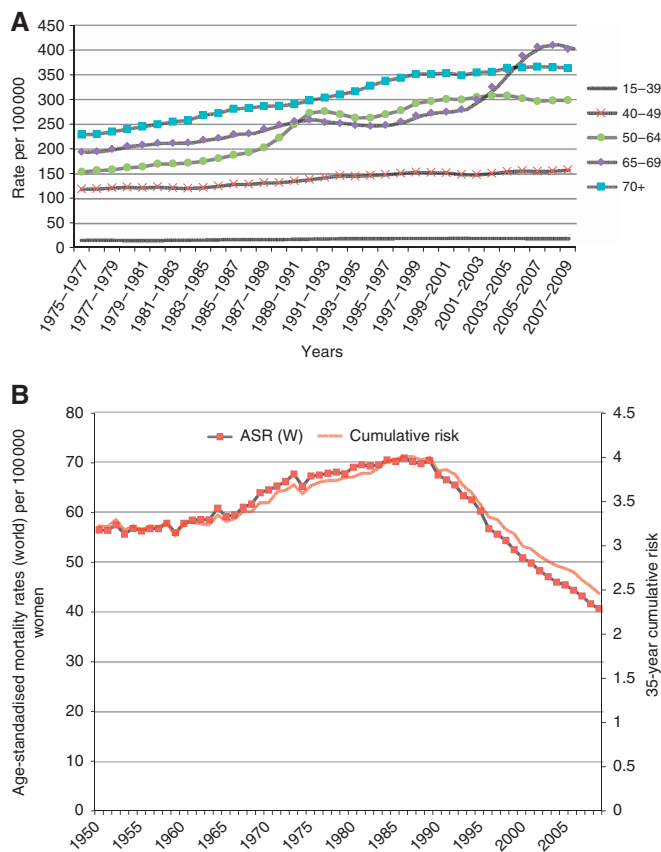
Figure 1. (**A**) European age-standardised incidence rates per 1 00 000 population, females, by age, Great Britain. (**B**) Breast cancer mortality at ages 35–69, UK, 1950–2009 (World Health Organization, 2012).

20 years. Improvements in treatment and service delivery are likely to have made the largest contribution to decreased mortality (Berry *et al*, 2005). Indeed, without effective treatment, screening for breast cancer is redundant. However, it is important to establish what contribution, if any, screening makes, given that it requires the use of substantial resources within the health system, and nearly two million women each year in England alone accept the invitation and agree to be screened (The NHS Information Centre, 2012).

2.5 Independent review of breast screening

It is within this context that Professor Sir Mike Richards, National Cancer Director, England and Dr Harpal Kumar, Chief Executive Officer of Cancer Research UK, asked Professor Sir Michael Marmot to chair an independent panel to review breast screening. The panel's terms of reference are shown in Appendix 1. This panel has reviewed the extensive literature and heard testimony from many of the experts in the field. This report details its findings and recommendations for the breast screening programme in England.

2.6 Independent review panel membership

The independent panel consisted of nationally and internationally recognised experts in epidemiology and/or medical statistics, as well as in current breast cancer diagnosis and treatment practices. A patient advocate was an integral member of the panel. No panel member had previously published on breast screening, thus helping to ensure an objective and independent assessment of the evidence.

The panel was chaired by Professor Sir Michael G Marmot, Director of the Institute of Health Equity, University College London; Chair, WHO Commission on Social Determinants of Health; Chair, Marmot Review – Strategic Review of Health

Inequalities in England after 2010; Chair, European Review on the Social Determinants of Health and the Health Divide; MRC Research Professor of Epidemiology and Public Health, University College London with long-standing research on social determinants of health and health inequalities.

The other panellists were:

Professor Douglas G Altman, Director of the Centre for Statistics in Medicine and Cancer Research UK Medical Statistics Group, University of Oxford. Doug's varied research interests include the use and abuse of statistics in medical research, studies of prognosis, regression modelling, systematic reviews, randomised trials, and studies of medical measurement. He is actively involved in efforts to improve the quality of scientific publications by promoting transparent and accurate reporting of health research.

Professor David A Cameron, Clinical Director of the Edinburgh Cancer Research Centre, Director of Cancer Services at NHS Lothian, and Professor of Oncology at Edinburgh University. Previously, David was the Director of the NIHR National Cancer Research Network and Professor of Oncology at Leeds University. His research interests are in translational and clinical trials in breast cancer, and he is the principal investigator of several clinical trials looking at treatment of early breast cancer. Before qualifying as a medical doctor, he completed an undergraduate degree in Mathematics.

Professor John A Dewar, Consultant and honorary Professor of Clinical Oncology. Until recently, John was Head of Oncology at Ninewells Hospital, Dundee. John has a long-standing interest in the management of patients with breast cancer and has been closely involved in clinical trials of both radiotherapy and systemic therapy for breast cancer.

Professor Simon G Thompson, Director of Research in Biostatistics at the University of Cambridge. Simon's research interests are in meta-analysis and evidence synthesis, clinical trial methodology, health economic evaluation, and cardiovascular epidemiology. He has collaborated on a number of major clinical trials, recently including all the major UK national trials of screening and treatment for abdominal aortic aneurysms.

Maggie Wilcox, patient advocate. Maggie was a health visitor for many years before working as Clinical Nurse Specialist in palliative care before her breast cancer diagnosis in 1997. After early retirement following her treatment, she became involved in patient advocacy in cancer services and research. She now provides a patient voice at national and local level as a member of various organisations, including the National Cancer Research Institute Breast Clinical Study Group and the Surrey, West Sussex and Hampshire Network Breast Site Specific Group.

2.7 Independent review process and role of secretariat

As set out in the review's terms of reference, the secretariat provided initial key literature on breast cancer screening, including publications recommended from both sides of the debate. The panel then called on a range of experts (see Appendix 1 for full list) to give evidence.

Cancer Research UK and the Department of Health provided the secretariat function for the review comprising:

- Dr Dulcie McBride, Consultant in Public Health Medicine, Department of Health
- Sara Hiom, Director of Information, Cancer Research UK
- Nick Ormiston-Smith, Data Analysis and Research Manager, Cancer Research UK
- Dr Martine Bomb, Programme Manager, Cancer Research UK
- Samantha Harrison, Programme Officer, Cancer Research UK

The secretariat acted purely as support to the panel in the practical, writing, and dissemination functions and having no say in the conclusions or recommendations. Further information can be found in Appendix 1.

3. THE EFFECT OF BREAST SCREENING ON MORTALITY

This section summarises the panel's views of the effect of breast screening on mortality. Specifically, the aim is to estimate the effect of the current national screening programmes in the United Kingdom on breast cancer mortality. Estimates of relative risk reduction, absolute risk reduction, and increase in life expectancy are discussed.

3.1 Introduction

Randomised controlled trials potentially provide the most reliable information about the effects of breast screening. Well-conducted RCTs are prone to fewer distorting effects, or biases, than observational studies. Systematic reviews and meta-analyses of RCTs are widely accepted as the highest level of evidence for guiding policy decisions on medical interventions. For this reason, our quantitative estimate of the benefits of breast screening comes from the randomised trials of breast screening. Given the wealth of observational studies on this issue, in section 3.6 we look to observational studies as a possible guide to more contemporary estimates of the effects of screening on mortality.

Randomised controlled trials, however, are not without their problems in practice. Lack of internal validity, for example, through failures in proper randomisation, losses to follow-up and misclassification of end points, can lead to biased estimates of effects. Differences between the trials and the current UK context, for example, in the type of screening undertaken or in the length of follow-up, lead to a lack of external validity. Both the internal and external validity of the RCTs of breast screening have been widely discussed.

A specific issue raised by some commentators is that most of the randomised trials of breast screening date from the 1980s or earlier. Treatment and overall management of breast cancer have improved considerably since that time. Are the trials still relevant? Such a question can be asked of any area of medical investigation and treatment; trials refer to the past and our use of interventions relates to the future. It is an important area of judgement and one that the panel kept at the forefront of its consideration.

The purpose of screening is to prolong survival, but length of survival from diagnosis of breast cancer to death cannot be used as an end point in the RCTs, because the cancers diagnosed by screening are diagnosed earlier than those diagnosed without screening. Thus, even in the absence of any therapy, a cancer diagnosed earlier by screening will have a better survival than the same cancer presenting later symptomatically. Mortality after invitation to screening is the appropriate end point. However, concerns have been raised about the use of breast cancer mortality. If the adjudication of a death as due to breast cancer is influenced by the woman's screening history, then the estimate of the effects on breast cancer mortality can become biased. For this reason, some have argued that death from all cancers, or indeed all-cause mortality, should be the primary outcome of interest in the trials. The panel disagrees with this view (section 3.5). We also comment on the estimation of absolute risk differences, as opposed to RRs, and the difference between the effects expressed per woman invited and per woman screened.

The panel's view is that although the trials are far from perfect, they offer the most reliable evidence on the RR reduction in breast cancer mortality to be derived from screening.

Table 1. Characteristics of the randomised trials of breast cancer screening

	New York HIP	Malmö I and II	Swedish Two County	Canada I and II	Stockholm	Göteborg	UK Age trial	Edinburgh
Start date	1963	1976	1977	1980	1981	1982	1991	1978
Randomisation method	Individual	Individual	Cluster	Individual	Day of birth	Day of birth[a]	Individual	Cluster
Population of women								
Source	IC	P	P	Various[b]	P	P	PC	PC
Number of women[c] (clusters)	62 000	60 076	133 065 (45)	89 835	60 800	52 222	160 921	54 654 (87)
Age group (years)	40–64	45–69 & 43–49	38–75	40–49 & 50–59	39–65	39–59	39–41	45–64
Invited group intervention	M & PE	M	M & SE	M & PE & SE	M	M	M	M & PE
Mammography								
No. of views	2	2 then 1/2	1	2	1	2 then 1	2 then 1	2 then 1
Screening interval	12 months	18–24 months	24–33 months	12 months	24–28 months	18 months	12 months	24 months
No. of screening rounds	4	6–8	2–4	4-5	2	4-5	8–10	2–4
Duration of screening	3 years	12 years	7 years	5 years	4 years	7 years	8 years	6 years
Attendance rate	65%	74%	85%	88%	82%	84%	81%	65%
Control group intervention	None	None	None	PE & SE[d]	None	None	None	None
Follow-up								
Controls invited for screening[e]	Not known	Never[f]	After 7y	Never	After 4y	After 7y	After 10y	After 10y
Cause of death determination	L	IEC, NS	L, IEC, NS	IEC, NS	IES, NS	NS	NS	NS

Abbreviations: HIP = health insurance plan; IC = insurance company register; IEC = independent endpoint committee; L = local; M = mammography; NS = national statistics/register; P = population register; PC = primary care register; PE = physical examination; SE = self-examination. Information taken from various publications, but principally the *Cochrane Review* (Gøtzsche and Nielsen, 2011; Nyström *et al*, 2002; Tábar *et al*, 2011). These summaries are sometimes simplifications of characteristics which differ between sub-trials or subgroups. There are also some discrepancies between different publications.
[a]Day of birth, and later individual.
[b]Includes P, IC, employee recruitment, and general publicity. Women were randomised after initial PE, and there is evidence that the women attending had a higher rate of breast cancer at that initial attendance than was expected from an age-matched population.
[c]Some of these numbers are approximate, as different numbers appear in different publications.
[d]After the initial assessment, only the women in Canada II underwent systematic PE during the screening period—in Canada I, they were taught how to do a physical examination.
[e]Systematic invitation of all controls.
[f]Applies to Malmö I ages 55–69.

3.2 Available randomised trials

Eleven randomised trials have been undertaken and reported (New York health insurance plan (HIP), Malmö I and II, Swedish Two County (Kopparberg and Östergötland), Canada I and II, Stockholm, Göteborg, UK Age trial, and Edinburgh; Table 1). The three trials with two parts have sometimes but not always been reported separately in publications. Three other randomised trials are mentioned in the *Cochrane Review* (Gøtzsche and Nielsen, 2011), but were excluded because they compared multiple interventions (not just mammography), or made major post-randomisation exclusions. We also exclude these three studies from our assessment.

All the trials compared women invited to screening with a control group not invited. However, they varied considerably, for example, in terms of the method of randomisation, age group of women invited, type of mammography employed, whether physical examination or self-examination was also used in either the invited or control groups, interval between screens, number of screens, length of follow-up, and system used for adjudicating breast cancer deaths (Table 1).

Randomisation

The invited and control groups in the trials were constructed either by randomising individuals, or by randomising clusters (geographical areas or general practices), or by allocation according to day of birth. Individual randomisation, with adequate allocation concealment, is rightly regarded as the most reliable method. For population screening studies, however, cluster randomisation can also be adequate, provided sufficient clusters are randomised and balance in social and other characteristics is achieved. Women are identified through existing registers, and so it is unlikely that participation bias, which afflicts some cluster trials (Puffer *et al*, 2003), would apply (for example, through women moving between areas in order to avoid or obtain an invitation to breast screening). Similarly, using allocation by day of birth would seem to be adequate for population screening trials. Of the trials considered, the Edinburgh trial suffered the most problems in terms of its cluster randomisation (Gøtzsche and Nielsen, 2011), with some re-allocations and post-randomisation exclusions of clusters, which led to severe baseline imbalances (26% of women in the control group and 53% in the invited group were in the highest socioeconomic group). For this reason, like the *Cochrane Review*, we exclude the Edinburgh trial from our main summary and comment on its results separately (section 3.5).

Age

The trials recruited women of different ages (Table 1). Most overlapped extensively with the age group 50–70 years, relevant to

the UK programmes, but some (e.g. UK Age trial, Malmö II) did not. We base our primary conclusions about RR on all the trials, as this appears fairly constant across age groups (Nyström *et al*, 2002). There is some evidence, however, that the RR may be attenuated in women under age 50 (Canadian Task Force on Preventive Health Care, 2011), so we also consider an analysis that excludes these women.

Duration of follow-up

Even in the pre-screening era, the median survival from diagnosis of breast cancer was several years, so any benefits of screening in terms of mortality are not immediate, but will accrue over time. So the best evidence would come from a trial with a long duration of follow-up, comparing the invited group with a control group who are never invited to screening. The data that come nearest to this are for the age group 55–69 in Malmö I, with a follow-up of 19 years. Most of the trials, however, started systematic screening of the control group after 4–10 years. Little effect on mortality is seen within the first 5 years of screening, so we regard a follow-up period of about 10–15 years after randomisation as providing the most reliable estimate of the RR. A shorter follow-up time would put too much weight on the early period after initial screening, whereas a longer period would include a greater diluting effect of screening in the control group. So we base our primary conclusions about breast cancer mortality on the data reported in the *Cochrane Review*, which provided results for 13 years of follow-up of the groups as randomised (Gøtzsche and Nielsen, 2011).

Adjudicating cause of death

Potential biases from classifying cause of death have been a major source of contention, especially in the Swedish trials. Ascribing a death as primarily due to breast cancer, or not due to breast cancer, is not always easy or reliable. So, when the screening history of a woman is known, or when a prior diagnosis of breast cancer has been made, this could influence the adjudicated cause of death. There are two ways in which this could distort the results of the trials. The first is overt bias, in which investigators closely involved with the trial adjudicate cause of death and tend to avoid ascribing the cause of death as breast cancer when the woman has been screened (and conversely if they had not). This would exaggerate any beneficial effect of screening. This bias (which may be subconscious) is avoided by the use of an independent end point committee to ascribe causes of death, or by the use of death certificates from national registries. These methods however do not avoid a second way in which a trial's results might be affected; screening increases the number of breast cancers diagnosed, and such a diagnosis may lead preferentially to classifying a subsequent death as due to breast cancer rather than any other cause. This second bias operates against any beneficial effect of screening.

Most trials used an independent end point committee to adjudicate causes of death or took the underlying cause of death from national registries (Table 1). Some of the Swedish trials were criticised for using trial investigators to ascribe cause of death, but subsequent evaluations were made using independent and consensus committees and national registry statistics (Nyström *et al*, 2002; Tábar *et al*, 2011). Although the exact numbers of deaths from breast cancer were not the same when adjudication was made using different methods, the overall estimates of RR of breast cancer mortality did not change very much. Thus, although this issue is certainly one of the major criticisms of the trials, the panel does not think it would exaggerate the estimates of RR reduction obtained from individual trials, or indeed from a

meta-analysis of trials. We comment on the use of other mortality end points in section 3.5.

Other issues

Many other aspects of the trials have been discussed in the literature, some of which we mention here. The numbers of women reported in each randomised group have not been identical across the multiple publications from certain trials. Although this is somewhat concerning, it is perhaps not surprising, given that population and other registers are not always fully reliable, and data checks over time reveal duplicates and other problems. Moreover, some publications are based on birth cohorts and others on exact age groups (Nyström *et al*, 2002). The trials report excluding women with a prior diagnosis of breast cancer. Although this is sensible, it can lead to problems if the exclusions are more easily made in the invited group (for example, because of more information obtained at screening) than in the control group. Some trials include physical examination or self-examination in either or both of the randomised groups. However, there is no evidence that these procedures influence breast cancer mortality (Canadian Task Force on Preventive Health Care, 2011).

Conclusion

We acknowledge the problems and biases discussed above, but judge them as unlikely to have had a major distorting effect on the overall result from a meta-analysis of the trials. Moreover, the biases considered do not all operate in the same direction, with some favouring screening and some acting against it. Although it is easy to be critical of many detailed aspects of the breast screening trials, the relevant judgement is whether the biases are so great as to make their results too misleading for guiding policy. The panel does not believe this to be the case, especially in contrast to the problems in interpreting the results from observational studies (section 3.6).

3.3 Meta-analysis of RRs

As discussed above, we focus on the deaths ascribed to breast cancer in 10 of the 11 randomised trials (excluding Edinburgh) and the meta-analysis conducted in the *Cochrane Review*, using 13 years of follow-up (analysis 1.2 in Gøtzsche and Nielsen, 2011). We do not distinguish the trials labelled 'adequately randomised' and 'sub-optimally randomised' in the *Cochrane Review*, but consider the totality of evidence across all the trials. We also use random-effects rather than fixed-effect meta-analysis to estimate an average effect across the trials. Using random effects acknowledges that the trials may be estimating different quantities, which is likely given their clinical heterogeneity, whereas a fixed-effect analysis estimates an assumed common effect across all the trials. The results are shown in Figure 2 along with the RRs of breast cancer mortality. The overall RR, comparing invited with control women, is 0.80 (95% CI 0.73–0.89). There was some heterogeneity in the RRs from different trials, but this was not statistically significant (Figure 2). Thus, the RR reduction in breast cancer mortality in the groups invited to screening is estimated as 20% (95% CI 11–27%).

The RR for women invited to screening is attenuated compared with that for women who actually attend screening (Cuzick *et al*, 1997). This is because some invited women do not attend, and they may be assumed to get no benefit from the invitation. If the underlying rate of breast cancer mortality in non-attenders is the same as in attenders, one may estimate the RR reduction in attenders as the RR reduction in those invited divided by the
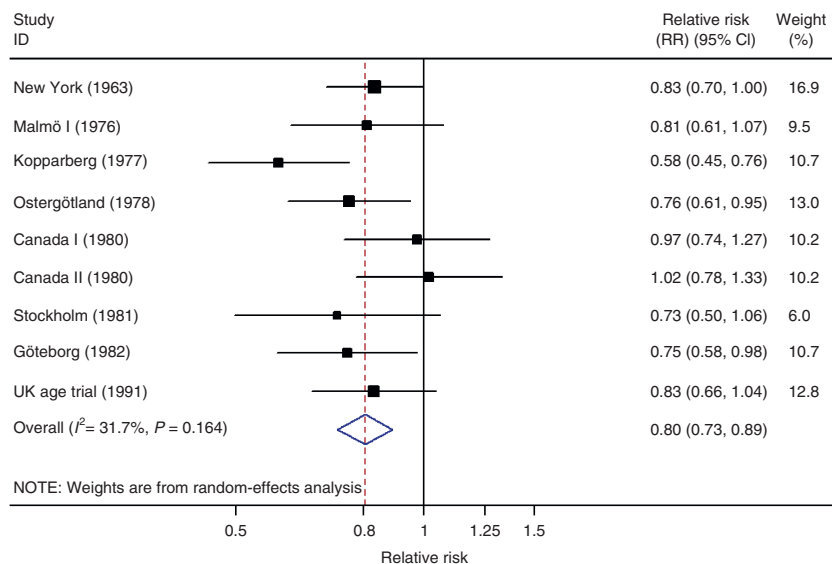
Figure 2. Meta-analysis of the breast cancer screening trials: RR of breast cancer mortality after 13 years of follow-up. Adapted from the *Cochrane Review* (Gøtzsche and Nielsen, 2011). *Note*: Malmö II is excluded because follow-up approximating 13 years was not available; the Swedish Two County (Kopparberg and Östergötland) and Canada I and II trials are split into their component parts; the Edinburgh trial is excluded because of severe imbalances between randomised groups.

(average) attendance rate. Taking the typical attendance in the trials as about 80% (Table 1), this would give 20% divided by 0.80, or 25%). However, this calculation is incorrect as the underlying risk is different in those not attending screening (Zackrisson *et al*, 2004; Moss *et al*, 2006). Without this extra information, which is not available for all trials, the calculation of the RR reduction in those attending screening is not possible. In contrast, the calculation can be made, irrespective of underlying risk differences, for the absolute risk reduction (section 3.4). We note that the coverage rate in the UK NHS screening programme is similar to that in the trials, at 77% (The NHS Information Centre, Public Health Indicators Team, 2011). Some non-systematic (opportunistic) screening occurred in the control groups of the trials, but detailed information is not available. This is ignored in our calculations, and will lead to the effect of attending screening being somewhat underestimated.

Other estimates of overall RR

Other meta-analyses of the breast cancer screening trials have given different estimates of the RR reduction. We summarise some of these below.

- The *Cochrane Review* undertook a fixed-effect meta-analysis of the above trials with 13 years follow-up, and reported an estimated RR of 0.81 (95% CI 0.74–0.87). As expected, the fixed-effect analysis gives a slightly narrower CI, but the estimated average RR reduction of 19% is similar to the figure of 20% above.
- If women < 50 years in the above trials are excluded, the overall RR reported in the *Cochrane Review* (analysis 1.6, Gøtzsche and Nielsen, 2011) is 0.77 (95% CI 0.69–0.86). So the RR reduction is estimated as 23%, slightly more than the 20% above based on all age groups.
- The *Cochrane Review* (Gøtzsche and Nielsen, 2011) focused on the Canada, Malmö, and UK Age trials as the only 'adequately randomised' trials. The estimated RR of breast cancer mortality over 13 years follow-up for invited *vs* control groups in these trials was 0.90 (95% CI 0.79–1.02), whereas in the trials considered 'sub-optimally randomised' it was 0.75 (0.67–0.83). As a compromise between these two estimates, the authors concluded that a 15% RR reduction was plausible.

- The US Task Force (Nelson *et al*, 2009) provided estimated RRs of breast cancer mortality of 0.86 (95% CI 0.75–0.99) for women aged 50–59 years invited to screening, and of 0.68 (95% CI 0.54–0.87) for those aged 60–69 years. These correspond to RR reductions of 14% and 32%, respectively, with an inverse variance weighted average of 19%.
- The Canadian Task Force (Canadian Task Force on Preventive Health Care, 2011) gave an estimate of the RR of breast cancer mortality for invited *vs* control groups of 0.79 (95% CI 0.68–0.90) for women aged 50–69 years, a RR reduction of 21%. Routinely screening for breast cancer with mammography every 2–3 years for this age group was rated as a weak recommendation, based on moderate-quality evidence according to GRADE criteria (Schünemann *et al*, 2011).
- A review by Duffy *et al* (2012) of all the trials and age groups gave an overall RR of 0.79 (95% CI 0.73–0.86) comparing invited with control groups, corresponding to a 21% RR reduction in breast cancer mortality.

Different meta-analyses include different trials, durations of follow-up, and definitions of outcome. Nevertheless, there is general agreement in their estimates, of about a 20% RR reduction in breast cancer mortality from invitation to screening.

Generalisability of RRs

A key issue is whether the RR reduction in breast cancer mortality observed in the trials may be taken as applying, at least approximately, to the current UK screening programmes. This is a judgement about external validity, rather than an issue for which much direct empirical evidence is available. As always in policy decision making, we need to use evidence from studies undertaken in the past to make an inference about what is likely in the future. Although RRs are often much more generalisable across contexts than absolute risk differences, it is clearly plausible that RRs could change in new situations. Of particular concern in breast screening is that many of the trials were undertaken a long time ago, that the techniques of mammography have changed considerably, that DCIS is now commonly diagnosed through screening (section 4.6), that the treatments for breast cancer, particularly the drug treatment that can eradicate microscopic spread, have become

more effective, and that the overall mortality rate from breast cancer has decreased in the United Kingdom and other countries. These points were put to the panel by some expert witnesses. One could therefore argue that breast screening is now less effective/relevant because even later stage cancers can be treated and/or cured, so there is less need to diagnose breast cancers earlier. However, there is a counter argument that because the systemic drug treatments are only partially effective, it could be that the major improvements that drug treatments have brought in cure rates are in fact in part due to breast screening: by diagnosing more cancers at an earlier stage, contemporary drug treatments have a better chance of eradicating microscopic disease, and thus the gains in survival would not have been as great if breast screening did not exist.

Both views have some supporting arguments, but the panel found no convincing evidence that one or other was more likely to be correct. Thus, the panel's view is that the appropriate manner in which to view the benefits of screening and those of better treatments are that these effects are independent, and thus that the estimates of the relative reduction in breast cancer mortality achieved with screening are the same now as 20 years ago. However, the uncertainty about whether there could be an interaction between the benefits of screening and of contemporary treatments is not a reason for stopping breast screening.

Particular aspects for which there is at least some evidence about the external validity of the trials relate to age, screening intensity, and follow-up time. The RR does not appear to change much across the age range 50–69 years (Nyström et al, 2002), but it may be reduced below the age of 50 (Canadian Task Force on Preventive Health Care, 2011). The RR does not appear to depend strongly on the number of screens, or the screening interval, at least across the ranges studied in the trials. The only randomised trial that compared different screening intervals is inconclusive (Breast Screening Frequency Trial Group, 2002). Reports from trials with long follow-up suggest that little benefit in terms of breast cancer mortality is seen in the first 5 years after starting screening, and that the benefit lasts for at least 10 years after cessation of screening. This is not surprising, given the slow progression rates of many breast cancers.

Conclusion

The panel concludes that the current screening programmes in the United Kingdom, which invite women aged 50–70 every 3 years to undergo mammography, are likely to deliver about a 20% reduction in breast cancer mortality at ages 55–79 years. Clearly, there is uncertainty in this figure. In addition to the uncertainty owing to the limited numbers of breast cancer deaths across the trials, there are potential biases in the trials and concerns about the generalisability of results from the trials to the current UK screening programmes. We note, however, that the level of disagreement in the literature about the RR reduction is minor in comparison to the controversy about the absolute risk reduction.

3.4 Absolute risk reduction

The above discussion suggests a natural way to estimate the absolute risk reduction that applies to the current screening programmes in the United Kingdom. For women aged 50 invited to screening, we assume no benefit in breast cancer mortality until age 55, a 20% reduction at ages 55–79, and no change in the rates of other causes of death. An estimated 1.70% of UK women aged 50 are currently expected to die from breast cancer between the ages of 55 and 79; this is calculated from UK mortality rates (2008–2010) and takes into account the risks of dying from other causes. Since the UK programme has existed since the late 1980s, one may assume that this risk has

already been reduced by 20% through screening. Hence, the risk without the screening programme would have been 2.13% (as 1.70/2.13 = 0.80), and the estimated absolute risk reduction is 2.13 − 1.70 = 0.43%.

The number of women needed to be invited for screening for 20 years starting at age 50 in order to prevent one death from breast cancer is therefore 1/0.43% = 235. An alternative way of expressing this is that, for every 10 000 women invited into the screening programme at age 50, about 43 deaths from breast cancer would be prevented.

The absolute risk reduction for women attending screening can be estimated as the absolute risk reduction in those invited divided by the average coverage rate in the NHS breast screening programme (77%), so about 0.43%/0.77 = 0.56%. The number of women needed to be screened for 20 years to prevent one death from breast cancer is then 1/0.56% = 180. For every 10 000 women attending screening from age 50–70 years, about 56 deaths from breast cancer would be prevented.

The above calculations are based on the same principles as those used in some publications (Advisory Committee on Breast Cancer Screening, 2006). Essentially, the RR reduction from the trials is regarded as approximately generalisable to the current UK screening programmes, and the corresponding absolute risk reduction is calculated by applying this RR reduction to the national rates of breast cancer mortality for an appropriate age group. The considerable uncertainty in the estimated RR reduction of 20%, as discussed in section 3.3, of course carries through to these estimates of absolute risk reduction.

The NHS screening programme estimates that 1400 lives are saved per year in England owing to breast screening (Advisory Committee on Breast Cancer Screening, 2006). For comparison and illustrative purposes, the panel estimates that for the 307 000 women (aged 50–52) who each year receive their first invitation to a 20-year screening programme (3-year average 2008/2009–2010/2011, The NHS Information Centre, Public Health Indicators Team, 2011), 0.43% of 307 000, or about 1300, deaths from breast cancer per year are prevented. This is close to the NHS screening programme's estimate.

Different methods and estimates in the literature

The marked difference in estimates of absolute risk reduction proposed in the literature is one of the greatest sources of controversy about the value of breast cancer screening (McPherson, 2010). The different estimates stem from the very varied methods used for their calculation. When calculations are made directly from the trials' data themselves, the absolute risk reduction depends overwhelmingly on the underlying risk of breast cancer, which is principally governed by the age groups considered, the length of follow-up, and the population studied. Although this is obvious, it has also been empirically shown by comparing different durations of follow-up in the Swedish Two County trial (Tábar et al, 2011).

- The Cochrane Review (Gøtzsche and Nielsen, 2011) focused on the Canada, Malmö, and UK Age trials as the only 'adequately randomised' trials. The absolute risk of breast cancer death in the control groups of these trials was low (overall rate of 0.33%), partly because of the inclusion of the large UK Age trial (women initially aged 39–41) and the 13-year follow-up period considered rather than the 25-year period from age 55–79, used above by the panel. With the Cochrane Review authors' estimated 15% RR reduction, this leads to an estimated absolute risk reduction of 0.05%, or equivalently that 2000 women need to be invited to screening to prevent one breast cancer death.
- An entirely different estimate is given by Duffy et al (2010) based on 22 years of follow-up for those aged 50–69 in the Swedish

Two County trial, which estimated a 38% reduction in breast cancer mortality. The calculation considers the absolute risk reduction per women screened across the 7 years of screening in the trial, and makes the strong assumption that the absolute benefits can simply be multiplied up to reflect the 20 years of screening in the UK programmes. This leads to an estimated absolute risk reduction of 0.88% in women screened, or equivalently that 113 women need to be screened to prevent one breast cancer death.

- The US Task Force (Nelson *et al*, 2009) considered a period of 7 years of invitation to screening and 13 years of follow-up after first invitation (Nelson *et al*, 2009). For ages 50–59 years, they estimated that 1339 women needed to be invited to prevent one death from breast cancer. For ages 60–69 years, their corresponding estimate was 377 women.
- The Canadian Task Force (Canadian Task Force on Preventive Health Care, 2011) estimated from the trials that screening 720 women aged 50–69 years once every 2–3 years for about 11 years would prevent one death from breast cancer.
- Beral *et al* (2011) summarised various published estimates of absolute risk reduction from the literature, and concluded that around one breast cancer death would be prevented in the long term for every 400 women aged 50–70 years regularly screened over a 10-year period, based on a previous review (Advisory Committee on Breast Cancer Screening, 2006).

From the above examples, it is clear that different methods of estimation give about a 20-fold difference in the estimates of absolute risk reduction. The panel's view is that to estimate the impact of the UK screening programmes on absolute risk of dying of breast cancer, it is necessary to consider the relevant underlying risk of breast cancer to which the RR reduction from the trials should apply. The panel believes this is best derived from the current UK national rate of breast cancer deaths for women aged 55–79 years. Calculations made directly from the absolute risks observed in the trials are heavily, and often misleadingly, influenced by the age groups included and the length of follow-up available (Beral *et al*, 2011). Estimates also depend on whether they are expressed per woman invited or per woman screened. We note, however, to the extent that the absolute rate of breast cancer mortality in the United Kingdom is currently declining, the absolute risk reduction from the UK screening programme would also be expected to decline correspondingly in the future.

Life expectancy gained

A reduction in the risk of breast cancer will lead to an increase in life expectancy. As breast cancer is only one of many causes of death, the average gain in life expectancy from the UK screening programme is likely to appear modest. An estimate can easily be derived by contrasting the life expectancy for women aged 50, using current national rates of breast cancer mortality and deaths from other causes, to that which would apply if the rates of breast cancer mortality were 25% higher in each year from age 55–79 years. (25% higher corresponds to the assumed 20% benefit from screening, as $1.25 = 1/0.80$.) This calculation leads to an estimate of 0.073 years (or 27 days) of life gained on average for each woman aged 50 invited to screening. To put this in perspective, the panel noted that abolition of all deaths from breast cancer completely would add 159 days on average to life expectancy for women aged 50.

We also note that this is a crude average of a zero gain for the vast majority of women and a substantial gain for a few. Alternative but equivalent ways of expressing this gain are as follows: (a) for the 307 000 women aged 50–52 who are invited for screening each year, about 22 000 years of life will be saved; (b) for each 10 000 women invited to screening, 730 years of life will be saved; (c) for each 10 000 attending screening, about 950 years of life will be saved; (d) given that 1 in about 180 women attending

screening avoid breast cancer death, such a woman would expect to gain on average an extra 17 years of life.

3.5 Other considerations

Edinburgh trial

The Edinburgh trial was the only UK trial in an age group that is within that of the national screening programme. However, as discussed in section 3.2, we excluded this trial because problems in the cluster randomisation led to a severe imbalance in socio-economic status of the women between the groups, and socio-economic status influences, in opposite directions, the risk of developing breast cancer and of dying from breast cancer. At 14 years of follow-up, the unadjusted results showed a 13% reduction in breast cancer mortality. However, on adjusting for socio-economic status, the rate ratio was 0.79 (95% CI 0.60–1.02), a RR reduction of 21% (Alexander *et al*, 1999). Thus, although doubts must remain about the validity of this latter estimate, we note that it very much in line with the figure of 20% we have used above.

Other outcomes

In the preceding sections, we have focused exclusively on breast cancer mortality. Owing to the concerns about whether such deaths are reliably adjudicated in the trials, some authors have suggested that this has led to exaggerated estimates of the RR reduction, and that the outcomes of death from any cancer, or death from any cause, are the appropriate ones for judging the impact of breast screening on mortality. The panel disagrees with this: evaluating all-cancer or all-cause deaths in the trials will lack power because breast cancer deaths represent only a small proportion within these categories. In particular, a 20% RR reduction in breast cancer deaths for ages 55–79 years would yield only 3.0% and 1.2% RR reductions in all-cancer and all-cause deaths, respectively. The trials are not of sufficient size (in terms of numbers of women and length of follow-up) to allow such small RR reductions to be reliably estimated. Hence, a statistically non-significant effect for all-cancer or all-cause deaths in the trials cannot be interpreted as evidence against a reduction in breast cancer deaths.

Some authors have argued that changes in the incidence of more advanced breast cancer, whether defined as above a certain tumour size or with spread to the ipsilateral axillary nodes, is a useful surrogate indicator of the effect of screening on breast cancer mortality in the trials, as the ultimate risk of dying of breast cancer depends in part on the stage of disease at first presentation. Although, on average, one could expect a breast cancer screening programme to lead to diagnosis of breast cancers at an earlier stage, this approach cannot, however, directly exclude lead time effects. The situation is further complicated by the issue of interval cancers, which have been shown in more than one study, as compared with screen-detected cancers, to be more often high grade, which is itself predictive of a poorer prognosis. However, what is less clear is whether the prognosis of a breast cancer is determined only by the stage when diagnosed, or whether in the absence of a screening programme the underlying biology is the main determinant of outcome, and this in turn influences when the cancers present. Thus, for those cancers diagnosed earlier by screening, it is not clear which, if any, of the clinical markers of prognosis (stage, size, grade etc.) is the best predictor of ultimate outcome; or is it some other fundamental characteristic only assessable by molecular biology?

Therefore, there appears to be little reason to use these surrogate outcomes as evidence for or against the benefits of screening, as substantial assumptions are needed to estimate the consequent effect on breast cancer mortality. Only if one wanted to disregard

completely the evidence about breast cancer mortality from the trials, would the use of such surrogate outcomes have value.

There are possibilities of specific harms of screening in terms of induction of other cancers through the X–rays used in mammography or the radiotherapy or drug therapy used to treat breast cancer, and of coronary damage and deaths through radiotherapy (especially of the left breast). These potential harms are discussed in section 5.2.

Statistical and other uncertainties

It is conventional that results from statistical analyses, including meta-analyses, are presented with a measure of statistical uncertainty such as 95% confidence limits. Although these are helpful in giving an impression of the possible influence of the play of chance (given the sample sizes that are available in the studies considered), they fail to represent the uncertainties because of possible biases (from lack of internal validity of the studies) or owing to generalisation from the trials to a new context (external validity). So, the CI given for the RR reduction of breast cancer mortality from a meta-analysis of the trials is an understatement of the uncertainty about the RR reduction that applies to the UK screening programmes. A RR reduction of 20% represents the panel's judgement of the evidence, and should be regarded as an approximate figure rather than a precise estimate.

3.6 Observational studies

In addition to the trials, the panel also considered the value of observational studies in estimating the impact of screening on breast cancer mortality. The RCTs of mammographic screening were conducted at least 20 years ago and most over 30 years ago. Observational studies may help to quantify the effects of screening in an era with major improvements in diagnostic imaging, clinical care, and patient outcomes, as many of the observational studies are more recent than the trials. Both proponents and critics of screening have suggested that the observational studies are more relevant today than the RCTs. However, these studies are beset by many more biases with consequent problems of interpretation. It is also possible that they are more prone to selective reporting than trials, in that the results obtained determine the enthusiasm of the authors and journals for publication.

The biases inherent in observational studies differ by type of study. All share the common problem of potential lack of comparability of screened and unscreened women. It is this feature that the RCTs are designed to address. Each observational study design has strengths and weaknesses and, within each class, specific studies vary in their methods and credibility. The relative merits and problems of the various observational study designs are hotly contested both in the literature and in the evidence the panel heard.

Ecological and time-trend studies

Some observational studies compare time trends for breast cancer mortality in countries or areas before and after the introduction of screening, or concurrently between areas with and without screening. In the first type of study, extrapolation of time trends demands that decisions are made, for example, about the linearity or otherwise of the trend, the choice of time periods considered as 'before' and 'after' screening, and the age groups included. In the second type of study, choices have to be made about the areas to include, the time period considered, and the age groups included. Such decisions, which can appear to have been made rather arbitrarily, can have a profound impact on the estimates obtained. Lack of comparability and different time trends in the

groups being contrasted could lead to substantial bias. For these reasons the panel does not consider that these types of studies provide reliable evidence on the effect of screening on breast cancer mortality, and amongst observational study designs we focus instead on case–control studies and incidence-based mortality studies.

Case–control studies

Case–control studies compare the history of breast screening attendance between women dying of breast cancer and control women who did not die of breast cancer. Case–control studies are prone to a number of potential biases. The main problem with case–control studies is that those attending breast screening are different from those who do not attend. This is referred to as self-selection bias or the 'healthy screened effect'. Attendance is influenced by social and demographic factors that are also likely to be related to the risk of dying from breast cancer, with the resulting bias potentially exaggerating the estimated effect of screening. Also, the existence of a breast screening programme in an area may be associated with better treatment of breast cancer. Therefore, women diagnosed with breast cancer in an area with a breast screening programme may also receive more effective treatment than women where there is no such programme. This would bias the study in favour of screening. Attempts are made to correct for the resulting biases by choice of controls and statistical adjustment (Connor et al, 2000; Duffy et al, 2002).

Some of the expert witnesses who gave evidence to the panel felt that case–control studies provided the most reliable form of observational data while others believed the opposite. The panel undertook a review of the individual characteristics of a number of case–control studies to assess the potential bias of each one (Appendix 3). In general, the studies matched controls to cases by both age and residence but some matched on just one of these variables. Self-selection bias was discussed in around three-quarters of the studies and statistically controlled for, using a variety of methods, in less than half of the studies (Appendix 3).

The case–control studies show more favourable benefit of screening compared with the trials. The panel believes that this is plausibly because of inadequate control for self-selection bias rather than in screening actually being far more beneficial now than in the trials. Attempts to correct for self-selection bias were based on information outside of the study itself (either from a previous time period, or from other geographical areas) that may not be fully relevant. When adjustment was made, the apparent benefit of screening was diminished. The bias that screening could be associated with better treatment was controlled for studies conducted in countries with uniform treatment services.

In conclusion, the panel notes that the beneficial effects of screening are in the same direction as those seen in the trials, but that control for self-selection bias may be inadequate in many of the studies.

Incidence-based mortality studies

Njor et al (2012) conducted a review of European studies on the impact of service mammography screening on breast cancer mortality using incidence-based mortality. In these studies, only breast cancer deaths occurring in women with breast cancer diagnosed after their first invitation to screening are included. They classified the studies according to type of comparison group. These were (1) women not yet invited, (2) historical data from the same region as well as from historical and current data from a region without screening, and (3) historical comparison group combined with data for non-participants.

They found that the effect of screening on breast cancer mortality varied across studies. The RRs were 0.76–0.81 in group 1;

0.75–0.90 in group 2; and 0.52–0.89 in group 3. Study databases overlapped in both Swedish and Finnish studies, adjustment for lead time was not optimal in all studies, and some studies had various other methodological limitations. There was less variability in the RRs after allowing for the methodological shortcomings. On the basis of evidence from the most reliable incidence-based mortality studies, they concluded that the most likely impact of European breast screening programmes was a breast cancer mortality reduction of 26% (95% CI 13–36%) among women invited for screening and followed up for 6–11 years.

Conclusion

Many observational studies have been published, and their conclusions hotly contested. In general, the more contemporaneous case–control and incidence-based mortality studies support the evidence from the trials that screening does have a beneficial effect on mortality. The panel's view is that the trials provide more reliable evidence for an estimate of mortality reduction. Nevertheless, the observational studies support the hypothesis that screening continues to be beneficial in an era of improved treatment.

4. OVERDIAGNOSIS

4.1 Introduction

The purpose of breast screening is to detect cancer early, before it has come to clinical attention. If all cancers would eventually be clinically recognised and treatment was the same and equally effective no matter when the tumour was diagnosed, then screening would be redundant. However, the understanding is that if the cancer is diagnosed earlier, then treatment will be more effective. This is the assumption on which screening is based. The evidence reviewed in section 3 supports that assumption.

As cancers are detected earlier because of screening, we expect the cancer incidence to be higher among screened women during the screening period (the time period between the detection of a cancer at screening and when it would have presented clinically is the '*lead time*' and is an inevitable part of screening). In principle, when screening ceases the incidence should fall back so that by the end of the screening period plus lead time, the cumulative incidence in the screened and control populations should be the same.

Some screen-detected cancers, however, may never progress to become symptomatic (clinically detectable) while some women would die from another cause before the cancer became evident. This adverse consequence (harm) of screening is called overdiagnosis or overdetection. It is variously defined as the 'detection of cancers on screening that would not have been found were it not for the screening test' (IARC, 2002), or 'that would never have clinically surfaced in the absence of screening' (Seigneurin *et al*, 2011) or 'that would not have presented clinically during the woman's lifetime (and therefore would not have been diagnosed in the absence of screening)' (Biesheuvel *et al*, 2007). Thus, it refers to all cancers, invasive or *in situ*.

Underpinning the concept of overdiagnosis is the belief that cancers grow at variable rates, as depicted, for example, in Figure 3A (Esserman *et al*, 2009; Elmore and Fletcher, 2012). Some screen-detected cancers may progress so slowly, that they would never have presented clinically; theoretically, some may be static or even regress but the practical effect is the same. Detection of these cancers turns women into patients, leads to surgery and other treatments that by definition are not beneficial for these women and can cause harm, and adversely affects their quality of life.

As cancers are diagnosed earlier owing to screening, we expect cancer incidence to be higher among screened than unscreened women during the screening period. However, when screening ceases, the incidence should fall back (sometimes referred to as the compensatory drop). If there is no overdiagnosis, the cumulative incidence in the screened and unscreened women will equalise after screening ceases, after a period equivalent to the lead time has elapsed. (Figure 3B, left). If there is overdiagnosis, however, the cumulative incidence will remain higher in the screened group and not equalise over time (Figure 3B, right).

Some overdiagnosis is seen as inevitable – some women will die before their screen-detected cancer would have presented symptomatically. Establishing its frequency is critically important in weighing up the benefits and harms of screening, both for populations and individual women. A big challenge is to get unbiased estimates of the risk. Opinions on the frequency of overdiagnosis range from it being trivial and unimportant to women to being very important and swamping any benefit of screening.

Whether a particular woman has had an overdiagnosed cancer, or whether individual tumours are overdiagnosed, cannot be judged. It is only possible to estimate frequency of overdiagnosis. The issue for the UK screening programmes is the magnitude of overdiagnosis in women who have been in a screening programme from age 50–70, then followed for the rest of their lives. There are no data to answer this question. Any estimate will therefore be, at best, provisional.

4.2 Sources of data on overdiagnosis

Overdiagnosis can be estimated from RCTs or observational studies. Valid estimates depend on similar underlying risks of breast cancer in the screened and unscreened women, and that the effect of lead time has been accounted for (Puliti *et al*, 2012). Overdiagnosed cancers are not all those detected earlier by screening but the subset that would not otherwise have been detected at all.

Randomised controlled trials have the advantage that by design they compare groups of women with the same average prognosis. There are disadvantages of the available RCTs though, including a screening phase that was always shorter than that employed in the NHS national screening programmes, and which varies across the RCTs.

The most reliable estimates of overdiagnosis are from those RCTs in which there was no screening of the control group at the end of the screening period. As screening advances detection of breast cancer, follow-up should extend beyond the screening period to allow a catch up of diagnoses in the unscreened group. In essence, this extended follow-up is needed to distinguish earlier diagnosis from overdiagnosis. If allowance is not made for such catch up, the extra cancers diagnosed in the screened group include some that would also have emerged without screening, albeit later. In principle, the extended period of follow-up should correspond to the lead time, but the average lead time is also the subject of debate, and the lead time is not the same for all cancers. As follow-up is extended well beyond the screening period, new cancers in both the screened and unscreened groups will be included regardless of screening, and the ratio of total numbers of diagnosed cancers will converge towards one (Puliti *et al*, 2012). An ideal follow-up would be to the end of women's lives. However, pragmatically an adequate follow-up is perhaps 5–10
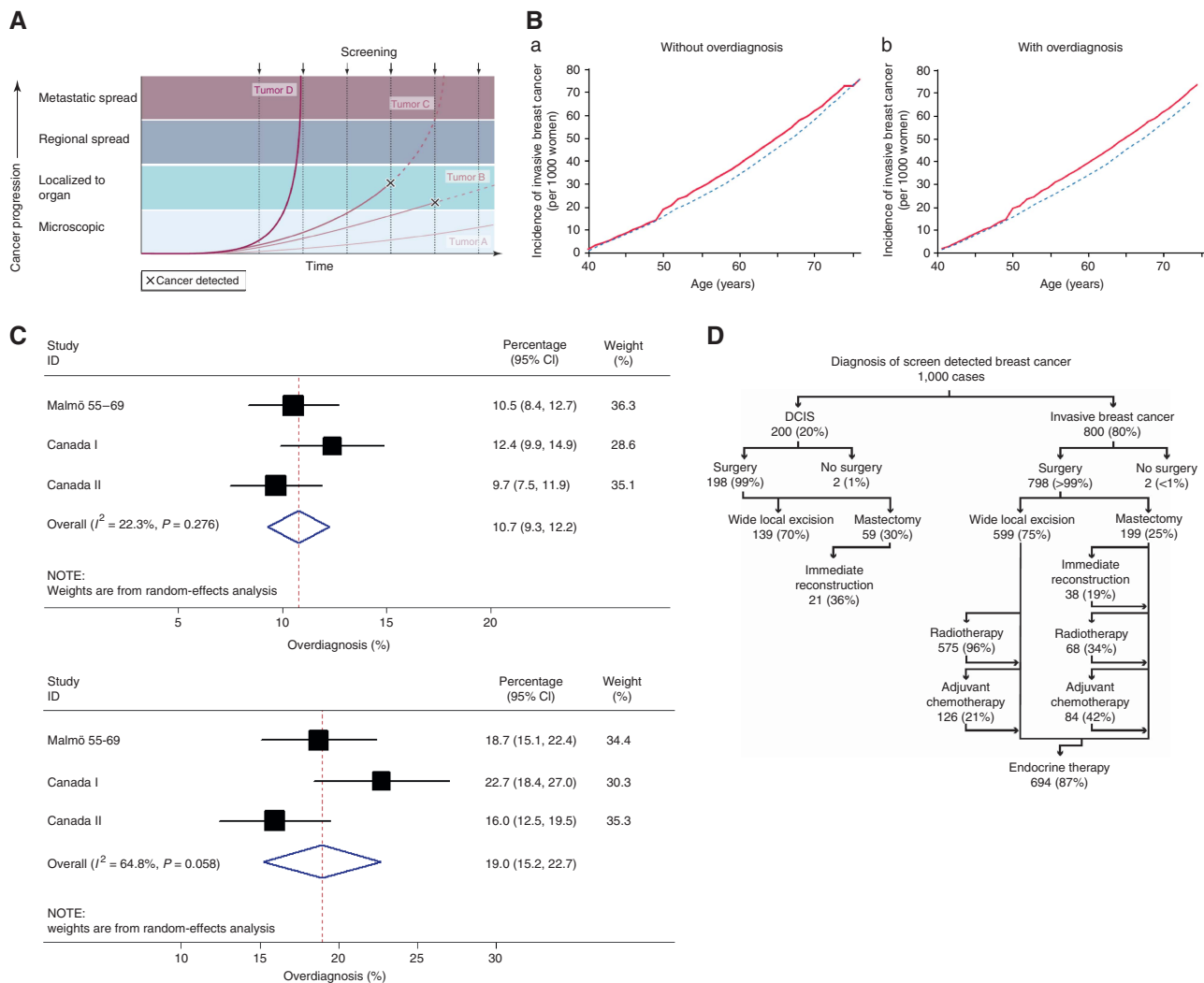
Figure 3. (**A**) Varying screen detection capability in relation to tumour growth rate (copied from Esserman *et al*, 2009). (**B**) Hypothetical cumulative incidence of breast cancer without (left) or with (right) overdiagnosis based on screening women between 50–68 years (copied from Biesheuvel *et al*, 2007; red line shows screened women and blue line unscreened women). (**C**) Meta-analysis of estimates of overdiagnosis: (a) excess cancers as a proportion of cancers diagnosed over whole follow-up period in women invited for screening, (b): excess cancers as a proportion of cancers diagnosed during screening period in women invited for screening. (**D**) Consequences of overdiagnosis.

years after the end of the intervention period (Biesheuvel *et al*, 2007; Puliti *et al*, 2011). The trials that clearly did not invite the control group for screening at the end of the screening phase were the two Canadian trials and the Malmö I trial for women aged 55–69 years (Miller *et al*, 2000, 2002; Zackrisson *et al*, 2006).

In the other RCTs, all the women in the control group were offered screening at the end of the active period of the trial. Estimates of overdiagnosis from these trials are problematic. Screening of women in the control group might itself be expected to lead to some overdiagnosis, and thus to an overall underestimate of overdiagnosis. Exclusion of cancers diagnosed at the end-of-trial screening of the control group would overestimate overdiagnosis, as the control women have not been followed long enough.

Besides the RCTs, there are many non-randomised (observational) studies that have attempted to estimate overdiagnosis. These studies raise many concerns, according to the study design, with the key concern being the likely non-comparability of groups, for example, in different geographical areas. As one contributor to overdiagnosis is the development of other diseases leading to death, the risk of overdiagnosis might be age-dependent. Estimates of overdiagnosis may thus be affected by the age distribution of the

screened group. For non-RCTs it is especially important that age distributions are comparable.

4.3 Estimating overdiagnosis

Overdiagnosis can be estimated by comparing the incidence of breast cancer in cohorts of screened and unscreened women who were followed for several years. Unfortunately, although there is agreement on the concept of overdiagnosis, there has been a wide divergence of views on how to estimate the amount of overdiagnosis, with the result that estimates of the frequency of overdiagnosis vary widely, from $\sim$ 0–50%.

The estimated amount of overdiagnosis depends greatly on the way the calculation is made, and many different methods exist. De Gelder *et al* (2011) (Appendix 4) described seven approaches, all of which have been applied in recent publications. The differences relate to which cases are included in the numerator and, especially, on the choice of denominator. The rate of overdiagnosis can be considered in relation to women invited to be screened, women actually screened, or cancers actually detected by screening. It can also relate to lifetime or the screening age range. It can be expressed as a percentage of the cancers

Table 2A. Estimates of overdiagnosis from the Malmö I trial, with and without inclusion of women age 45–54.

Description	Calculation	Estimated overdiagnosis
All women (age 45–69)		
A. Excess cancers as a proportion of cancers diagnosed over whole follow up period in unscreened women[a]	115/1205	9.5%
B. Excess cancers as a proportion of cancers diagnosed over whole follow up period in women invited for screening	115/1320	8.7%
C. Excess cancers as a proportion of cancers diagnosed during screening period in women invited for screening	115/741	15.5%
D. Excess cancers as a proportion of cancers detected by screening in women invited for screening (64% of all cancers detected in screened group)[b]	115/474	24.3%
Older women (age 55–69)		
A. Excess cancers as a proportion of cancers diagnosed over whole follow up period in unscreened women[a]	82/698	11.7%
B. Excess cancers as a proportion of cancers diagnosed over whole follow up period in women invited for screening	82/780	10.5%
C. Excess cancers as a proportion of cancers diagnosed during screening period in women invited for screening	82/438	18.7%
D. Excess cancers as a proportion of cancers detected by screening in women invited for screening (64% of all cancers detected in screened group)[b]	82/282	29.1%

[a]Result reported by Zackrisson et al (Zackrisson, 2006)
[b]i.e. excluding all interval cancers.

diagnosed in the screening group or as the percentage excess over that seen in the unscreened group. Also, it can be expressed as a relative increase or an absolute increase. Clearly, the different estimates address different questions. Understanding published estimates of overdiagnosis percentages requires identification of exactly how those estimates were derived.

The panel believes that there is no single best way to estimate overdiagnosis. For RCTs, the main options are:

- From the population perspective, the proportion of all cancers diagnosed during the screening period and for the rest of the woman's lifetime in women invited to screening who are over-diagnosed (not including any diagnosed before the age of screening). This probability can be estimated using the difference in cumulative numbers of newly diagnosed breast cancers in groups invited or not invited to be screened, expressed either as a percentage of the number of cancers in the control group (excess risk) or as a percentage of the number of cancers in the screening group (proportional risk). This probability will diminish over time as the number of newly diagnosed cancers increases in both groups.
- From the perspective of a woman invited to be screened, the probability that a cancer diagnosed during the screening period represents overdiagnosis (Welch et al, 2006; Harris et al, 2011). This probability can be estimated using the difference in cumulative numbers of newly diagnosed breast cancers in groups invited or not invited to be screened, expressed as a percentage of the cancers diagnosed during the screening phase of the trial for women in the invited group. The cases in the invited group can also be restricted to those actually detected at a screening visit – that is, excluding interval cancers or cancers among women who did not attend for screening.

These approaches use the same numerator but varying denominators. The panel considers that the appropriate calculations should include DCIS cases, but notes that some studies have reported estimates of overdiagnosis in relation to invasive cancers only.

The panel illustrates how different approaches yield various estimates using data from the Malmö trial (Andersson et al, 1988; Zackrisson et al, 2006), partly following Welch (Welch et al, 2006; Welch and Black, 2010). All cancers, both invasive and non-invasive DCIS, are considered. Also, for transparency, the calculations are expressed in terms of numbers of women whereas some authors have reported rates per 1000 woman years of follow-up.

The Malmö I trial included women aged 45–69 at entry. Cancer incidence was reported after an average of 15 years of follow-up (to December 2001) (Zackrisson et al, 2006). In the active screening period up to 1990, there were 741 cancers diagnosed detected in the screening group and 591 in the control group, an excess of 150. In the period from 1990 to 2001, a further 579 and 614 new cancers were diagnosed, respectively, showing a catching up of 35 cancers. The total numbers of cancers in the screened and control groups were 1320 and 1205, respectively, showing an overall excess of 115 cancers diagnosed among screened women. Zackrisson et al (2006) reported a RR of 1.10 and interpreted these data as showing an estimated overdiagnosis of 10% (95% CI 1–18%). Reporting such a percentage requires consideration of the denominator: 10% of what (Fletcher, 2011)? In fact, the figure of 10% represents the estimated excess risk of a diagnosis of breast cancer among women who had been invited to be screened, and were followed for 15 years after the trial ended. The figure of 10% thus addresses the first key question stated above – population impact.

The panel calculated four estimates of percentage overdiagnosis from the Malmö I trial (Table 2A). The younger women (age 45–54) were offered screening at the end of the study period so the estimates are shown both for all women (age 45–69 at enrolment) and only for women aged 55–69. Different definitions of overdiagnosis lead to estimates ranging from 9 to 29%, although they are based on the same trial.

To answer the second key question – from the perspective of a woman being screened, what is the probability that a cancer diagnosed during the screening period represents overdiagnosis – it is important to include screen-detected cancers and interval cancers. Among women being screened, whether in a trial or a routine screening programme, not all of the diagnosed cancers will be detected at the routine screening; many cancers will be picked up between screens, as 'interval' cancers and might have presented symptomatically in the absence of screening. The relative proportion of interval to screen-detected cancers will increase as the screening interval increases (Breast Screening Frequency Trial Group, 2002) – in general more screen-detected cancers implies fewer interval cancers – so excluding interval cancers will give an estimate of overdiagnosis subject to screening frequency. Further, clinical experience suggests that suspicion of cancer may encourage a woman to accept the invitation to screen. The panel therefore prefers to use, as a denominator for the risk of overdiagnosis among women invited for screening, the second key question, the number of cancers diagnosed in invited women throughout the period of screening.

Table 2B. Randomised trials without screening invitation to the control group (Miller *et al*, 2000; Miller *et al*, 2002; Zackrisson *et al*, 2006)

	A	B	C	D
Malmö I 55–69	11.7% (82/698)	10.5% (82/780)	18.7% (82/438)	29.1% (82/282)
Canada 1	14.1% (82/581)	12.4% (82/663)	22.7% (82/361)	29.4% (82/279)
Canada 2	10.7% (67/626)	9.7% (67/693)	16.0% (67/420)	19.8% (67/338)

4.4 Estimates of overdiagnosis

The literature on overdiagnosis has been reviewed by several authors since 2005. They used different study inclusion criteria, but gave most attention to data from RCTs. Moss (2005) calculated overdiagnosis for eight RCTs as did Gøtzsche (2004) for six of the same trials. Biesheuvel *et al* (2007) reviewed the literature with particular attention given to the RCTs and the two former reviews. Recently, Puliti *et al* (2012) reviewed the European literature covering observational studies. Biesheuvel and Puliti both considered the issue of bias in each of the studies, specifically in relation to adjustment for lead time and case-mix.

Moss (2005) and Gøtzsche (2004) produced very different estimates of overdiagnosis from the same trials. Biesheuvel *et al* (2007) converted all their estimates to a common measure of overdiagnosis (method A described below), but important discrepancies remained. Biesheuvel *et al* (2007) reported that in the studies they considered least biased, overdiagnosis estimates ranged from − 4 to 7.1% for women aged 40–49 years, 1.7 to 54% for women aged 50–59 years, and 7 to 21% for women aged 60–69 years (Biesheuvel *et al*, 2007). Similar large variations have been seen in the estimates of overdiagnosis from observational studies (Puliti *et al*, 2012). Some of the variation seen in these age-specific estimates stems from very small numbers of cases within age groups within trials.

Given the wide variation in both the methods used and the estimates obtained, the panel calculated four estimates of percentage overdiagnosis:

A. Excess cancers as a proportion of cancers diagnosed over whole follow-up period in unscreened women
B. Excess cancers as a proportion of cancers diagnosed over whole follow-up period in women invited for screening
C. Excess cancers as a proportion of cancers diagnosed during screening period in women invited for screening
D. Excess cancers as a proportion of cancers detected at screening in women invited for screening

RCTs without screening of control group at the end of the trial

The most reliable estimates of overdiagnosis come from RCTs in which women in the control group were not offered screening at the end of the trial. Three trials clearly meet this criterion: Malmö I, for women aged 55–69 years, and the two Canadian trials that screened women for 5 years and reported follow-up data at 11 years (i.e., about 6 years after the end of screening; Miller *et al*, 2000, 2002). The estimates of overdiagnosis from these two trials were quite similar to those from Malmö I.

The situation with the HIP study was less clear from the available literature, so the panel excluded this study for the purposes of the estimate of overdiagnosis. In addition, the panel had difficulty from the published literature extracting the data on the numbers of cancer cases in the two arms using the same definition of cases as the other three studies. In particular, the first report of the HIP study included both DCIS and lobular cancer *in situ* (LCIS) in the non-invasive cases (Shapiro, 1977; Shapiro

et al, 1982), but thereafter we could not determine whether LCIS cases had been included in the subsequent incidence data, nor whether non-invasive cases had been included in the process of cross-checking with the New York Cancer registry data and National Death index (Chu *et al*, 1988). Estimates of overdiagnosis from the Malmö I and the two Canadian trials using the four methods already described are shown in Table 2B. The estimates from the three RCTs are quite similar.

Opportunistic screening in the control group would lead to an underestimate of overdiagnosis. In the Malmö and Canadian trials, about 25% (26% and 17%, respectively, in the two Canadian trials) of the women in the control group reported having received a mammogram both during the active trial period and follow-up period. No allowance has been made in the above calculations for that effect.

All four methods use the same numerator, derived from the difference in newly diagnosed cases of breast cancer in the group invited for screening and the control group. Methods A and B differ in whether they compare the excess against the number of cancers diagnosed in the control group or the screening group. Many published estimates use the former (method A).

None of the methods are wrong – they just address different questions. The panel's preferred measures are method B to address the population perspective and method C for the perspective of an individual woman. Figure 3C shows the results from random effects meta-analyses for these two estimates of overdiagnosis.

As many have noted, these three RCTs offer the most reliable evidence for an estimate of overdiagnosis. The combined data suggest a risk of overdiagnosis of about 11% with a population perspective and 19% from the individual woman's perspective.

The panel considers the data consistent with overdiagnosis of about 5–15% from the population perspective and 15–25% from the individual woman's perspective. These estimates are subject to the same sources of uncertainty as noted for the estimates of mortality from the RCTs. In addition, the estimates are not tailored to the UK screening scheme or a 20-year screening period.

In total, these three trials included only 1200 cancers diagnosed during the screening period of which an estimated 243 were overdiagnosed. Given these small numbers, it is important to consider other estimates from other RCTs and the higher quality observational studies. However, those studies clearly provide less reliable estimates.

RCTs with screening of control group at the end of the trial

In several RCTs, all the women in the control group were offered screening at the end of the active phase of the trial. Estimates of overdiagnosis from these trials are problematic. Exclusion of cancers detected at the end-of-trial screening of the control group would overestimate overdiagnosis, as the control women have not been followed long enough. Such an effect is clearly seen in the RCTs without end-of-trial screening. On the other hand, inclusion of cases detected at the end-of-trial screen of women in the control group means that screening is not being compared with no screening. Also, some of the cancers detected by that screen would themselves be overdiagnosed. Thus, including these cancers would lead to an underestimate of overdiagnosis.

Although for several trials both calculations just described are possible, the estimates obtained generally vary widely. For example, for the Stockholm trial using method B, the estimate of overdiagnosis varies from − 2.6% from all diagnosed cancers to + 39% if cancers detected at the end-of-trial screen of the control group are excluded. Although it is reasonable to believe that these two estimates bracket the desired answer (had there been no extra screen and with extended follow-up), the panel believes it is impossible to get useful and reliable estimates of overdiagnosis

Table 2C. Breast cancer incidence rates per 100 000 women in screened and non-screened areas, before and after screening started, and during the last three years of observation (Jørgensen et al, 2009)				
	Screened areas		Non-screened areas	
	1971–1990 No screening	1991–2003 Screening	1971–1990	1991–2003
Age group				
35–49	120	128	111	120
50–69	214	386	198	286
70–79	273	340	264	309

from these trials. An alternative approach is to estimate the effect of lead time and adjust for it. That approach makes very strong, unverifiable assumptions, and the panel is not persuaded that such an adjustment can be made reliably.

Observational studies

Overdiagnosis can be estimated from some non-RCTs, but as always with observational studies there are serious concerns about comparability. Numerous observational studies have adopted a variety of study designs to compare screened and unscreened women or, more often, women who were or were not invited to screening.

There is a considerable body of literature examining the effects of screening in populations and trying to assess the degree of overdiagnosis. Even in the absence of screening, breast cancer incidence rates are not stable over time in populations, and the wide variation in quoted overdiagnosis rates reflects this variation as well as different lengths of follow-up, different statistical assumptions, and different ways of accounting for lead time.

When screening is introduced there will be a short-term rise in the incidence of newly diagnosed cancers. If that rise is solely due to advancing the time when some cancers are diagnosed the increase should fall back to pre-screening levels after some years. A failure to do so may be interpreted as evidence of a degree of overdiagnosis (Esserman et al, 2009). Time trends can also be examined for women of different age groups: before, during, and after the screening programme age range. Such data are shown in Figure 1A in section 2 for breast cancer incidence in the United Kingdom. The increase in incidence associated with the introduction of population screening is clearly seen, first for women aged 50–64 and later for women aged 65–69.

Some studies have compared post-screening incidence with a projection of previous incidence trends in the screened population. Those studies have resulted in very different estimates of overdiagnosis. The panel asked Cancer Research UK to review a set of plausible assumptions made in the literature and to produce estimates based on these assumptions (Jørgensen and Gøtzsche, 2009a; Duffy et al, 2010). The panel found that by changing each of the assumptions, one could get a vast range of estimates of overdiagnosis (Appendix 6). The results of the modelling produced a range of estimates for the impact of the current NHS breast screening programme in England from 0 to >6550 women (aged ≥45) per year in England. Ten per cent of the results were <1150 and ten per cent >4115. As there appears to be no a priori reason to favour one set of assumptions over another, the panel do not think that approaches based on extrapolation offer a robust method to estimate overdiagnosis.

Several groups have compared breast cancer incidence trends over time in screened and unscreened countries or regions over the same time period (Jørgensen and Gøtzsche, 2009). The difficulty with these studies is distinguishing true overdiagnosis from the

excess incidence of breast cancer that results from screening, bringing forward the time of diagnosis. Given that overdiagnosis is defined as a cancer that would not have come to attention in the woman's life span, long follow-up after cessation of screening is essential. The difficulties can be illustrated by studies of comparisons of incidence rates in regions within a single country that did or did not introduce population screening. A study from Denmark is illustrative, as only 20% of the Danish population was offered organised mammography screening over a long time-period (Jørgensen et al, 2009). Screening was introduced in Copenhagen in 1991 and in Funen in 1993 for women aged 50–69. The authors noted that the population in those areas has distributions of age and socioeconomic status comparable with the rest of Denmark.

Table 2C shows the numbers of breast cancers diagnosed per 100 000 women in screened and non-screened areas of Denmark for 20 years before and 13 years after the introduction of screening in 1991. Incidence rates of breast cancer were higher in the screened areas than in the non-screened areas before screening began, suggesting some non-comparability of the areas. During the 13 years of screening, the incidence in women aged 50–69 rose both in the screened areas and the non-screened areas, but more in the screened areas. Incidence also rose in women aged 70–79. One way to estimate overdiagnosis is to compare the ratio of new cancers in screened and unscreened groups in the two periods. In the pre-screening period, the ratio was 1.08 (214/198) and for the screening period it was 1.35 (386/286). The authors say that these data indicate 35% overdiagnosis, but if we adjust for the pre-screening difference the excess is 25% (1.35/1.08 = 1.25). These simple calculations ignore the underlying rise in cancer incidence throughout the period. The authors used regression modelling to take account of incidence trends and age differences, giving an estimate of 33%. As noted earlier, such analyses make additional assumptions that are not verifiable. Studies such as this do not indicate the likely effect of long-term follow-up in reducing the excess in the incidence rate in the screened compared with the unscreened populations.

There have been many other observational studies, but most have the type of problem illustrated here in distinguishing overdiagnosis from the expected increase in breast cancer incidence due to screening and require many assumptions to derive estimates of overdiagnosis. A recent review of 13 observational studies showed overdiagnosis to vary in the range of 0–54%. Adjustment for lead time and breast cancer risk yielded overdiagnosis estimates in the range of 1–10% (Puliti et al, 2012).

The panel's judgement is that the best estimates will come from long-term follow-up of RCTs, as reviewed above.

Statistical and other uncertainties

As noted in section 3, it is conventional that results from statistical analyses, including meta-analyses, are presented with a measure of statistical uncertainty such as 95% confidence limits. Although these are helpful in giving an impression of the possible influence of the play of chance (given the sample sizes that are available in the studies considered), they fail to represent the uncertainties due to possible biases (internal validity of the studies) or to generalisation from the studies to a new context (external validity). So the CIs given for the estimated percentage overdiagnosis are an understatement of the uncertainty about the risk of overdiagnosis associated with the UK screening programmes. Estimates of overdiagnosis have additional uncertainties relating to which estimate to use, and the data are not available for all studies to calculate overdiagnosis in the suggested ways.

Conclusion

The panel believes that overdiagnosis occurs, and that women need to be aware that screening carries a risk of detecting cancers, invasive and

in situ, which would not have troubled them in their lifetime. Tumours that represent overdiagnosis cannot be identified clinically and so will have to be managed according to current clinical protocols.

The panel considers that the data from three of the RCTs without end-of-trial screening of controls provide the most reliable estimates of the extent of overdiagnosis, but notes that there is a rather limited amount of data and numerical estimates are subject to several uncertainties in common with estimates of mortality benefit.

As noted for the estimated benefit for mortality (see section 3.2), the overdiagnosis rates estimated from old RCTs may not reflect those in current screening programmes. There is, however, no clear evidence to suggest that the current rate of overdiagnosis would be lower or higher than in the original trials. The panel thinks that the best estimate of overdiagnosis for a population invited to be screened is of the order of 11%, defined as the percentage excess incidence in the screening population above the long-term expected incidence in the absence of screening.

An alternative definition addresses the answer to the question 'if I am invited to enter into the screening programme and am given a cancer diagnosis during the screening period, what is the likelihood of overdiagnosis'? The panel views the evidence as suggesting that this probability is of the order of 19%.

4.5 Consequences of overdiagnosis

As previously stated, detection of overdiagnosed cancers turns women into patients, leads to surgery and other treatments that are not therapeutically beneficial for these women and can cause harm, and adversely affects their quality of life. As cancers that would not go on to cause cancer death cannot be individually identified, they are treated according to the current treatment protocols. Figure 3D summarises the management of UK screen-detected cancers, both invasive and non-invasive, in 2010/2011 (NHS Breast Screening Programme & Association of Breast Surgery-West Midlands Cancer Intelligence Unit, 2012).

One cannot, however, assume that the overdiagnosed cancers would be managed in the same proportional way as the generality of screen-detected cancers. That the patient dies before the cancer would have presented clinically, implies that such tumours:

- would tend to be more slowly growing, as a more rapidly growing tumour would be more likely to present clinically within a shorter time-frame;
- would be relatively small, as larger tumours would be more likely to present symptomatically.

Thus, overdiagnosed cancers would tend to be more likely to be:

- DCIS (and the relative excess of DCIS in screen-detected cancers would support this), and possibly more likely to be low/intermediate rather than high grade.
- Grade 1 or grade 2 invasive rather than grade 3.

Thus, compared with the diagram, patients with cancers that are overdiagnosed would be:

- relatively more likely to have been treated on the DCIS side than the invasive; and as more likely to be low/intermediate grade, less likely to have had radiotherapy;
- if invasive, more likely to be managed by WLE and radiotherapy than mastectomy as likely to be small
- if an invasive cancer, less likely to have had chemotherapy, as patients having chemotherapy are more likely to have had grade 3 and/or node-positive cancers (NHS Breast Screening Programme & Association of Breast Surgery-West Midlands Cancer Intelligence Unit, 2012);

- if an invasive cancer, more likely to have had endocrine therapy, as oestrogen positivity is associated with older age and lower grade invasive cancers.

Evidence in support of this tendency for overdiagnosed cancers to be of potentially better prognosis, and thus given less aggressive therapy can be seen, for example, in the reports of the nature of cancers found in the two arms of randomised screening trials. Table 2D shows such data for the Malmö I trial.

4.6 Ductal carcinoma *in situ* (DCIS)

There is evidence that breast screening has led to an increase in the identification of DCIS (IARC, 2002). It has been suggested that DCIS is a relatively benign condition that would not cause harm, and therefore diagnosis of DCIS contributes significantly to the magnitude of overdiagnosis.

Definition

DCIS is a malignant process that arises from the epithelial tissues of the breast, and consists of neoplastic cells, which do not, however, infiltrate beyond the limiting basement membrane, and thus remain within the ducts where they arose. Classification is based on the morphological features: architectural growth pattern and the cytological characteristics of the malignant cells. It is usually grouped by grade into high, intermediate, or low grade (IARC, 2002). Along with LCIS, it is classified as non-invasive breast cancer, and although the cells have the appearance of malignancy, they do not show invasiveness, so carcinoma *in situ* is not in itself a life threatening condition. The concern is that at least some have the capacity to progress to invasive malignancy.

DCIS is most commonly detected mammographically as microcalcification. Less commonly, DCIS will present with a symptomatic lump.

Incidence

Table 2E adapted from 'The non-invasive breast cancer report' (National Cancer Intelligence Network, 2011), shows the frequency of non-invasive breast cancer for different age groups and presentations in England for the two years 2006 and 2007.

The majority (about 90%) of non-invasive cancers diagnosed are DCIS. It is apparent that the majority are screen-detected but, nevertheless, 38% were diagnosed symptomatically. Some of the symptomatic tumours may have been detected incidentally when patients presented with a different problem (e.g. microcalcifications found in the contralateral breast when the woman has presented with a benign problem in the one breast). Thus, the detection and management of non-invasive disease is not exclusively a problem of the screening programme. Nevertheless, within the screening age group (age 50–70), the majority (79%) of the DCIS is screen-detected. For 2009–2010, of all screen-detected cancers, about one in five were non-invasive, being a little higher (24%) for the prevalent round and lower (19%) for the incident rounds (The NHS Information Centre, Public Health Indicators Team, 2011). Thus, a mammographic screening programme will detect DCIS. In some cases, (about one in five) (Evans, 2012) investigation of what is radiologically DCIS will lead to the detection of an invasive carcinoma – the larger the area of DCIS, the more likely that there will be a frankly invasive component.

Natural history of DCIS

Before introduction of the screening programme, DCIS was a relatively uncommon tumour. Since it is frequently a marker of associated invasive cancer, it has been investigated and usually

Table 2D. Characteristics of cancers diagnosed in Malmö I trial (Andersson et al, 1988)		
After 5 complete rounds of screening	**Mammography group**	**Control group**
Non-invasive	93 (16%)	53 (11%)
Stage 1	296 (51%)	162 (37%)
Stage 2	142 (25%)	172 (39%)
Stage 3	26 (4%)	27 (6%)
Stage 4	22 (4%)	32 (7%)
Given chemotherapy	26 (5%)	41 (9%)

Table 2E. Frequency of non-invasive breast cancer, England 2006 & 2007 (National Cancer Intelligence Network, 2011)					
	Non-invasive breast cancer			**Total invasive.**	
Age	**Symptomatic (%)**	**Screen detected (%)**	**Total non-invasive**	**No.**	**% non-invasive[a]**
<50	1060 (94)	70 (6)	1130	15 589	7
50–70	1241 (21)	4654 (79)	5895	43 379	12
>70	738 (76)	227 (24)	965	23 930	4
All ages	3039 (38)	4951 (62)	7990	82 899	9

[a]% is non invasive as a % of invasive and non-invasive tumours combined.

excised, and hence it is not possible to know what would have happened if it had been left undisturbed and untreated. Given that the screening programme is diagnosing much more DCIS than presents symptomatically, the relevant questions are:

How common is DCIS?

As above, it represents about 1 in 5 of screen-detected cancers, but only 1 in 20 of all symptomatic cases (National Cancer Intelligence Network, 2011). In reports of small series (IARC, 2002) of women without known breast cancer who underwent postmortems (hospital-based or forensic), invasive cancer was found in about 1% and DCIS in 9%, but there was wide variation in the series, presumably reflecting differences in the women selected and methodologies for examining the breast.

How often does it progress to invasive cancer?

The data from trials of therapy (radiotherapy and/or tamoxifen) after WLE of DCIS shows that both interventions reduce the risk of local relapse (similar to the findings for invasive cancer after WLE). Relevant to the UK screening programme is the UK, Australia, New Zealand (UK/ANZ) trial (Cuzick et al, 2011), in which after WLE of screen-detected DCIS, without any further treatment, relapse in the breast occurred in about 19% of cases, in half of which the relapse was invasive. Progression appears to occur slowly – for example, one series of screen-detected DCIS (Wallis et al, 2012) showed the median time to invasive progression for high-grade DCIS was 76 months, and for low/intermediate grade 131 months.

Is there any way of identifying those cases of DCIS that will or will not progress/relapse as invasive cancer?

DCIS is classified histologically on the basis of excised specimens, and there is currently no certain means of identifying lesions that would not progress. The risk of invasive relapse is higher with high- or intermediate-grade DCIS. Low-grade DCIS seems to pursue a more indolent course, and when invasive relapse occurs it is likely to be a grade-1 tumour. There is ongoing work (Pinder et al, 2010; Reeves et al, 2012) looking at histological and molecular markers to identify those most likely to progress, especially to invasive disease.

Does DCIS affect survival?

The follow-up of patients with DCIS usually shows excellent survival. For example, in the UK/ANZ trial of 1701 women with a median follow-up of 12.7 years, only 179 (11%) had died, of which 39 (2% of all cases) died of breast cancer. Long-term follow-up (NHS Breast Screening Programme & Association of

Breast Surgery-West Midlands Cancer Intelligence Unit, 2012) of 1603 cases of screen-detected non-invasive breast cancer (nearly all DCIS) showed a 20-year relative survival of 97.2% (95% CI 93.6–100.6), with 7.2% of the 493 deaths being due to breast cancer. However, these series are of patients who have had the DCIS treated: what is unclear is what the risk of dying of breast cancer would have been had it been left untreated.

Conclusions

The main question is whether DCIS is a marker of malignancy requiring active treatment or a benign condition of no clinical significance. On the one hand, DCIS (particularly high grade) can certainly serve as a marker for invasive cancer – either because it is associated with the presence of invasive disease at the time of detection, or because its presence indicates an increased risk of invasive disease developing subsequently – in about 10% of cases at 10 years after WLE only. On the other hand, autopsy series and screening programmes both demonstrate that DCIS can be found in the breast of middle-aged women at a greater frequency than presents symptomatically.

Part of the explanation is time. Breast cancer has a long natural history and in patients with invasive cancer, the evolution of metastatic spread and ultimate death may take place over decades. If one also considers the progression of DCIS to invasive cancer as part of this process, the evolution is even longer. In other words, the relevant question is not whether DCIS progresses to invasive cancer (it can), but whether it might have progressed to an invasive cancer that causes symptoms within the lifetime of the women concerned. This will depend mainly on the age of the woman, her life expectancy at the point of diagnosis, and perhaps other factors that could affect progression (hormonal exposure, obesity, etc.). Current series do not show a significant impact of DCIS on survival, after treatment, even at 20 years, but increasing survival may mean that for women in their 50s and even 60s, the diagnosis of DCIS may impact on their long-term survival. Long-term data are needed.

Thus, in diagnosing DCIS via a screening programme, there is a balance to be struck between the potential benefits for some women of identifying and treating a pre-invasive cancer, and the risks for others of treating something that would never have affected the woman in her lifetime. It is not simply the case that DCIS represents overdiagnosis, although it undoubtedly is a contribution to the cases of overdiagnosis.

5. OTHER CONSIDERATIONS

5.1 Introduction

Beside the benefit of breast screening for mortality and its harm in terms of overdiagnosis, the panel considered other relevant issues. These include additional harms through invitation, screening, diagnosis, and treatment, as well as women's perceptions and cost effectiveness. Although the panel has not made a systematic appraisal of evidence in all these areas, being outside its terms of reference (Appendix 1), it has drawn together comments on each of these issues as they should not be neglected when considering the overall impacts of breast screening.

5.2 Harms associated with breast screening

Mammography

Radiation exposure: Mammography uses X-rays and thus exposes women to very low doses of ionising radiation that could cause breast cancers. The actual dose of radiation depends on several factors including the number of views of each breast and whether film or digital mammography is used.

The Health Protection Agency (Health Protection Agency 2001) has suggested that the lifetime additional cancer risk for each mammography examination is between 1 in 1 00 000 and 1 in 10 000.

Although these doses are lower than those for which cancer is directly induced (Preston *et al*, 2002), screening a large population on a regular basis may cause harm. The NHS Breast Screening Programme (2011) in 2006 stated that for every 14 000 women in the age range 50–70 years screened by the NHSBSP three times over a 10-year period, the associated exposure to X-rays will induce about one potentially fatal breast cancer. (NHS Breast Screening Programme & Association of Breast Surgery-West Midlands Cancer Intelligence Unit, 2012). A more recent estimate is that screening women every 3 years from age 47–73 would cause 3–6 cancers per 10 000 women screened (Berrington de Gonzalez, 2011). This risk is incorporated in estimates of the benefit of screening (see section 3). Digital mammography, which uses a lower radiation dose, is increasingly being used in the English screening programme. Therefore, it is likely that the risk of exposure will be reduced.

Pain: During the process of mammography, the breast is compressed and flattened in order to create a uniform density, which improves the image and reduces the radiation dose. A substantial proportion of women find this painful and some studies (Nelson *et al*, 2009; Gøtzsche and Nielsen, 2011) have shown that the pain and discomfort of mammography deters them from attending for further screening (Gøtzsche and Nielsen, 2011).

The assessment process

Figure 4 summarises the process and numbers for women recalled after routine screening mammograms Figure 4.

Many women take part in the screening programme; it is often argued that for many the benefit will be reassurance (Welch *et al*, 2011). With that reassurance, however, must come the knowledge that all screening tests have errors of false positives and false negatives. The mammogram may sometimes appear to show an abnormality that requires further investigation to determine whether or not it is a cancer-requiring treatment or fail to detect a cancer that is present.

False-positive mammogram

In Figure 4, 2522 women (i.e., 3105 recalled minus the 583 diagnosed with cancer = 2522: 3.36% of all the women screened) were
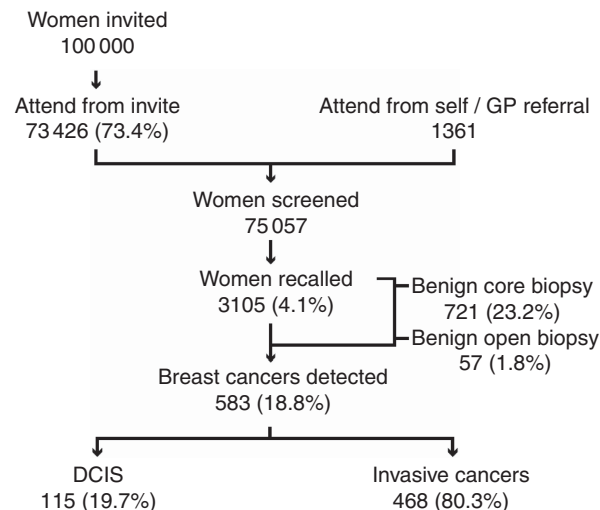

Figure 4. The assessment process. Overall cancer detection rate is 583/75 057 = 7.8 cancers per 1000 women screened. Of the 3105 women recalled, 583 (18.7%) will be diagnosed with invasive or *in situ* cancer. Data extracted for women aged 50–70 for year 2009/10 (NHS Breast Screening Programme, 2011; NHS Breast Screening Programme & Association of Breast Surgery-West Midlands Cancer Intelligence Unit, 2012).

recalled and found not to have cancer. This is called a false-positive result. Of the women recalled and found not to have cancer, the majority (1744/2522 = 69%) had only further imaging (mammography, ultrasound) but a minority (778/2522 = 31%) had a biopsy, which was core biopsy under local anaesthetic in all except 2.3% (57/2522) who had a formal biopsy under general anaesthetic. The latter group represents only 0.076% (57/75 057) of all women screened.

Numerous studies have assessed the psychological impact of a false-positive result on women (Brett *et al*, 1998; Brett and Austoker, 2001). The studies' results are conflicting but a recent systematic review of the literature (Bond *et al*, 2012) concluded that, in the population at general risk of breast cancer, a false-positive result can cause breast cancer-specific psychological distress, which may endure for up to 3 years. The degree of distress is associated with the level of invasiveness of subsequent assessment. Some studies found that the distress caused by a false-positive result deterred some women from re-attending for breast screening, which would reduce any benefit they would otherwise have got from being offered screening in the first place. The level of distress can be mitigated by providing women with clearly worded information about the recall and appropriate support from clinical staff in before and during assessment (Bond *et al*, 2012).

False-negative results

No screening test is completely accurate and sometimes mammography will not detect a cancer. This may because the cancer is not mammographically visible or develops between screening rounds and women are warned of this possibility in the screening literature. When women present with an interval cancer, the previous mammograms are reviewed blind to assess whether a suspicious abnormality was visible on the previous screening

mammogram. If so, such cases are classified as a true false-negative mammogram, that is, the suspicious abnormality was missed at the first screen. For women attending at three yearly intervals, the false-negative rate is 0.2/1000 women screened (Lawrence, 2012; c.f. the cancer detection rate by screening of 7.8 cancers/1000 women screened).

Diagnostic testing

Core biopsy carries a risk of local haemorrhage and, rarely, reaction to local anaesthetics. Open surgical biopsy involves a general anaesthetic but it is regarded as a low-risk procedure.

Psychological consequences of a positive diagnosis

The psychological consequences of a breast cancer diagnosis and subsequent treatment have been well documented. In terms of harms of screening, these consequences are particularly relevant to those women who have been overdiagnosed. Although these women will not know that the cancer would not have caused them any harm they will have suffered unnecessary psychological trauma associated with a cancer diagnosis

Two studies (Yousaf et al, 2005; Schairer et al, 2006) have shown a small but significant increased risk of suicide in patients diagnosed with breast cancer. The risk increases with advancing stage of the disease and therefore may be less relevant for those who are overdiagnosed. However, two further studies (Jamison et al, 1978; de Leo et al, 1991) have found suicidal ideation to be present in some patients post-mastectomy. Although these risks are small they should not be overlooked when assessing the benefits and harms of breast screening.

Potential morbidity and mortality from treatment

Breast surgery: As with any surgical procedure, there are hazards from the anaesthetic and the surgical procedure itself. Although the surgery can be extensive (especially if it involves reconstructive surgery as well), the surgery is elective, patients are assessed pre-operatively, serious complications are rare. The most extensive surgery is mastectomy and reconstruction for which the mortality is estimated to be <0.3% (The NHS Information Centre, Public Health Indicators Team, 2011). In contrast, following mastectomy, 10% of patients will have some sort of complication (e.g. infection, fluid accumulation) (The NHS Information Centre, Public Health Indicators Team, 2011).

Radiotherapy: Acutely, radiotherapy can cause skin reactions and uncommonly radiation pneumonitis. Both of these are short-lived and usually not severe.

Radiotherapy can cause other long-term harms (Early Breast Cancer Trialists' Collaborative Group (EBCTCG), 2005). There is, at 15 years, a small excess risk of non-breast cancer mortality (15.9 vs 14.6%, an absolute difference of 1.3%). This is mainly due to heart disease (so seen more in left- than right-sided cases because more of the heart is irradiated), lung, and oesophageal cancers. These estimates are derived from trials of radiotherapy performed mostly during or before the 1970s; since then radiotherapy techniques have changed especially with the introduction of CT planning, so reducing the volume of heart and lung irradiated, which should reduce, but not eliminate, such complications. Data from the Surveillance Epidemiology and End Results (SEER) database (Giordano et al, 2005) shows that the risk of death from ischaemic heart disease due to radiotherapy has diminished from 1973 to 1989 (risk from right-sided tumours unchanged, left sided decreased).

The last published Oxford overview (Clarke et al, 2005) showed that there is a reduction in mortality from the reduction in local recurrence of invasive cancer by radiotherapy. Essentially, for every four recurrences prevented at 5 years, there will be one death prevented at 15 years. For illustration, the local recurrence rate in the radiotherapy START trial (in which many patients had screen-detected cancers) was 3.5% at 5 years, which, given radiotherapy reduces local recurrence by about two-thirds, would correspond to a 5-year local recurrence rate of about 10.5% without radiotherapy. This gain of 7% in local control should correspond to a reduction in mortality of just under 2%.

Adjuvant hormone therapy: The most extensive experience is with tamoxifen. Trials of adjuvant tamoxifen for 5 years have shown that for patients with hormone receptor-positive breast cancer, breast cancer mortality is reduced by about 33% (Early Breast Cancer Trialists' Collaborative Group (EBCTCG) 2005), translating into an absolute reduction in mortality at 10 years of 5.3% and 12.2% for node-negative and node-positive patients, respectively. Tamoxifen does have some long-term hazards in that it carries an increased risk of uterine cancer and thromboembolic disease. Their effect on mortality is of the order of 0.2% per decade and is outweighed by the modest but positive effect of tamoxifen on ischaemic heart disease (possibly because it lowers cholesterol) (Dewar et al, 1992). Aromatase inhibitors are increasingly used instead of tamoxifen, but their overall effect on mortality is very similar to that of tamoxifen.

Cytotoxic chemotherapy: Adjuvant cytotoxic chemotherapy reduces both overall and breast cancer-specific mortality. Use of an anthracycline- or taxane-containing regime yields a RR reduction of about one third in breast cancer mortality (Peto et al, 2012). The absolute benefit depends on the risk profile but will often be of the order of 6–7% at 10 years. There are acute toxicities associated with giving chemotherapy — such as alopecia, nausea and vomiting, which are all unpleasant but non-fatal. Acute neutropenic sepsis can be fatal but this is a rare event in the adjuvant setting. There is an increased risk of thromboembolism. Mortality rates during adjuvant chemotherapy have been reported at around 0.3% (Cameron et al, 2003). The main long-term risks are (Azim et al, 2011):

- *Cardiac*: Anthracyclines can cause a cardiomyopathy, the incidence being dose related and increasing with age. Trials suggest an absolute excess mortality of up to 1%, but this may be an underestimate as the incidence of cardiac failure may be higher and can occur many years after treatment.
- *Second cancers*: The main risk with chemotherapy, particularly anthracycline-based, appears to be acute myeloid leukaemia and myelodysplastic syndrome. At standard doses, the risk is probably of the order of 0.5% but may be higher if the doses (especially of alkylating agents and anthracyclines) are increased.
- *Neurotoxicity and premature menopause*: Both are very real causes of morbidity but not of mortality.

Conclusion

We know that within the NHS screening programmes, of patients found to have invasive or non-invasive cancer, 99% have surgery (of whom 5.7% have mastectomy and immediate reconstruction), 72% have radiotherapy, 72% have adjuvant hormone therapy, and 27% adjuvant chemotherapy (NHS Breast Screening Programme & Association of Breast Surgery-West Midlands Cancer Intelligence Unit, 2012). From the above, assuming a worst case scenario, it would be reasonable to assume no adverse mortality effect for hormone therapy, no net effect of radiotherapy on mortality, a maximum of 0.2 per 1000 dying because of surgery (0.3% of those having reconstruction) and 1.3 per 1000 dying because of chemotherapy (0.5% of the 27% who

have chemotherapy), giving an adverse mortality rate of 0.15%. For patients who have an 'overdiagnosed' cancer, the risk is likely to be lower as it is unlikely that they would have received chemotherapy (see section 4).

- The panel concludes that the excess mortality from the investigation and treatment of invasive breast cancer is small and outweighed by the benefits of the treatment.

For DCIS, the benefits of radiotherapy or hormone therapy are in terms of recurrence rather than a reduction in mortality, but the absolute risks of such treatment in terms of mortality are likely to be very small. For patients with screen-detected breast cancer, there is no evidence that these risks are any greater than in the symptomatic population, but for women diagnosed with a breast cancer, that if it were certain would never be symptomatic, there is nevertheless a real, but very small, mortality risk from being screened.

5.3 Women's perceptions of screening

The development of new information to accompany cancer screening invitations was not in scope for this review and is being dealt with separately. Women's perspectives on overdiagnosis and whether they see it as a key issue in their screening decisions had not previously been investigated, so Cancer Research UK commissioned some qualitative research to investigate this. The findings, from one focus group attended by panel members, are presented briefly here for information (Appendix 5), but academic papers, focusing on a larger sample of qualitative research, will follow publication of this report.

These women understood the concept of screening and most had attended. Although they understood breast cancer, and many knew people who had had it, they had little concept of DCIS and overdiagnosis. Their opinions are not mainly informed by the screening leaflet, and it would appear many do not read it in detail. Thus, informing women about screening will involve much more than simply re-writing the leaflet.

5.4 Cost-effectiveness of breast screening

It was not in the panel's remit to review the data relating to the costs or the cost-effectiveness of breast cancer screening. The

Department of Health in England has provided funds of about £100 million per year to deliver the current screening programme (NHS Breast Screening Programme, 2012).

If one were to take the well-founded cost-effectiveness approach such as that employed by the National Institute for Health and Clinical Excellence (NICE) when reviewing a health technology, it would be important to establish the costs not only of the intervention, but of all subsequent interventions, both in those invited to be screened and those not offered screening. No such data are available for any of the randomised trials, and thus this panel is not in a position to consider the full costs of a breast screening programme, including the financial costs to the NHS of any overdiagnosed cancers.

Thus, although it has been estimated that the UK NHSBSP comes within the NICE cost/quality-adjusted life year threshold of £20 000–30 000 (Advisory Committee on Breast Cancer Screening, 2006), the panel is not able to comment on this, as it has not been able to scrutinise the costs of treatment with and without screening, including the costs of treating the cancers that are overdiagnosed.

We can, however, make general comparisons with other interventions and see that, in terms of lives saved per year, breast cancer screening is of a similar order of magnitude as cervical screening, bowel cancer screening using faecal occult blood testing and, the use of statins (Table 3).

Table 3. Comparison of public health interventions

Intervention	Number lives saved per year
Breast screening for women aged 50–69	1300
Cervical screening (with liquid based cytology)	4500 (Peto et al, 2004)
Bowel screening with biennial FOBT	1800–2400 (Parkin et al, 2008)
Use of statins	9700 (Department of Health, 2006)

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Recommendations for further research

The panel's review of the randomised trials of breast screening leads to the following recommendations about future research priorities:

- An individual participant data meta-analysis of the breast screening trials is in progress. This should help resolve some (but not all) of the concerns that have been raised about individual trials and their combined interpretation. The panel supports this enterprise, and is disappointed that it had already not been done a long time ago.
- The impact of breast screening outside the ages 50–69 years is very uncertain. The panel supports the principle of the ongoing trial in the United Kingdom for randomising women under age 50 and above age 70 to be invited for breast screening.

The panel's review of overdiagnosis leads to their support for further research into DCIS, in particular:

- A proposed study to examine the need for treatment of low-grade DCIS

- Continued support for the Sloane project, which has an extensive database of screen-detected cases of DCIS, and the long-term follow-up of these cases may well improve our understanding of this condition (The Sloane Project 2010, 2011).
- Current mammographic screening techniques now detect many more cases of DCIS than in the trials. The appropriate treatment of these is uncertain, because there is limited information on their natural history (section 4.6). The panel supports studies to elucidate the appropriate treatment of screen-detected DCIS.
- Work on improved screening and pathological techniques that can predict prognosis more effectively.

The panel also supports:

- A re-evaluation of the cost-effectiveness of the NHS breast cancer screening programme that takes into account the conclusion of this report.

6.2 Conclusions

Breast screening extends lives. The panel's review of the evidence on benefit – the older RCTs, and those more recent observational studies judged to be relevant – point to a 20% reduction in mortality in women invited to screening. A great deal of uncertainty surrounds this estimate but it represents the panel's overview of the evidence. This corresponds to one breast cancer death averted for every 235 women invited to screening, and one death averted for every 180 women who attend screening.

The breast screening programmes in the United Kingdom, inviting women aged 50–70 every 3 years, probably prevent about 1300 breast cancer deaths a year, equivalent to about 22 000 years of life being saved; a most welcome benefit to women and to the public health.

But there is a cost to women's well-being. In addition to extending lives by early detection and treatment, mammographic screening detects cancers, proven to be cancers by pathological testing, that would not have come to clinical attention in the woman's life were it not for screening - called overdiagnosis. The consequence of overdiagnosis is that women have their cancer treated by surgery, and in many cases radiotherapy and medication, but neither the woman nor her doctor can know whether this particular cancer would be one that would have become apparent without screening and could possibly lead to death, or one that would have remained undetected for the rest of the woman's life.

The answer the panel sought was to the question of the level of overdiagnosis in women screened for 20 years and followed to the end of their lives. Estimates abound of overdiagnosis, from near to zero to 50%, but there are no reliable data to answer this question. There has not even been agreement on how to measure it. On the basis of follow-up of three RCTs, the panel estimated that in women invited to screening, about 11% of the cancers diagnosed in their lifetime constitute overdiagnosis, and about 19% of the cancers diagnosed during the period that women are actually in the screening programme. However, the panel emphasises, these figures are the best estimates from a paucity of reliable data. Any excess mortality stemming from investigation and treatment of breast cancer is considered by the panel to be minimal and considerably outweighed by the benefits of treatment.

Putting together benefit and overdiagnosis from the above figures, the panel estimates that for 10 000 UK women invited to screening from age 50 for 20 years, about 681 cancers will be found of which 129 will represent overdiagnosis, and 43 deaths from breast cancer will be prevented. In round terms, therefore, for each breast cancer death prevented about three overdiagnosed cases will be identified and treated. Of the ∼307 000 women aged 50–52 who are invited to screening each year, just over 1% would have an overdiagnosed cancer during the next 20 years. Given the uncertainties around the estimates, the figures quoted give a spurious impression of accuracy.

6.3 Policy recommendations

The panel concludes that the UK breast screening programmes confer significant benefit and should continue. The greater the proportion of women who accept the invitation to be screened, the greater is the benefit to population health in terms of reduction in mortality from breast cancer. However, for each woman the choice is clear: on the plus side, screening confers reduction in the risk of mortality from breast cancer because of early detection and treatment. On the negative side, is the knowledge that she has perhaps a 1% chance of having a cancer diagnosed and treated that would never have caused problems had she not been screened.

Evidence from a focus group the panel conducted, and in line with previous similar studies, was that screening was an offer many women will feel is worth accepting: the treatment of overdiagnosed cancer may cause suffering and anxiety but that suffering is worth the gain from the potential reduction in breast cancer mortality. Clear communication of these harms and benefits to women is of utmost importance and goes to the heart of how a modern health system should function. There is a body of knowledge on how women want information presented, and this should inform the design of information to the public.

Terms of Reference for the breast screening review

The overall aim of the review is to develop an up-to-date (2012) assessment of both the benefits and harms associated with population breast screening programmes. This is a rigorous review of the evidence by an independent panel; it is not a formal systematic review.

The review has been commissioned by Professor Sir Mike Richards, National Cancer Director, England, and Dr Harpal Kumar, Chief Executive of Cancer Research UK.

Up to six independent experts will be appointed to undertake the review. These experts will be nationally and internationally recognised for their expertise in epidemiology and/or medical statistics as well as in current breast cancer diagnosis and treatment, but will not have previously published on the topic of breast screening.

The reviewers will be supported in their work by a small team based at Cancer Research UK who will assist with the collation of relevant research papers and will facilitate the work of the panel. Additional funding for the review will be provided from the Department of Health.

Scope

The reviewers will be asked to consider both the evidence from RCTs of breast screening and from observational studies, including prospective follow-up and case–control studies, of the impact of breast screening programmes both in the United Kingodm and elsewhere.

Evidence for the review will be limited to research that has been published or accepted for publication.

The most recent publications from each source (e.g., RCT or cohort) should be considered, but the reviewers may also choose to consider earlier publications from the same source.

In addition to considering individual studies, the reviewers will be asked to consider published systematic reviews and/or meta-analyses of breast screening in different countries/jurisdictions. The reviewers will also be asked to consider published methodology papers to assess benefits and harms in breast screening studies.

The reviewers will be expected to understand the arguments that have been made in various articles and opinion pieces regarding the benefits and harms of breast screening. The focus, however, will be on the evidence, thus articles providing opinions (either for or

against breast screening) that do not contain original data or meta-analyses will not be expected to have a bearing on the conclusions of the review.

Outputs

The key outputs of the review will be:

- An estimate of the likely benefits of breast screening and the range of uncertainty in this estimate.
- An estimate of the likely harms of breast screening and in particular the risks of overdiagnosis and the range of uncertainty in this estimate (i.e., patients being diagnosed and treated for cancer that would not have caused problems during their lifetime).

The review will comment on overall effectiveness of screening which, in addition to the above, will depend on participation rates and developments in effective treatment.

If the available evidence permits, assessments of benefits and harms should also be made for different age groups and for different subgroups (e.g., DCIS diagnoses, and socioeconomic, and ethnic groups).

Approach to the review

It will be for the independent reviewers to determine how they wish to conduct the review. In addition to reviewing the published evidence, it is likely they will choose to consult a range of experts in breast screening who have published in this field. These experts could include epidemiologists, statisticians or clinicians. Consultation may be via written communications, interviews, or workshops.

Regular updates on the process of the review will be made available by Cancer Research UK (through a dedicated page on the Cancer Research UK website (http://www.cruk.org.uk/breastscreeningreview).

Reporting

The reviewers will be asked to prepare a report for publication by Cancer Research UK. The outputs of the report may also be published in peer-reviewed journals. On completion, the report will be shared with the UK National Screening Committee (NSC) and with Ministers in England, but the NSC will have no input to the content of the report.

It is expected that the initial report will be published by spring/summer 2012. However, it is conceivable that, at that stage, the over-riding view of the independent review group is that further primary research is needed to develop definitive conclusions. If this is the case, the review will make recommendations regarding the balance of evidence as it currently stands.

It is further expected that the review will make recommendations on key messages regarding risks and uncertainty that need to be considered when drafting new communications materials regarding the breast screening programme. This will include considerations of effectiveness.

Review process

As outlined in section 2, the panel called on a range of experts to give evidence. The expert witnesses who have presented evidence to the panel and debated points relevant to the review are:

Philippe Autier, Vice President, Population Research, International Prevention Research Institute (iPRI), Lyon, France

Michael Baum, current Director of the Clinical Trials Group at University College London,Professor Emeritus of surgery and visiting Professor of medical humanities, University College London

Dame Valerie Beral, Professor of Epidemiology and Director, Cancer Epidemiology Unit, University of Oxford

Susan Bewley, Consultant Obstetrician and Honorary Senior Lecturer at King's College London

Stephen Duffy, Professor of Cancer Screening, Wolfson Institute of Preventative Medicine, at Barts and the London School of Medicine and Dentistry, part of Queen Mary University London, UK.

Harry de Koning, Professor of Screening Evaluation, Erasmus MC, Rotterdam, the Netherlands

Ian Ellis, Professor of Cancer Pathology, University of Nottingham

Peter Gøtzsche, Director, Nordic Cochrane Centre, Copenhagen, Denmark

Klim McPherson, Emeritus Fellow, Visiting Professor of Public Health Epidemiology, Oxford University

Albert Mulley, Director, The Dartmouth Centre for Health Care Delivery Science and Professor of Medicine, Dartmouth Medical School, Dartmouth, USA

Lennarth Nyström, Associate Professor, Department of Public Health and Clinical Medicine, Umea University, Sweden

Julietta Patnick, Director, NHS Cancer Screening Programmes and Visiting Professor, University of Oxford

Sir Richard Peto, Professor of Medical Statistics & Epidemiology, Co-director of the Clinical Trial Service Unit, University of Oxford

Paul Pharoah, Professor of Cancer Epidemiology, University of Cambridge

Sir Nick Wald, Institute Director, Wolfson Institute of Preventive Medicine, Barts and the London Medical School

Jane Wardle, Professor in Clinical Psychology and Director, Health Behaviour Unit, University College London

Robin Wilson, Consultant Radiologist, The Royal Marsden, London

These expert witnesses also suggested additional scientific evidence for consideration by the panel and provided follow-up information on their evidence, if requested by the panel. The secretariat organised and attended each witness session but did not participate in any discussions. The commissioners of the independent review, Professor Sir Mike Richards and Dr Harpal Kumar attended some of these sessions but only as observers; they did not participate in any discussions or pose any questions to either the panel or the expert witnesses.

Role of the secretariat

Cancer Research UK and Department of Health provided the secretariat, acting purely as support to the panel in the practical, writing, and dissemination functions, and having no say in the conclusions or recommendations.

In addition, the secretariat collated a bibliography of all scientific research papers and reports that had been brought to the panel's attention by experts from both sides of the screening debate. The secretariat also provided additional specific research papers that the panel wished to consider. In addition to providing the modelling study, showing the impact the various assumptions used to calculate the level of 'overdiagnosis' can have on these estimates (Appendix 6), Nick Ormiston-Smith provided cancer incidence, mortality, and survival statistics and ran statistical analyses as requested and instructed by panel members. The secretariat also organised a focus group with women of screening age in collaboration with the Cancer Research UK Health Behaviour Research Centre at University College London, as requested by the panel.

APPENDIX 2. CHANGES IN BREAST CANCER MANAGEMENT AND MORTALITY

Since the late 1980s, there have been three main changes in breast cancer management:

- Treatment
- Organisation of services
- Population screening

Treatment

Surgery: There has been a shift from mastectomy to breast conservation (lumpectomy and radiotherapy), formal staging of the axilla, latterly by sentinel node biopsy.

Radiotherapy: Trials have established the role of radiotherapy, following lumpectomy and, for selected patients, following mastectomy.

Adjuvant systemic therapy: Trials have established that for patients with oestrogen receptor (ER)-positive invasive breast cancer, tamoxifen (or for postmenopausal patients, aromatase inhibitors) reduce the risk of relapse and improve long-term survival. Adjuvant chemotherapy was initially introduced for high-risk premenopausal patients, using the CMF regime, then, as its benefits were appreciated, postmenopausal and lower-risk patients, were also treated and anthracycline- and/or taxane-containing regimens were also used (with further benefit). More recently, for the minority of women with HER2-positive breast cancer treated with chemotherapy, trials have confirmed that the addition of trastuzumab further improves survival.

Organisation of services

The management of breast cancer in the United Kingdom was considered part of general surgery, pathology, radiology, and oncology. There has been shift (in part due to the setting up of specialist screening services) to all breast cancer patients being seen in specialist units and decisions about management being considered at specialist multidisciplinary team meetings. This was an incremental process, and the improvements in, for example, surgical staging have helped better targeting of treatment (e.g., knowledge of the nodal status assists selecting patients for chemotherapy, knowing the ER status selects patients for adjuvant hormone therapy).

Screening

Most of the issues about screening are discussed elsewhere in this report.

Changes in mortality

The graph (Figure A2) shows the changes in breast cancer mortality for different age groups over a 40-year period.

The increases in mortality seen in the early period (1971 to mid 1980s) presumably follow on from the increase in incidence. The pattern by age follows that of the incidence changes with the peak incidence in older patients and little change in the under 50s. This is what one would expect if there was little change in treatment (as was the case).

The fall in breast cancer mortality starts in the late 1980s, affects all age groups, and has continued to 2007. The figures in the table below examine these changes in more detail (GROS 2010, ONS 2011, NISRA 2011). Deaths from breast cancer can be expressed as standardised mortality rates, which are very useful for comparing populations, but also as absolute numbers of deaths, reflecting the realities in terms of the burden on the health service.

Table A2. Changes in breast cancer mortality and the contribution of breast cancer as a cause of death for the years 1998–90 and 2008–10, by age groups

	Age groups (years)				
	40–49	50–59	60–69	70–79	80–84
Average numbers of breast cancer deaths/year					
1988/90	1415	2590	3516	3720	1662
2008/10	1000	1681	2354	2655	1498
Reduction (%)	415 (29)	909 (35)	1162 (33)	1065 (29)	164 (10)
Breast cancer mortality					
1988/90	39.0	81.5	115.1	151.2	214.0
2008/10	21.1	43.2	68.9	107.1	167.3
Reduction (%)	17.9 (46)	38.3 (47)	46.2 (40)	44.1 (29)	46.7 (22)
No. of deaths (female)/year					
1988/90	6712	15 947	45 567	89 289	64 641
2008/10	6659	13 612	28 347	57 884	50 557
Reduction (%)	53 (1)	2335 (15)	17 220 (38)	31 405 (35)	14 084 (21.8)
Breast cancer deaths as % of all deaths					
1988/90	21.1	16.2	7.7	4.2	2.6
2008/10	15.0	12.3	8.3	4.6	3.0
Change	− 6.1	− 3.9	+ 0.6	+ 0.4	+ 0.4

GROS, 2010; ONS, 2011; NISRA, 2011

The figures confirm a reduction in both the breast cancer mortality and the absolute number of deaths from breast cancer in all age groups. It should, however, be noted that:

- the reduction in breast cancer mortality is most marked in the under 70s.
- breast cancer mortality rises markedly with age, but the relative contribution of breast cancer deaths to the total number of
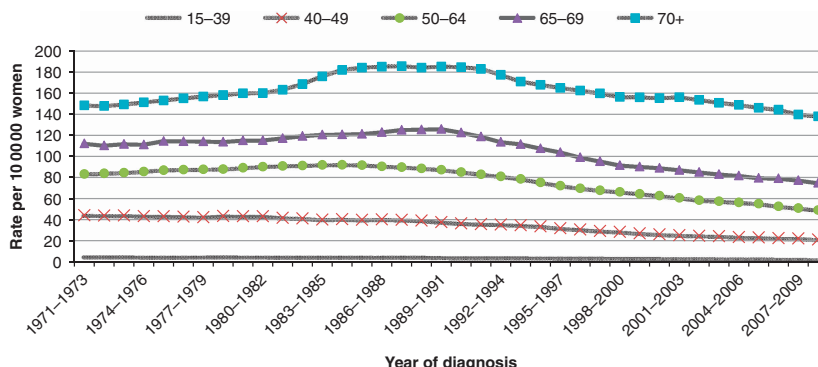

Figure A2. European age-standardised mortality rates, breast cancer, women, by age, UK 1971–2010. (Cancer Research UK, 2012).

deaths falls with age (reflecting the increase in other causes of death with age).

- accompanying the fall in breast cancer mortality and deaths, there has been a marked fall in other causes of death, particularly in the 60–79 age group.
- the net effect is that the relative contribution of breast cancer to total deaths has fallen in those <59 years but has (modestly) risen in those aged ⩾60.

The screening programme would be expected to impact only on the deaths of women ⩾55 years, and it is apparent from the figures that the overall effect on mortality will be attenuated by the impact of other causes of death. (e.g., in the 70–79 age group, breast cancer accounts for <5% of all female deaths).

It should also be noted that breast cancer mortality rates vary by country. Breast cancer mortality in the United Kingdom in the late 1980s was about 40 (per 100 000 population), whereas in Sweden (where many of the screening trials were carried out) and Norway it was about 26. The undoubted improvements in the United Kingdom mortality have only brought it down to the starting level

in Sweden/Norway (where it was down to 22 by 2006). Comparisons of mortality need to include both absolute levels and changes and understanding of the reasons for baseline differences.

Conclusion
Many of the early trials of adjuvant therapy occurred concurrently with the screening trials, and so the changes mentioned above were introduced concurrently with the introduction of screening. It is thus difficult to disaggregate the individual contributions of each to the undoubted improvements in breast cancer mortality. The organisational changes are not the product of RCTs and are multifaceted, so their contribution, although probably real (Kesson *et al*, 2012), is the most difficult to quantify. There has been a significant improvement in breast cancer mortality in all ages, part of which is certainly due to improvements in treatment. Changes in mortality also reflect factors affecting incidence as well as presentation and organisational arrangements. Crude mortality statistics are the summation if these factors but do not of themselves indicate the relative contributions.

APPENDIX 3. CASE–CONTROL STUDIES

Case–control study selection
General medical literature was searched using PubMed for the period 1970 to present in order to identify case–control studies that assessed the effect of screening mammography on breast cancer mortality. The following search terms were used in locating the articles, 'breast cancer', 'screening', 'screening mammography', 'breast cancer mortality', 'breast cancer death', 'screening case–control' and 'screening case referent'. A total of 21 case–control studies (see Table A3) were identified as considering breast cancer mortality and compared against a list of

case–control references shared by expert witnesses. The case–control studies showed breast screening to confer a greater benefit than did the trials. Although these studies, in general, attempted to control for non-comparability of screened and unscreened women, the panel was concerned that residual bias could inflate the estimate of benefit. There were also a larger number of case–control studies investigating comparisons between lifestyle risk factors and detection of abnormalities at screening, and early- and late-stage cancers. These latter studies were not used, as they did not directly provide an estimate of screening benefit.

Table A3. Overview of the characteristics of the case–control studies considered and their methods for correcting for self-selection bias

Study/reference year	Age group studied	Controls	Self-selection adjustment	Unadjusted odds ratio of breast cancer mortality in women attending screening *vs* those who did not (95% CI) age groups not reported where combined figure available	Adjusted odds ratio of breast cancer mortality in women attending screening *vs* those that did not (95% CI)
Gullberg *et al*, 1991	Not reported	Matched by age	Adjustment considered but not made	0.42 (0.22–0.78)	N/A
Palli *et al*, 1989	40–69 at diagnosis	5 Matched by year of birth and residence	Adjustment has been attempted using proxy variables	Age 40–49: 0.63 (0.24–1.64) Age 50–69: 0.51 (0.29–0.89)	Age 40–49: 1.46 (0.36–5.93) Age 50–69: 0.35 (0.14–0.85)[a]
Puliti *et al*, 2008	Residents aged 50–74	4 Matched by date of birth and residence	Adjustment using women in areas where the breast screening was not underway.	0.46 (0.38–0.56)	0.55 (0.36–0.85)[b]
Collette *et al*, 1984	50–64 at study entry	3 Matched by year of birth and residence	Adjustment considered but not made	0.30 (0.13–0.70)	N/A
Miltenburg *et al*, 1998	50–69 at first invitation	3 Matched by year of birth, invitation and residence	Adjustment considered but not made	0.54 (0.37–0.79)	N/A
Verbeek *et al*, 1984	35–65 at first invitation	5 Matched by year of birth	Adjustment considered but not made	0.48 (0.23–1.00)	N/A
Verbeek *et al*, 1985	35–65 at first invitation	5 Matched by year of birth	Adjustment considered but not made	0.51 (0.26–0.99)	N/A
Van Dijck *et al*, 1996	65–92 at last invitation	5 Matched by age and invitation	Adjustment considered but not made	0.56 (0.28–1.13)	N/A
Broeders *et al*, 2002	35+ at first invitation	5 Matched by age, invitation and residence	Adjustment considered but not made	Age 40–49: 0.84 (0.30–2.29) Age 50–59: 0.65 (0.30–1.42) Age 60–69: 0.65 (0.31–1.28) Age 70–79: 0.70 (0.32–1.54)	N/A
Van Schoor *et al*, 2010	40–69 at invitation			Age 40–49: 0.50 (0.30–0.82) Age 50–59: 0.54 (0.35–0.85) Age 60–69: 0.65 (0.38–1.13)	N/A

Table A3. (Continued)

Study/reference year	Age group studied	Controls	Self-selection adjustment	Unadjusted odds ratio of breast cancer mortality in women attending screening vs those who did not (95% CI) age groups not reported where combined figure available	Adjusted odds ratio of breast cancer mortality in women attending screening vs those that did not (95% CI)
Van Schoor et al, 2011	50–69 at invitation	5 Matched by residence	Adjustment using the Duffy method	—	0.65 (0.49–0.87)
Paap et al, 2010	50–75 at invitation	1 Matched by year of birth and residence	Adjustment using the Duffy method	0.30 (0.14–0.63)	0.24 (0.10–0.58)[b]
Otto et al, 2012	49–75 at first invitation	1 Matched by invitation and year of birth	Adjustment using women in same area before breast screening.	0.45 (0.37–0.54)	0.51 (0.40–0.66)[b]
Friedman and Dubin, 1991	40–64 at entry	4 Matched by survival and year of birth	Adjustment considered but not made	0.75 (0.60–0.93)	N/A
Norman et al, 2007	40–64 at entry	Matched within original random control group	Adjustment considered but not made	Age 40–49: 0.89 (0.65–1.23) Age 50–64: 0.47 (0.35–0.63)	N/A
Elmore et al, 2005	40–65 at diagnosis	1 Matched by age, risk and health plan	Adjustment considered but not made	0.92 (0.79–1.08)[c]	0.92 (0.79–1.08)[c]
Moss et al, 1992	45–64 at entry	5 Matched on age at entry to trial and residence	Adjustment considered but not made	0.51 (0.27–0.98)[d]	N/A
Fielder et al, 2004	50–74 at diagnosis	5 Matched on year of birth and GP Practise	Adjustment using the Duffy method	0.49 (0.36–0.66)	0.75 (0.49–1.14)[b]
Allgood et al, 2008	50–70 at diagnosis	2 Matched on year of birth	Adjustment using the Duffy method	0.35 (0.24–0.50)	0.52 (0.32–0.84)[b]
Roder et al, 2008	45–80 at death of cases	3 Matched on date of birth	Adjustment has been attempted using socio-economic status		0.59 (0.47–0.74)
Gabe et al, 2007	⩾40 at death	3/4 Matched by date of birth and residence	Adjustment using the Duffy method	0.59 (0.41–0.84)	0.65 (0.39–1.09)[b]

Abbreviations: CI = confidence interval; IMPACT, DOM, CARE = Women's Contraceptive and Reproductive Experiences Study; HIP = health insurance plan; OR = odds ratio; TEDBC = trial of early detection of breast cancer. Odds ratios for age groups have been presented where an overall odds ratio is not available.
[a]At least 2 examinations and last screen <30 months.
[b]Corrected for self-selection bias.
[c]Combined risk groups, screening by mammogram.
[d]Outcome of the analysis conducted solely in the screening district.

APPENDIX 4. FORMULAE USED TO CALCULATE OVERDIAGNOSIS

Formulae for calculating overdiagnosis from de Gelder et al (2011):

$(E − D)/T_{0, \text{age } 0–100 \text{ years}}$: the relative increase in breast cancers due to overdiagnosis $(E − D)$ compared with the predicted number of breast cancers in the female population aged 0–100 years in a situation without screening.

$(E − D)/T_{0, \text{screening age and older}}$: the relative increase in breast cancers due to overdiagnosis $(E − D)$ compared with the predicted number of breast cancers in women of the screening age and older in a situation without screening.

$(E − D)/T_{0, \text{screening age}}$: the relative increase in breast cancers due to overdiagnosis compared with the predicted number of breast cancers in women of the screening age in a situation without screening.

$(E − D)/T_{1, \text{screening age}}$: the fraction of overdiagnosed cancers of all diagnosed breast cancers in women of the screening age in a situation with screening.

$(E − D)/SD$: the fraction of all screen-detected (SD) cancers that is overdiagnosed.

$T_{1, \text{screening age}}/T_{0, \text{screening age}}$: the RR of breast cancer for women of the screening age in a situation with screening compared with the predicted number of breast cancers in women of the same age in a situation without screening. The estimator can be corrected for lead time for instance, by shifting the predicted incidence without screening forward in time.

$T_{1, \text{screening age}}/(T_{1, \text{screening age, corrected}})$: the RR of breast cancer for women of the screening age in a situation with screening compared with the predicted number of tumors in a situation with screening if no overdiagnosis would take place ($T_{1, \text{screening age, corrected}}$).

Abbreviations:

D: number of deficit breast cancers in the age groups exceeding the screening limit, calculated as the difference in the number of breast cancers without and with screening; DCIS: ductal cacinoma in situ; E: number of excess breast cancers in the screening ages, calculated as the difference in the number of breast cancers with and without screening; SD: number of screen-detected cancers; T_0: predicted number of breast cancers in the absence of screening; T_1: modelled total number of breast cancers in the presence of screening; $T_{1, \text{corr}}$: total number of breast cancers in the presence of screening minus the number of overdiagnosed cancers.

APPENDIX 5. FOCUS GROUP

Overview

Nine women from the London area and within the breast screening age range (50–71 years) were invited to join a focus group to discuss their reasons for attending screening or not, and to comment on information on the risk of overdiagnosis and DCIS. All women spoke fluent English, had no previous personal history of cancer and came from a range of socioeconomic backgrounds. These women were recruited from a market research recruitment database, hosted by Saros (http://www.saros-research-recruitment.com). Saros screened eligible participants via email or phone. The group was facilitated by Dr Jo Waller from the Cancer Research UK Health Behaviour Research Centre at University College London and observed by two panel members.

Findings

Accepting or declining an invitation to screen: Seven out of the nine attendees had accepted an invitation to be screened in the past, which is close to the current UK average. The main reasons expressed for attending were an assumed feeling that attending screening is beneficial (the perceived benefits are that finding the disease earlier means better outcomes and that a negative scan provides peace of mind), awareness of breast cancer ('lots of people are getting it') and the fact that you receive a specific invitation to attend. The main reason mentioned for not attending was the anticipated pain of the actual mammogram and the embarrassment of the technique. Some of the women who had previously attended screening had experienced discomfort and mentioned that this might deter them from accepting another invitation in the future. There was a general consensus among the group that no screening programme can be perfect, and that some cancers may be missed but the women had not been particularly aware of DCIS and overdiagnosis. The screening attendees felt that the screening programme was well organised, but most agreed that they would be less likely to attend if not specifically invited.

The knowledge that you may be diagnosed with and treated for a slow growing tumour that would never have caused you problems in your lifetime did not appear to change this group's intention to accept another screening invitation. There was a general consensus, in this group and others, that attending screening and possible subsequent decisions on treatment if cancer or DCIS are found were two separate issues.

The women were surprised, however, to learn that doctors cannot always tell whether a tumour is likely to cause harm or not, but felt that the treatment decision was one to be made by the woman after discussions with their consultant. There was a feeling that doctors would not recommend treatment for cancer if they did not think it was appropriate. This is in line with findings from other qualitative research (Dr Jo Waller, personal communication).

There was more concern about the potential radiation risk: 'For every 14 000 women screened regularly for 10 years, one woman may develop breast cancer she will die from because of the radiation from the mammograms' (NHS Breast Screening leaflet, 2011) and the high number of women recalled for further tests after the initial mammogram (about one in every 20 screened), rather than the concept of overdiagnosis.

Information about screening: Some of the women in this focus group expected the information in the breast screening leaflet to boost uptake of screening invitations, and felt that it should therefore be written in non-alarmist terms. Many could not recall whether they had read the leaflet when they were last invited and nobody could remember what information it actually contained. The women indicated that their decision to accept or decline an invitation to screen was unlikely to be influenced by information in this leaflet. This group also expected the leaflet to focus on what to expect when attending for a screen, notably the procedure of the mammogram. However, they also felt that some basic information about risks and benefits should be included for those women who wanted it.

Table A5. Focus group characteristics

Ethnicity	No.	Current employment	No.
Black Caribbean	2	Employed full time	1
White British	7	Employed part time	3
Other	0	Unemployed	0
Marital status		Self-employed	1
Single/never married	2	Full-time homemaker	0
Married/living with partner	1	Retired	3
Divorced	5	Still studying	0
Widowed	1	Disabled or too to work	0
		No answer	1
Highest level of education qualification obtained		Does your household own a car or van?	
Degree or higher degree	5	No	2
Higher education qualification below degree level	1	Yes, one	5
A-levels or highers	0	Yes, two or more	1
ONC/BTEC	0	No answer	1
O level or GCSE equivalent (Grade A–C)	2	Have you been screened for breast cancer? (i.e., had a mammogram)	
O level or GCSE (Grade D–G)	1	Yes	6
No formal qualifications	0	No	2
Current living arrangements		When was your last mammogram? (Open text)	
Home owned outright	3	No answer	3
Home owned with mortgage	3	Cannot remember	1
Rent from local authority/housing association	2	2007	2
rent privately	0	2010	3
Other (e.g., living with family/friends)	0	Have you ever been called back for further tests after your mammogram?	
No answer	1	Yes	2
		No	6
		No answer	1

APPENDIX 6. MODELLING OVERDIAGNOSIS USING TIME TRENDS

Introduction

The most reliable estimates of overdiagnosis come from three RCTs in which women in the control group were not offered screening at the end of the trial. However, these randomised trials of breast screening date from the 1980s or earlier. The setting of the trials is not necessarily directly comparable with the current screening programme (for example, because of different technology used in mammography and changes in the underlying breast cancer risk). Many researchers have therefore attempted to use observational studies to estimate the extent of overdiagnosis contemporaneously.

One method of estimating the level of overdiagnosis is the extrapolation method (ONS, 2012). This method predicts the expected level of breast cancer if there were to be no mammography screening, and compares it with the actual observed level.

Estimates produced using this basic extrapolation method differ for a number of reasons, but one reason for different estimates is the choice of assumptions used in the modelling (ONS, 2012). A number of models were therefore run to consider the effect of applying different assumptions and the impact this has on the estimate of overdiagnosis due to breast screening.

Methods

The method predicts the expected level of breast cancer diagnosis in the age group targeted for screening in the absence of screening using a regression model, and then calculates the difference between this expected level and the observed data; the excess due to screening (Figure A6.1). Similar analysis for the older-age group was then undertaken to calculate the size of a compensatory drop. Compensatory drop is the relative decrease in the incidence of a cancer in a screened population compared with an unscreened population, once screening stops. This is because screening detects the cancer earlier, so cases that would have presented symptomatically have already been diagnosed during screening. The overall estimate of overdiagnosis is therefore the number of excess cases due to screening minus the size of this compensatory drop. This was calculated for each year in the analysis and the average taken so that the results from all the models with different periods could be compared.

This extrapolation method assumes that the risk of breast cancer has increased at a constant rate, as the period used to estimate the expected level of breast cancer ends. In addition, it assumes that the quality of case ascertainment by registries and diagnostic methods has remained stable over time.

Models

In total, 2250 regression models, using both linear and Poisson regression, were applied to the age-specific incidence rates in England from 1975 to 2004 using a combination of different assumptions, and a range of different overdiagnosis estimates were produced. The assumptions examined were:

- the pre-screening era period, this varied from 1975–1984 to 1975–1988 (and all intermediate years).
- the target-screening age group, this included age groups 50–64, 45–64 and Poisson regression models, which considered two age categories separately 45–49 and 50–64
- the post-screening age group, this included age groups 65–74, all women ⩾65 and Poisson regression models, which considered two age categories 65–69 and all women ⩾70.
- the screening era period, the start of the screening period was allowed to vary between 1989 and 1993 and the end from 2002 to 2004.

Table A6.1 set outs the different model specifications used in the modelling.

Table A6.1. The different assumptions in the published models

| Model spec	Assumptions
Regression technique	Linear regression on rates and Poisson regression on counts
Pre-screening era year end	1984 through to 1988
Screening age group	50–64, 45–49 and 50–64, and 45–64
Post-screening age group	65–74, 65–69 and 70 and over, and 65 plus
Screening era	All combinations starting with 1989–1993 and ending in 2002–2004
Compensatory drop adjustment	Rate ratio and difference of counts
Risk adjustment	None and for increasing trend in under 45s
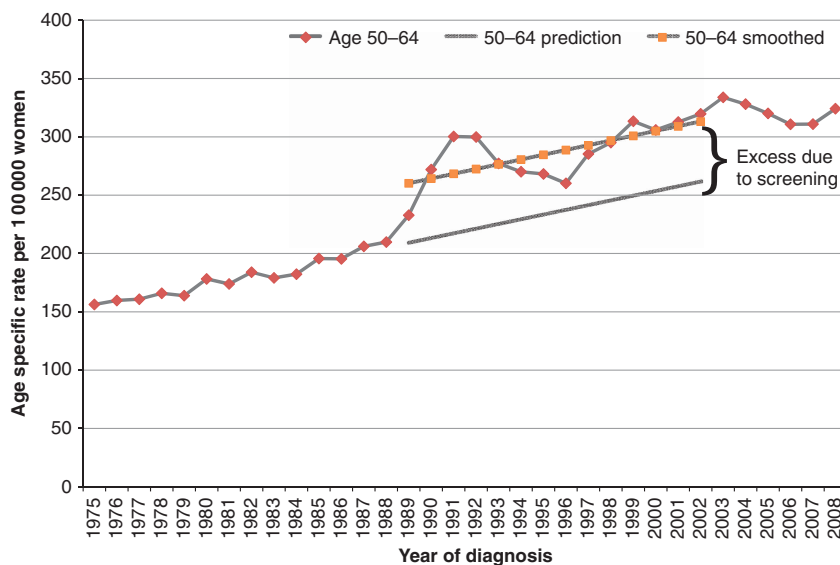
Figure A6.1. Breast cancer age-specific incidence rates, England 1975–2008. Description to include: breast cancer age-specific incidence rates, England 1975–2008, age 50–64 with expected, observed and smoothed data.

Standard Models: All of the combinations of age groups and time periods, described above, were modelled using both linear regression (300 models) and Poisson regression (675 models). These analyses are known here as the standard regression models.

Calculation of the Compensatory Drop: In addition to the standard models described above, the calculation of the compensatory drop (using either count or rate ratio) was investigated. The results of the standard regression models have been expressed using the difference of counts in the observed and expected data. To investigate the effect of calculating the compensatory drop using a different method the outcome of the linear regression standard models using the difference of counts were compared with the same models, but the compensatory drop was calculated using the rate ratio of the observed and expected data. This analysis was repeated but the rate ratio was only applied to the final year in each set of linear regression results.

Model Adjustment: Finally, an adjustment was applied to the standard Poisson regression model estimates to take account of increasing incidence in the mostly unscreened under 45 age group. These results were compared with the standard Poisson regression model estimates.

Results

The results of the 2250 regression models produced a range of estimates from 0 to 6552 or 214 per 10 000 women invited to screening (Figure A6.2). A total of 10% of the results were <1150 (or 37 per 10 000 women invited to screening) and 10% above 4115 (or 134 per 10 000 women invited to screening). These figures are not directly comparable with the 129 per 10 000 women invited to screening in the main report because they do not include DCIS. Three assumptions had the biggest effect on the results of the modelling: the adjustments made to the regression technique, end of pre-screening era, and target age group.

Table A6.2. Number of additional cases of overdiagnosed per year when pre-screening era ends using the standard linear and Poisson regression models

Year when pre-screening era ends	Number of additional overdiagnosed cases per year compared with 1988: linear regression models	Number of additional overdiagnosed cases per year compared with 1988: Poisson regression models
1984	1083	1314
1985	664	749
1986	468	565
1987	191	208
1988	0	0

Standard Models: Choosing a standard Poisson regression model rather than a standard linear regression model increases the estimates of over diagnosis by an average of 316 women per year (range 80–476). The choice of target age group 50–64 rather than the age group 45–64 reduced the estimates of over diagnosis by an average of 379 across all of the regression models. Increasing the length of the pre-screening era from 1984 to 1988 reduced the estimates of overdiagnosis by an average of 1083, using the linear regression models and 1314 in the Poisson regression models. The length of the pre-screening era, which varied from 1975 to 1984 or 1988 (and all years in between) has an impact because there has been an increase in the rate of breast cancer diagnosis in this 5-year period.

The effects of the length of the screening era and the age groups used in the post-target age group had much less of an effect. Starting the screening era at 1991, on average, had the biggest effect and increased the over diagnosis estimate by 123 women per year compared with 1993 (the lowest year) in the standard linear regression model. In the standard Poisson regression model, 1991 is the highest year that was 158 higher than 1993 (the lowest estimate), on average, in the standard Poisson regression model.

Calculation of the compensatory drop: When the rate ratio was applied rather than the difference of counts to calculate the compensatory drop on a set of standard linear regression estimates, the estimates of overdiagnosis were increased by, on average, 295 women per year. If only the last year of the screening period was used rather than the average across all years, then this method increased the estimate overdiagnosis by 1469 women per year on average. This difference was primarily driven by the choice of the pre-screening era (1984 through to 1988), which in the standard linear regression model can differ by up to 1083 women (Table A6.2) but using the rate ratio method and the last year the estimate varied by an average of 2420 women.

Model adjustment: Applying an adjustment to take account of increasing incidence in women under 45 years to the standard Poisson regression model estimates reduced the estimates, on average, by 1161 women. This was also driven by the length of the pre-screening era, but the results were not incremental in the same way as the linear regression results. The models that used 1984 and 1985 as the end of the pre-screening era had very low estimates and the models that used 1986, 1987, and 1988 had relatively larger estimates.

Conclusion

The best method of assessing both the positives and negatives of breast screening would be a randomised control trial. However, in the absence of an RCT and with publically available data, the level of overdiagnosis can be estimated by extrapolation. However, the results are sensitive to the assumptions used to set up the model, and are limited by the age extension roll out between 2002 and 2004.
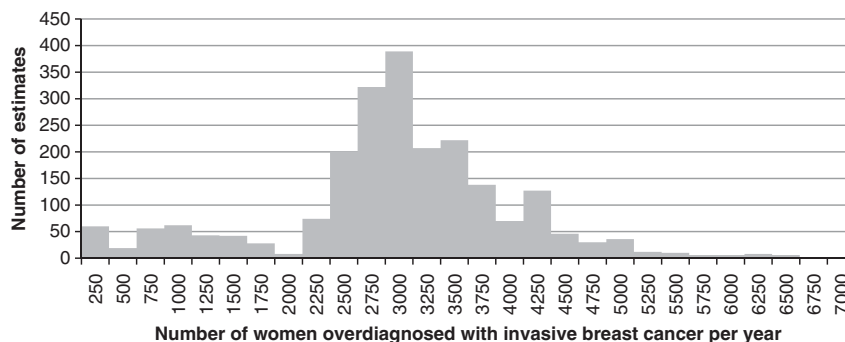

Figure A6.2. Histogram of the range of estimates.

The decision on how to adjust the regression modelling has the biggest impact on the results. However, the adjustments to the model that best represents the level of breast cancer would be in the absence of screening is unclear.

This extrapolation method assumes that the risk of breast cancer has increased at a constant rate as the period used to estimate the expected level of breast cancer ends. In addition, it assumes that the quality of case ascertainment by registries and diagnostic methods has remained stable over time. Although in theory it would be possible to adjust for these effects, how to adjust for them in practice would create further uncertainty in the estimates produced because different methods would create a further range of possible overdiagnosis estimates.

REFERENCES

References for Section 2

Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke l, Zelen M, Mandelblatt JS, Yakovlev AY, Habbema JD, Feuer EJ (2005) Effect of screening and adjuvant therapy on mortality from breast cancer. *N Engl J Med* **353**(17): 1784–1792.

Canadian Task Force on Preventive Health Care (2011) Recommendations onscreening for breast cancer in average-risk women aged 40–74 years. *Can Med Assoc J* **183**(17): 1991–2001.

Cancer Research UK (2012) *Breast Cancer Incidence Statistics*. <http://info.cancerresearchuk.org/cancerstats/types/breast/incidence/#sex> (accessed 15 September 2012).

Forrest P (1986) *Breast Cancer Screening - Report to the Health Ministers of England, Wales, Scotland and Northern Ireland* by a working group chaired by Professor Sir Patrick Forrest. Printed in the United Kingdom for Her Majesty's Stationery Office (HMSO), www.cancerscreening.nhs.uk/breastscreen/publications/forrest-report.html.

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

IARC (2002) *IARC Handbooks of Cancer Prevention - Breast Cancer Screening*. International Agency for Research on Cancer; World Health Organisation. IARC Press: Lyon.

The NHS Information Centre (2012) *Screening and Immunisations: Breast screening programme England 2010–2011* <http://www.ic.nhs.uk/webfiles/publications/008_Screening/brstscreen1011/bst_scr_prog_eng_2010_11_rep.pdf> (accessed 24 September 2012).

Woolf SH (2010) The 2009 breast cancer screening recommendations of the US Preventive Services Task Force. *JAMA* **303**(2): 162–163.

World Health Organization (2012) *Health Statistics and Health Information: Mortality Database*. <http://www.who.int/whosis/mort/download/en/index.html> (accessed 7 October 2012).

References for Section 3

Advisory Committee on Breast Cancer Screening (2006) Screening for Breast Cancer in England: Past and Future. *NHSBSP Publication*. No. 61, **2006**: pp 1–60.

Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, Muir BB, Prescott RJ, Smith A (1999) 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* **353**(9168): 1903–1908

Beral V, Alexander M, Duffy S, Ellis IO, Given-wilson R, Holmberg L, Moss SM, Ramirez A, Reed MW, Rubin C, Whelehan P, Wilson R, Young KC (2011) The number of women who would need to be screened regularly by mammography to prevent one death from breast cancer. *J Med Screen* **18**(4): 210–212.

Canadian Task Force on Preventive Health Care (2011) Recommendations on screening for breast cancer in average-risk women aged 40–74 years. *Can Med Association J* **183**(17): 1991–2001.

Connor RJ, Boer R, Prorok PC, Weed DL (2000) Investigation of design and bias issues in case-control studies of cancer screening using microsimulation. *Am J Epidemiol* **151**(10): 991–998

Cuzick J, Edwards R, Segnan N (1997) Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med* **16**(9): 1017–1029.

Duffy SW, Cuzick J, Tabar L, Vitak B, Chen TH-H, Yen MF, Smith RA (2002) Correcting for non-compliance bias in case-control studies to evaluate cancer screening programmes. *Appl Stat* **51**: 235–243.

Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen TH, Yen AM, Smith RA (2010) 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* **17**(1): 25–30.

Duffy SW, Yen Ming-Fang A, Chen TH, Chen SL, Chui SY, Fan JJ, Smith RA, Vitak B, Tábar L (2012) Long-term benefits of breast screening. *Breast Cancer Manage* **1**(1): 31–38.

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

McPherson K (2010) Screening for breast cancer–balancing the debate. *BMJ* **340**: c3106.

Moss SM, Cuckle H, Evans A, Johns L, Waller M, 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**(9552): 2053–2060.

Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L (2009) Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med* **151**(10): 727–737, W237–W242.

Njor S, Nyström L, Moss S, Paci E, Broeders M, Segnan N, Lynge E (2012) Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. *J Med Screen* **19**(Suppl 1): 33–41.

Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjold B, Rutqvist LE (2002) Long-term effects of mammography screening: updated overview of the Swedish randomised trials. *Lancet* **359**(9310): 909–919.

Puffer S, Torgerson D, Watson J (2003) Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* **327**(7418): 785–789.

Schünemann H, Brozek, Guyatt G, Oxman A (eds) (2011) *GRADE Handbook for Grading the Quality of Evidence and the Strength of Recommendations*. The GRADE Working Group. Available at http://www.gradepro.org/gradepro/.

Tábar L, Vitak B, Chen TH, Yen AM, Cohen A, Tot T, Chiu SY, Chen SL, Fann JC, Rosell J, Fohlin H, Smith RA, Duffy SW (2011) Swedish two-county trial: impact of mammographic screening on breast cancer mortality during 3 decades. *Radiology* **260**(3): 658–663.

The NHS Information Centre, Public Health Indicators Team (2011) *Breast screening programme England 2009–2010*. <http://www.ic.nhs.uk/statistics-and-data-collections/screening/breast screening> (accessed 24 September 2012).

United Kingdom Co-ordinating Committee On Cancer Research (2002) The frequency of breast cancer screening: results from the UKCCCR Randomised Trial. *Eur J Cancer* **38**(11): 1458–1464.

Zackrisson S, Andersson I, Manjer J, Janzon L (2004) Non-attendance in breast cancer screening is associated with unfavourable socio-economic circumstances and advanced carcinoma. *Int J Cancer* **108**(5): 754–760.

References for Section 4

Andersson I, Aspegren K, Janzon L, Landberg T, Lindholm K, Linell F, Ljungberg O, Ranstam J, Sigfusson B (1988) Mammographic screening and mortality from breast cancer: the Malmo mammographic screening trial. *BMJ* **297**(6654): 943–948.

Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L (2007) Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *Lancet Oncol* **8**(12): 1129–1138.

Chu KC, Smart CR, Tarone RE (1988) Analysis of breast cancer mortality and stage distribution by age for the Health Insurance Plan clinical trial. *J Natl Cancer Inst* **80**(14): 1125–1132.

Cuzick J, Sestak I, Pinder SE, Ellis IO, Forsyth S, Bundred NJ, Forbes JF, Bishop H, Fentiman IS, George WD (2011) 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* **12**(1): 21–29.

De Gelder R, Heijnsdijk EA, Van Ravesteyn NT, Fracheboud J, Draisma G, De Koning HJ (2011) Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* **33**(1) 111–121.

Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen TH, Yen AM, Smith RA (2010) 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* **17**(1): 25–30.

Elmore JG, Fletcher SW (2012) Overdiagnosis in breast cancer screening: time to tackle an underappreciated harm. *Ann Intern Med* **156**(7): 536–537.

Esserman L, Shieh Y, Thompson I (2009) Rethinking screening for breast cancer and prostate cancer. *JAMA* **302**(15): 1685–1692.

Evans A (2012) Professor of Breast Imaging, Medical School, University of Dundee, Dundee DD1 9SY. Personal communication.

Fletcher SW (2011) Breast cancer screening: a 35-year perspective. *Epidemiol Rev* **33**(1): 165–175.

Gøtzsche PC (2004) On the benefits and harms of screening for breast cancer. *Int J Epidemiol* **33**(1): 56–64 (discussion 69–73).

Harris R, Yeatts J, Kinsinger L (2011) Breast cancer screening for women ages 50 to 69 years a systematic review of observational evidence. *Prev Med* **53**(3): 108–114.

IARC (2002) *IARC Handbooks of Cancer Prevention - Breast Cancer Screening*. International Agency for Research on Cancer; World Health Organisation.

Jørgensen KJ, Gøtzsche PC (2009) Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. *BMJ* **339**: b2587.

Jørgensen KJ, Zahl PH, Gøtzsche PC (2009) Overdiagnosis in organised mammography screening in Denmark. A comparative study. *BMC Womens Health* **9**: 36.

Miller AB, To T, Baines CJ, Wall C (2000) Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50–59 years. *J Natl Cancer Inst* **92**(18): 1490–1499.

Miller AB, To T, Baines CJ, Wall C (2002) The Canadian National Breast Screening Study-1: breast cancer mortality after 11 to 16 years of follow-up. A randomized screening trial of mammography in women age 40 to 49 years. *Ann Intern Med* **137**(5): 305–312.

Moss S (2005) Overdiagnosis and overtreatment of breast cancer: overdiagnosis in randomised controlled trials of breast cancer screening. *Breast Cancer Res* **7**(5): 230–234.

National Cancer Intelligence Network (2011) *The Non-invasive Breast Cancer Report. An Analysis of Non-invasive Breast Cancers Diagnosed in England from 1 January 2006 to 31 December 2007.* <http://www.cancerscreening.nhs.uk/breastscreen/non-invasive-breast cancer-report.pdf> (accessed 14 September 2012)

The NHS Information Centre, Public Health Indicators Team (2011) *Breast Screening Programme England 2009–2010.* <http://www.ic.nhs.uk/statistics-and-data-collections/screening/breast screening> (accessed 24 September 2012).

NHS Breast Screening Programme & Association of Breast Surgery-West Midlands Cancer Intelligence Unit (2012) *An Audit of Screen Detected Breast Cancers for the Year of Screening April 2010 to March 2011.* <http://www.cancerscreening.nhs.uk/breastscreen/publications/baso2010 2011.pdf> (accessed 13 September 2012).

Pinder SE, Duggan C, Ellis IO, Cuzick J, Forbes JF, Bishop H, Fentiman IS, George WD (2010) A new pathological system for grading DCIS with improved prediction of local recurrence: results from the UKCCCR/ANZ DCIS trial. *Br J Cancer* **103**(1): 94–100.

Puliti D, Duffy SW, Miccinesi G, De Koning H, Lynge E, Zappa M, Paci E (2012) Overdiagnosis in mammographic screening for breast cancer in Europe: a literature review. *J Med Screen* **19**(Suppl 1): 42–56.

Puliti D, Miccinesi G, Paci E (2011) Overdiagnosis in breast cancer: design and methods of estimation in observational studies. *Prev Med* **53**(3): 131–133.

Reeves GK, Pirie K, Green J, Bull D, Beral V (2012) Comparison of the effects of genetic and environmental risk factors on in situ and invasive ductal breast cancer. *Int J Cancer* **131**(4): 930–937.

Seigneurin A, Francois O, Labarere J, Oudeville P, Monlong J, Colonna M (2011) Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. *BMJ* **343**: d7017.

Shapiro S (1977) Evidence on screening for breast cancer from a randomized trial. *Cancer* **39**: 2772–2782.

Shapiro S, Venet W, Strax P, Venet L, Roeser R (1982) Ten- to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst* **69**(Suppl 6): 349–355.

United Kingdom Co-ordinating Committee on Cancer Research (2002) The frequency of breast cancer screening: results from the UKCCCR randomised trial. *Eur J Cancer* **38**(11): 1458–1464.

Wallis MG, Clements K, Kearins O, Ball G, Macartney J, Lawrence GM (2012) The effect of DCIS grade on rate, type and time to recurrence after 15 years of follow-up of screen-detected DCIS. *Br J Cancer* **106**(10): 1611–1617.

Welch HG, Black WC (2010) Overdiagnosis in cancer. *J Natl Cancer Inst* **102**(9): 605–613

Welch HG, Schwartz LM, Woloshin S (2006) Ramifications of screening for breast cancer: 1 in 4 cancers detected by mammography are pseudocancers. *BMJ* **332**(7543): 727.

Zackrisson S, Andersson I, Janzon L, Manjer J, Garne JP (2006) Rate of over-diagnosis of breast cancer 15 years after end of Malmö mammographic screening trial: follow-up study. *BMJ* **332**(7543): 689–692.

References for Section 5

Advisory Committee on Breast Cancer Screening (2006) Screening for Breast Cancer in England: Past and Future. *NHSBSP Publication*. No. 61, **2006**: pp 1–60

Azim HA, JR, De Azambuja E, Colozza M, Bines J, Piccart MJ (2011) Long-term toxic effects of adjuvant chemotherapy in breast cancer. *Ann Oncol* **22**(9): 1939–1947.

Berrington de Gonzalez A (2011) Estimates of the potential risk of radiation-related cancer from screening in the UK. *J Med Screen* **18**(4): 163–164.

Bond M, Pavey T, Welch K, Cooper C, Garside R, Dean S, Hyde CJ (2012) Psychological consequences of false-positive screening mammograms in the UK. *Evid Based Med* **18**(2): 54–61.

Brett J, Austoker J (2001) Women who are recalled for further investigation for breast screening: psychological consequences 3 years after recall and factors affecting re-attendance. *J Public Health Med* **23**: 292–300.

Brett J, Austoker J, Ong G (1998) Do women who undergo further investigation for breast screening suffer adverse psychological consequences? A multi-centre follow-up study comparing different breast screening result groups five months after their last breast screening appointment. *J Public Health Med* **20**: 396–403.

Cameron DA, Massie C, Kerr G, Leonard RC (2003) Moderate neutropenia with adjuvant CMF confers improved survival in early breast cancer. *Br J Cancer* **89**(10): 1837–1842.

Clarke M, Collins R, Darby S, Davies C, Elphinstone P, Evans E, Godwin J, Gray R, HICKS C, James S, Mackinnon E, Mcgale P, Mchugh T, Peto R, Taylor C, Wang Y (2005) Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet* **366**(9503): 2087–2106.

de Leo D, Predieri M, Melodia C, Vella J, Forza G de Bertolini C (1991) Suicide attitude in breast cancer patients. *Psychopathology* **24**(2): 115–119.

Department of Health (2006) *The Coronary Heart Disease National Service Framework: Shaping the Future – Progress Report for 2006.* <http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/Publications olicyAndGuidance/DH_063168> (accessed 24 October).

Dewar JA, Horobin JM, Preece PE, Tavendale R, Tunstall-Pedoe H, Wood RA (1992) Long term effects of tamoxifen on blood lipid values in breast cancer. *BMJ* **305**(6847): 225–226.

Early Breast Cancer Trialists' Collaborative Group (EBCTCG) (2005) Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet* **365**(1687): 717.

Giordano SH, Kuo YF, Freeman JL, Buchholz TA, Hortobagyi GN, Goodwin JS (2005) Risk of cardiac death after adjuvant radiotherapy for breast cancer. *J Natl Cancer Inst* **97**(6): 419–424.

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

Health Protection Agency (2001) X-rays: How safe are they? <http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1194947388410> (accessed 24 September 2012)

Jamison KR, Wellisch DK, Pasnau RO (1978) Psychosocial aspects of mastectomy: I. the women's perspective. *Am J Psychiatry* **135**(4): 432–436.

Lawrence G (2012) West Midlands data. Personal communication.

Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L (2009) Screening for breast cancer: an update for the U.S. Preventive Services Task Force. *Ann Intern Med* **151**(10): 727–737, W237–W242.

NHS Breast Screening Programme & Association of Breast Surgery-West Midlands Cancer Intelligence Unit (2012) *An Audit of Screen Detected Breast Cancers for the Year of Screening April 2010 to March 2011.* <http://www.cancerscreening.nhs.uk/breastscreen/publications/baso2010 2011.pdf > (accessed 13 September 2012)

NHS Breast Screening Programme (2011) *Programme Annual Review 2011.* <http://www.cancerscreening.nhs.uk/breastscreen/publications/nhsbsp annualreview2011.pdf> (accessed 24 October 2012)

NHS Breast Screening Programme (2012) *Programme Logistics 2012.* <http://www.cancerscreening.nhs.uk/breastscreen/cost.html > (accessed 18 October 2012).

The NHS Information Centre (2011) *National Mastectomy and Breast Reconstruction Audit – 4th Annual Report 2011.* <http://www.ic.nhs.uk/ services/national-clinical-audit-support-programme ncasp/audit-reports/ mastectomy-and-breast-reconstruction> (accessed 13 September 2012)

Parkin DM, Tappenden P, Olsen AH, Patnick J, Sasieni P (2008) Predicting the impact of the screening programme for colorectal cancer in the UK. *J Med Screen* **15**(4): 163–174.

Peto J, Gilham C, Fletcher O, Matthews FE (2004) The cervical cancer epidemic that screening has prevented in the UK. *Lancet* **364**(9430): 249–256.

Peto R, Davies C, Godwin J, Gray R, Pan HC, Clarke M, Cutter D, Darby S, Mcgale P, Taylor C, Wang YC, Bergh J, Di Leo A, Albain K, Swain S, Piccart M, Pritchard K (2012) Comparisons between different polychemotherapy regimens for early breast cancer: meta-analyses of long-term outcome among 100,000 women in 123 randomised trials. *Lancet* **379**(9814): 432–444.

Preston DL, Mattsson A, Holmberg E, Shore R, Hildreth NG, Boice Jr JD, (2002) Radiation effects on breast cancer risk: a pooled analysis of eight cohorts. *Radiat Res* **158**(2): 220–235.

Schairer C, Brown LM, CHEN BE, Howard R, Lynch CF, Hall P, Storm H, Pukkala E, Anderson A, Kaijser M, Andersson M, Joensuu H, Fossa SD, GANZ PA, Travis LB (2006) Suicide after breast cancer: an international population-based study of 723,810 women. *J Natl Cancer Inst* **98**(19): 1416–1419.

Welch G, Schwartz L, Woloshin S (2011) *Overdiagnosed: Making People Sick in Pursuit of Health. 'We Look Harder for Breast Cancer'.* Beacon Press: Boston.

Yousaf U, Christensen ML, Engholm G, Storm HH (2005) Suicides among Danish cancer patients 1971–1999. *Br J Cancer* **92**(6): 995–1000.

Reference for Section 6

The Sloane Project 2010 (2011) *UK Prospective Audit of Screen Detected Non Invasive Carcinomas and Atypical Hyperplasias of the Breast: Progress Report 2008/09 and 2009/10.* <http://www.sloaneproject. co.uk/ documents/Sloane%20Annual%20Report%202010%20web.pdf> (accessed 12 October 2012).

References for Appendix 2

Cancer Research UK. *Breast Cancer Mortality Stat.* http://www. cancerresearchuk.org/cancer info/cancerstats/types/breast/mortality/ #trends (accessed 25 October 2012)

General Register Office for Scotland (GROS) (2012) Deaths Time Series Data, Deaths in Scotland in 2010. Edinburgh, <http://gro-scotland.gov.uk/ statistics/theme/ vital-events/general/ref tables/2010/deaths.html> (accessed 15 September 2012)

Kesson EM, Allardice GM, George WD, Burns HJ, Morrison DS (2012) Effects of multidisciplinary team working on breast cancer survival: retrospective, comparative, interventional cohort study of 13 722 women. *BMJ* **344**: e2718.

Office for National Statistics (ONS). 2012 Mortality Statistics: deaths registered in 2010, England and Wales (PDF 798KB) 2010, *Natl Stat*: London 2011, <http://www.ons.gov.uk/ons/publications/all releases. html?definition=tcm%3A77-27475> (accessed 15 September 2012)

Northern Ireland Statistics and Research Agency (NISRA) (2011) Registrar General Annual Report—2010 Northern Ireland Statistics and Research Agency, <http://www.nisra.gov.uk/demography/default.asp50.html > (accessed 15 September 2012)

References for Appendix 3

Allgood PC, Warwick J, Warren RM, Day NE, Duffy SW (2008) A case-control study of the impact of the East Anglian breast screening programme on breast cancer mortality. *Br J Cancer* **98**: 206–209.

Broeders MJ, Verbeek AL, Straatman H, Peer PG, Jong PC, Beex LV, Hendriks JH, Holland R (2002) Repeated mammographic screening reduces breast cancer mortality along the continuum of age. *J Med Screen* **9**(4): 163–167.

Collette HJ, Day NE, Rombach JJ, De Waard F (1984) Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case–control study. *Lancet* **1**(8388): 1224–1226.

Elmore JG, Reisch LM, Barton MB, Barlow WE, Rolnick S, Harris EL, Herrinton LJ, Geiger AM, Beverly RK, Hart G, Yu O, Greene SM, Weiss NS, Fletcher SW (2005) Efficacy of breast cancer screening in the community according to risk level. *J Natl Cancer Inst* **97**(14): 1035–1043.

Fielder HM, Warwick J, Brook D, Gower-Thomas K, Cuzick J, Monypenny I, 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**(4): 194–198.

Friedman DR, Dubin N (1991) Case–control evaluation of breast cancer screening efficacy. *Am J Epidemiol* **133**(10): 974–984.

Gabe R, Tryggvadottir L, Sigfusson BF, Olafsdottir GH, Sigurdsson K, Duffy SW (2007) A case–control study to estimate the impact of the Icelandic population-based mammography screening program on breast cancer death. *Acta Radiol* **48**(9): 948–955.

Gullberg B, Andersson I, Janzon L, Ranstam J (1991) Screening mammography. *Lancet* **337**(8735): 244.

Miltenburg GA, Peeters PH, Fracheboud J, Collette HJ (1998) Seventeen-year evaluation of breast cancer screening: the DOM project, The Netherlands. Diagnostisch Onderzoek (investigation) Mammacarcinoom. *Br J Cancer* **78**(7): 962–965.

Moss SM, Summerley ME, Thomas BT, Ellman R, Chamberlain JO (1992) A case–control evaluation of the effect of breast cancer screening in the United Kingdom trial of early detection of breast cancer. *J Epidemiol Community Health* **46**(4): 362–364.

Norman SA, Russell Localio A, Weber AL, Coates RJ, ZHOU L, Bernstein L, Malone KE, Marchbanks PA, Weiss LK, Lee NC, Nadel MR (2007) Protection of mammography screening against death from breast cancer in women aged 40-64 years. *Cancer Causes Control* **18**(9): 909–918.

Otto SJ, Fracheboud J, Verbeek AL, Boer R, Reijerink-verheij JC, Otten JD, Broeders MJ, De Koning HJ (2012) Mammography screening and risk of breast cancer death: a population-based case-control study. *Cancer Epidemiol Biomarkers Prev* **21**(1): 66–73.

Paap E, Holland R, Den Heeten GJ, Van Schoor G, Botterweck AA, Verbeek AL, Broeders MJ (2010) A remarkable reduction of breast cancer deaths in screened versus unscreened women: a case-referent study. *Cancer Causes Control* **21**(10): 1569–1573.

Palli D, Rosselli Del Turco M, Buiatti E, Ciatto S, Crocetti E, Paci E (1989) Time interval since last test in a breast cancer screening programme: a case-control study in Italy. *J Epidemiol Community Health,* **43**(3): 241–248.

Puliti D, Miccinesi G, Collina N, De Lisi V, Federico M, Ferretti S, Finarelli AC, Foca F, Mangone L, Naldoni C, Petrella M, Ponti A, Segnan N, Sigona A, Zarcone M, Zorzi M, Zappa M, Paci E (2008) Effectiveness of service screening: a case-control study to assess breast cancer mortality reduction. *Br J Cancer* **99**(3): 423–427.

Roder D, Houssami N, Farshid G, Gill G, Luke C, Downey P, Beckmann K, IOSIFIDIS P, Grieve L, Williamson L (2008) Population screening and intensity of screening are associated with reduced breast cancer mortality: evidence of efficacy of mammography screening in Australia. *Breast Cancer Res Treat* **108**(3): 409–416.

Van Dijck JA, Verbeek AL, Beex LV, Hendriks JH, Holland R, Mravunac M, Straatman H, Werre JM (1996) Mammographic screening after the age of 65 years: evidence for a reduction in breast cancer mortality. *Int J Cancer* **66**(6): 727–731.

Van Schoor G, Moss SM, Otten JD, Donders R, Paap E, DEN Heeten GJ, Holland R, Broeders MJ, Verbeek AL (2011) Increasingly strong reduction in breast cancer mortality due to screening. *Br J Cancer* **104**(6): 910–914.

Van Schoor G, Moss SM, Otten JD, Donders R, Paap E, Den Heeten GJ, Holland R, Broeders MJ, Verbeek AL (2010) Effective biennial mammographic screening in women aged 40-49. *Eur J Cancer* **46**(18): 3137–3140.

Verbeek AL, Hendriks JH, Holland R, Mravunac M, Sturmans F (1985) Mammographic screening and breast cancer mortality: age-specific effects in Nijmegen Project, 1975-82. *Lancet* **1**: 865–866.

Verbeek AL, Hendriks JH, Holland R, Mravunac M, Sturmans F, Day NE (1984) Reduction of breast cancer mortality through mass screening with modern mammography. First results of the Nijmegen project, 1975-1981. *Lancet* **1**: 1222–1224.

References for Appendix 4

De Gelder R, Heijnsdijk EA, Van Ravesteyn NT, Fracheboud J, Draisma G, De Koning HJ (2011) Interpreting overdiagnosis estimates in population-based mammography screening. *Epidemiol Rev* **33**(1): 111–121.

References for Appendix 5

NHS Breast screening leaflet 2011 < http://www.cancerscreening.nhs.uk/ breastscreen/publications/nhsbsp.pdf > (accessed 13 September)

References for Appendix 6

Office of National Statistics. Data provided on request. Similar data can be found at < http://www.ons.gov.uk/ons/search/index.html?newquery= cancer + registrations > (accessed 15 September 2012)

Other Source Material:

Autier P, Boniol M (2012) The incidence of advanced breast cancer in the West Midlands, United Kingdom. *Eur J Cancer Prev* **21**(3): 217–221.

Autier P, Boniol M, Gavin A, Vatten LJ (2011) Breast cancer mortality in neighbouring European countries with different levels of screening but similar access to treatment: trend analysis of WHO mortality database. *BMJ* **343**: d4411.

Autier P, Boniol M, La Vecchia C, Vatten L, Gavin A, Hery C, Heanue M (2010) Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. *BMJ* **341**: c3620.

Autier P, Boniol M, Middleton R, Dore JF, Hery C, Zheng T, Gavin A (2011) Advanced breast cancer incidence following population-based mammographic screening. *Ann Oncol* **22**(8): 1726–1735.

Autier P, Hery C, Haukka J, Boniol M, Byrnes G (2009) Advanced breast cancer and breast cancer mortality in randomized controlled trials on mammography screening. *J Clin Oncol* **27**(35): 5919–5923.

Autier P, Koechlin A, Smans M, Vatten L, Boniol M (2012) Mammography screening and breast cancer mortality in sweden. *J Natl Cancer Inst* **104**(14): 1080–1093.

Baines CJ, Miller AB, Kopans DB, Moskowitz M, Sanders DE, Sickles EA, To T, Wall C (1990) Canadian National Breast Screening Study: assessment of technical quality by external review. *AJR Am J Roentgenol* **155**(4): 743–747; discussion 748—749.

Bakacs T, Mehrishi JN (2010) Breast and other cancer dormancy as a therapeutic endpoint: speculative recombinant T cell receptor ligand (RTL) adjuvant therapy worth considering? *BMC Cancer* **10**: 251.

Banks E, Beral V, Reeves G (2004) Published results on breast cancer and hormone replacement therapy in the Million Women Study are correct. *Climacteric* **7**(4): 415–416; author reply 416–417.

Barratt A, Howard K, Irwig L, Salkeld G, Houssami N (2005) Model of outcomes of screening mammography: information to support informed choices. *BMJ* **330**(7497): 936.

Baum M (2004) Commentary: false premises, false promises and false positives—the case against mammographic screening for breast cancer. *Int J Epidemiol* **33**(1): 66–67; discussion 69–73.

Beral V, Alexander M, Duffy S, Ellis IO, Given-Wilson R, Holmberg L, Moss SM, Ramirez A, Reed MW, Rubin C, Whelehan P, Wilson R, Young KC (2011) The number of women who would need to be screened regularly by mammography to prevent one death from breast cancer. *J Med Screen* **18**(4): 210–212.

Beral V, Bull D, Green J, Reeves G (2007) Ovarian cancer and hormone replacement therapy in the Million Women Study. *Lancet* **369**(9574): 1703–1710.

Beral V, Bull D, Reeves G (2005) Endometrial cancer and hormone-replacement therapy in the Million Women Study. *Lancet* **365**(9470): 1543–1551.

Berry D (2004) Commentary: screening mammography: a decision analysis. *Int J Epidemiol* **33**(1): 68; discussion 69–73.

Bewley S (2011) The NHS breast screening programme needs independent review. *BMJ* **343**: d6894.

Blanks RG, Moss SM, McGahan CE, Quinn MJ, Babb PJ (2000) 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**(7262): 665–669.

Bonuccelli G, Tsirigos A, Whitaker-Menezes D, Pavlides S, Pestell RG, Chiavarina B, Frank PG, Flomenberg N, Howell A, Martinez-Outschoorn UE, Sotgia F, Lisanti MP (2010) Ketones and lactate 'fuel' tumor growth and metastasis: evidence that epithelial cancer cells use oxidative mitochondrial metabolism. *Cell Cycle* **9**(17): 3506–3514.

Brinkley D, Haybrittle JL (1975) The curability of breast cancer. *Lancet* **2**(7925): 95–97.

Bulliard JL, Levi F (2012) Mammography screening: time to reevaluate its impact? *Eur J Cancer Prev* **21**(3): 222–226.

Burton RC, Bell RJ, Thiagarajah G, Stevenson C (2012) Adjuvant therapy, not mammographic screening, accounts for most of the observed breast cancer specific mortality reductions in Australian women since the national screening program began in 1991. *Breast Cancer Res Treat* **131**(3): 949–955.

Caan B, Habel L, Quesenberry C, Kushi L, Herrinton L. (2008) Re: Declines in invasive breast cancer and use of postmenopausal hormone therapy in a screening mammography population. *J Natl Cancer Inst* **100**(8): 597–598; author reply 99.

Caldarella A, Puliti D, Crocetti E, Bianchi S, Vezzosi V, Apicella P, Biancalani M, Giannini A, Urso C, Zolfanelli F, Paci E (2013) Biological characteristics of interval cancers: a role for biomarkers in the breast cancer screening. *J Cancer Res Clin Oncol* **139**(2): 181–185.

Castello-Cros R, Bonuccelli G, Molchansky A, Capozza F, Witkiewicz AK, Birbe RC, Howell A, Pestell RG, Whitaker-Menezes D, Sotgia F, Lisanti MP (2011) Matrix remodeling stimulates stromal autophagy, 'fueling' cancer cell mitochondrial metabolism and metastasis. *Cell Cycle* **10**(12): 2021–2034.

Cockburn J, Staples M, Hurley SF, De Luise T (1994) Psychological consequences of screening mammography. *J Med Screen* **1**(1): 7–12

Coleman MP, Esteve J, Damiecki P, Arslan A, Renard H (1993)Trends in cancer incidence and mortality. *IARC Sci Publ* **1**(121): 1–806.

Coleman MP. Trends in breast cancer incidence, survival, and mortality. (2000) *Lancet* **356**(9229): 590–591.

Collaborative Group on Hormonal Factors in Breast Cancer. (2001) Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet* **358**(9291): 1389–1399.

Collins LC, Tamimi RM, Baer HJ, Connolly JL, Colditz GA, Schnitt SJ (2005) Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: results from the Nurses' Health Study. *Cancer* **103**(9): 1778–1784

Coronary Drug Project Research Group. (1980) Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* **303**(18): 1038–1041.

Cuzick J, Sestak I, Baum M, Buzdar A, Howell A, Dowsett M, Forbes JF (2010) Effect of anastrozole and tamoxifen as adjuvant treatment for early-stage breast cancer: 10-year analysis of the ATAC trial. *Lancet Oncol* **11**(12): 1135–1141.

de Gelder R, Fracheboud J, Heijnsdijk EA, den Heeten G, Verbeek AL, Broeders MJ, Draisma G, de Koning HJ (2011) Digital mammography screening: weighing reduced mortality against increased overdiagnosis. *Prev Med* **53**(3): 134–140.

de Gelder R, van As E, Tilanus-Linthorst MM, Bartels CC, Boer R, Draisma G, de Koning HJ (2008) Breast cancer screening: evidence for false reassurance? *Int J Cancer* **123**(3): 680–686.

de Haes JC, de Koning HJ, van Oortmarssen GJ, van Agt HM, de Bruyn AE, van Der Maas PJ (1991) The impact of a breast cancer screening programme on quality-adjusted life-years. *Int J Cancer* **49**(4): 538–544.

de Koning HJ, van Ineveld BM, van Oortmarssen GJ, de Haes JC, Collette HJ, Hendriks JH, van der Maas PJ (1991) Breast cancer screening and cost effectiveness; policy alternatives, quality of life considerations and the possible impact of uncertain factors. *Int J Cancer* **49**(4): 531–537.

de Roos MA, van der Vegt B, de Vries J, Wesseling J, de Bock GH (2007) Pathological and biological differences between screen-detected and interval ductal carcinoma in situ of the breast. *Ann Surg Oncol* **14**(7): 2097–2104.

Demissie K, Mills OF, Rhoads GG (1998) Empirical comparison of the results of randomized controlled trials and case-control studies in evaluating the effectiveness of screening mammography. *J Clin Epidemiol* **51**(2): 81–91.

Duffy SW (2002) The controversy over breast cancer screening. *Hosp Med* **63**(12): 708–709.

Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Bjorneld L, Myles JP, Warwick J (2005) Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Res* **7**(6): 258–265.

Duffy SW, Lynge E, Jonsson H, Ayyaz S, Olsen AH (2008) Complexities in the estimation of overdiagnosis in breast cancer screening. *Br J Cancer* **99**(7): 1176–1178.

Duffy SW, Tabar L, Vitak B, Yen MF, Warwick J, Smith RA, Chen HH (2003) The Swedish two-county trial of mammographic screening: cluster randomisation and end point evaluation. *Ann Oncol* **14**(8): 1196–1198.

Eilbert KW, Carroll K, Peach J, Khatoon S, Basnett I, McCulloch N (2009) Approaches to improving breast screening uptake: evidence and experience from Tower Hamlets. *Br J Cancer* **101**(Suppl 2): S64–S67.

Esserman L, Thompson I (2010) Solving the overdiagnosis dilemma. *J Natl Cancer Inst* **102**(9): 582–583.

Esserman LJ, Shieh Y, Rutgers EJ, Knauer M, Retel VP, Mook S, Glas AM, Moore DH, Linn S, van Leeuwen FE, van 't Veer LJ (2011) Impact of mammographic screening on the detection of good and poor prognosis breast cancers. *Breast Cancer Res Treat* **130**(3): 725–734.

Etzioni R, Urban N, Ramsey S, McIntosh M, Schwartz S, Reid B, Radich J, Anderson G, Hartwell L (2003) The case for early detection. *Nat Rev Cancer* **3**(4): 243–252.

Euroscreen Working Group. (2012) Summary of the evidence of breast cancer service screening outcomes in Europe and first estimate of the benefit and harm balance sheet. *J Med Screen* **19**(s1): 5–13.

Fallowfield LJ, Hall A (1991) Psychosocial and sexual impact of diagnosis and treatment of breast cancer. *Br Med Bull* **47**(2): 388–399.

Feig SA (2012) Number needed to screen: appropriate use of this new basis for screening mammography guidelines. *AJR Am J Roentgenol* **198**(5): 1214–1217.

Fisher B, Gunduz N, Coyle J, Rudock C, Saffer E (1989) Presence of a growth stimulating factor in serum following primary tumor removal in mice. *Cancer Res* **49**(8): 1996–2001.

Fletcher SW, Elmore JG (2011) More on screening mammography. *N Engl J Med* **364**(3): 283 285–286.

Fracheboud J, Groenewoud JH, Boer R, Draisma G, de Bruijn AE, Verbeek AL, de Koning HJ (2006) Seventy-five years is an appropriate upper age limit for population-based mammography screening. *Int J Cancer* **118**(8): 2020–2025.

Freedman DA, Petitti DB, Robins JM (2004) On the efficacy of screening for breast cancer. *Int J Epidemiol* **33**(1): 43–55.

Gabe R, Duffy SW (2005) Evaluation of service screening mammography in practice: the impact on breast cancer mortality. *Ann Oncol* **16** (Suppl 2): ii153–ii162.

Garcia-Closas M, Sherman ME, Brinton LA (2006) Clarifying breast cancer risks associated with menopausal hormone therapy. *Lancet Oncol* **7**(11): 885–886.

Gilbert FJ, Cordiner CM, Affleck IR, Hood DB, Mathieson D, Walker LG (1998) Breast screening: the psychological sequelae of false-positive recall in women with and without a family history of breast cancer. *Eur J Cancer* **34**(13): 2010–2014.

Giordano L, Cogo C, Patnick J, Paci E. (2012) Communicating the balance sheet in breast cancer screening. *J Med Screen* **19** (Suppl 1):67–71.

Giordano L, von Karsa L, Tomatis M, Majek O, de Wolf C, Lancucki L, Hofvind S, Nyström L, Segnan N, Ponti A, Van Hal G, Martens P, Danes J, von Euler Chelpin M, Aasmaa A, Anttila A, Becker N, Pentek Z, Budai A, Madai S, Fitzpatrick P, Mooney T, Zappa M, Ventura L, Scharpantgen A, Seroczynski P, Morais A, Rodrigues V, Bento MJ, Gomes de Carvalho J, Natal C, Prieto M, Sanchez-Contador Escudero C, Zubizarreta Alberti R, Fernandez Llanes SB, Ascunce N, Ederra Sanza M, Sarriugarte Irigoien G,

Salas Trejo D, Ibanez Cabanell J, Wiege M, Ohlsson G, Tornberg S, Korzeniewska M, Fracheboud J, Patnick JJ, Ducarroz S, Suonio E (2012) Mammographic screening programmes in Europe: organization, coverage and participation. *J Med Screen* **19** (Suppl 1): 72–82.

Glass AG, Lacey Jr JV, Carreon JD, Hoover RN (2007) Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. *J Natl Cancer Inst* **99**(15): 1152–1161.

Glasziou P, Houssami N (2011) The evidence base for breast cancer screening. *Prev Med* **53**(3): 100–102.

Gompel A, Plu-Bureau G (2010) Is the decrease in breast cancer incidence related to a decrease in postmenopausal hormone therapy? *Ann N Y Acad Sci* **1205**: 268–276.

Gompel A, Santen RJ (2012) Hormone therapy and breast cancer risk 10 years after the WHI. *Climacteric* **15**(3): 241–249.

Gøtzsche PC (2011) Relation between breast cancer mortality and screening effectiveness: systematic review of the mammography trials. *Dan Med Bull* **58**(3): A4246.

Gøtzsche PC (2004) On the benefits and harms of screening for breast cancer. *Int J Epidemiol* **33**(1): 56–64.

Gøtzsche PC (2006) Ramifications of screening for breast cancer: overdiagnosis in the Malmo trial was considerably underestimated. *BMJ* **332**(7543): 727.

Gøtzsche PC, Jørgensen KJ (2011) The breast screening programme and misinforming the public. *J R Soc Med* **104**(9): 361–369.

Gøtzsche PC, Jørgensen KJ, Zahl PH, Maehlen J (2012) Why mammography screening has not lived up to expectations from the randomised trials. *Cancer Causes Control* **23**(1): 15–21.

Gøtzsche PC, Nielsen M (2009) Screening for breast cancer with mammography. *Cochrane Database Syst Rev* (4): CD001877.

Gøtzsche PC (2004) Misleading quotations and other errors persist in rejoinder on breast cancer screening. *Int J Epidemiol* **33**(6): 1404.

Gøtzsche PC (2012) Mammography screening: truth, lies, and controversy. *Lancet* **380**(9838): 218.

Green J, Cairns BJ, Casabonne D, Wright FL, Reeves G, Beral V (2011) Height and cancer incidence in the Million Women Study: prospective cohort, and meta analysis of prospective studies of height and total cancer risk. *Lancet Oncol* **12**(8): 785–794.

Green J, Roddam A, Pirie K, Kirichek O, Reeves G, Beral V (2012) Reproductive factors and risk of oesophageal and gastric cancer in the Million Women Study cohort. *Br J Cancer* **106**(1): 210–216.

Groenewoud JH, Otten JD, Fracheboud J, Draisma G, van Ineveld BM, Holland R, Verbeek AL, de Koning HJ (2007) Cost-effectiveness of different reading and referral strategies in mammography screening in the Netherlands. *Breast Cancer Res Treat* **102**(2): 211–218.

Hakama M, Pukkala E, Heikkila M, Kallio M (1997) Effectiveness of the public health policy for breast cancer screening in Finland: population based cohort study. *BMJ* **314**(7084): 864–847.

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

Hirsch BR, Lyman GH (2011) Breast cancer screening with mammography. *Curr Oncol Rep* **13**(1): 63–70.

Hofvind S, Ponti A, Patnick J, Ascunce N, Njor S, Broeders M, Giordano L, Frigerio A, Tornberg S, Van Hal G, Martens P, Majek O, Danes J, von Euler-Chelpin M, Aasmaa A, Anttila A, Becker N, Pentek Z, Budai A, Madai S, Fitzpatrick P, Mooney T, Zappa M, Ventura L, Scharpantgen A, Seroczynski P, Morais A, Rodrigues V, Bento MJ, Gomes de Carvalho J, Natal C, Prieto M, Sanchez Contador Escudero C, Zubizarreta Alberti R, Fernandez Llanes SB, Ederra Sanza M, Sarriugarte Irigoien G, Salas Trejo D, Ibanez Cabanell J, Wiege M, Ohlsson G, Korzeniewska M, de Wolf C, Fracheboud J, Lancucki L, Ducarroz S, Suonio E (2012) False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes. *J Med Screen* **19** (Suppl 1): 57–66.

Jonsson H, Johansson R, Lenner P (2005) Increased incidence of invasive breast cancer after the introduction of service screening with mammography in Sweden. *Int J Cancer* **117**(5): 842–847.

Jonsson H, Nyström L, Tornberg S, 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**(3): 152–160.

Jørgensen KJ (2011) Flawed methods explain the effect of mammography screening in Nijmegen. *Br J Cancer* **105**(4): 592–593.

Jørgensen KJ (2012) Is the tide turning against breast screening? *Breast Cancer Res* **14**(4): 107.

Jørgensen KJ, Keen JD, Gøtzsche PC (2011) Is mammographic screening justifiable considering its substantial overdiagnosis rate and minor effect on mortality? *Radiology* **260**(3): 621–627.

Jørgensen KJ, Keen JD, Zahl PH, Gøtzsche PC (2012) The Two-County breast screening trial cannot provide a reliable estimate of the effect of breast cancer screening. *Radiology* **262**(2): 729–730.

Jørgensen KJ, Zahl PH, Gotzsche PC (2010) Breast cancer mortality in organised mammography screening in Denmark: comparative study. *BMJ* **340**: c1241.

Junod B, Zahl PH, Kaplan RM, Olsen J, Greenland S (2011) An investigation of the apparent breast cancer epidemic in France: screening and incidence trends in birth cohorts. *BMC Cancer* **11**: 401.

Kalager M, Adami HO, Bretthauer M, Tamimi RM (2012) Overdiagnosis of invasive breast cancer due to mammography screening. *Ann Intern Med* **157**(3): 221–222.

Kalager M, Zelen M, Langmark F, Adami HO (2010) Effect of screening mammography on breast-cancer mortality in Norway. *N Engl J Med* **363**(13): 1203–1210.

Kerlikowske K, Cook AJ, Buist DS, Cummings SR, Vachon C, Vacek P, Miglioretti DL (2010) Breast cancer risk by breast density, menopause, and postmenopausal hormone therapy use. *J Clin Oncol* **28**(24): 3830–3837.

Keyzer-Dekker CM, van Esch L, de Vries J, Ernst MF, Nieuwenhuijzen GA, Roukema JA, van der Steeg AF (2012) An abnormal screening mammogram causes more anxiety than a palpable lump in benign breast disease. *Breast Cancer Res Treat* **134**(1): 253–258.

Knox G (1991) Case-control studies of screening procedures. *Public Health* **105**(1): 55–61.

Larsson LG, Nyström L, Wall S, Rutqvist L, Andersson I, Bjurstam N, Fagerberg G, Frisell J, Tabar L (1996) The Swedish randomised mammography screening trials: analysis of their effect on the breast cancer related excess mortality. *J Med Screen* **3**(3): 129–132.

Logan RF, Patnick J, Nickerson C, Coleman L, Rutter MD, von Wagner C and English Bowel Cancer Screening Evaluation Committee (2012) Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut* **61**(10): 1439–1446.

Lynge E, Braaten T, Njor SH, Olsen AH, Kumle M, Waaseth M, Lund E (2011) Mammography activity in Norway 1983 to 2008. *Acta Oncol* **50**(7): 1062–1067.

Magnus MC, Ping M, Shen MM, Bourgeois J, Magnus JH (2011) Effectiveness of mammography screening in reducing breast cancer mortality in women aged 39-49 years: a meta-analysis. *J Womens Health (Larchmt)* **20**(6): 845–852.

Malmgren JA, Parikh J, Atwood MK, Kaplan HG (2012) Impact of mammography detection on the course of breast cancer in women aged 40-49 years. *Radiology* **262**(3): 797–806.

Mandelblatt JS, Cronin KA, Berry DA, Chang Y, de Koning HJ, Lee SJ, Plevritis SK, Schechter CB, Stout NK, van Ravesteyn NT, Zelen M, Feuer EJ (2011) Modeling the impact of population screening on breast cancer mortality in the United States. *Breast* **20** (Suppl 3): S75–S81.

Martinez-Alonso M, Vilaprinyo E, Marcos-Gragera R, Rue M (2010) Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Res* **12**(4): R58.

McCann J, Stockton D, Godward S (2002) Impact of false-positive mammography on subsequent screening attendance and risk of cancer. *Breast Cancer Res* **4**(5): R11.

Miller AB (1993) The costs and benefits of breast cancer screening. *Am J Prev Med* **9**(3): 175–180.

Miller AB (2011) Breast cancer screening: commentary and conclusions. *Prev Med* **53**(3): 147–148.

Miller AB (2011) Breast cancer screening: introduction. *Prev Med* **53**(3): 99.

Miller AB, Baines CJ, To T, Wall C (1992) Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. *CMAJ* **147**(10): 1459–1476.

Miller AB, Baines CJ, To T, Wall C (1992) Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. *CMAJ* **147**(10): 1477–1488.

Miller AB (2004) Commentary: a defence of the Health Insurance Plan (HIP) study and the Canadian National Breast Screening Study (CNBSS). *Int J Epidemiol* **33**(1): 64–65.

Morrell S, Barratt A, Irwig L, Howard K, Biesheuvel C, Armstrong B (2010) Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. *Cancer Causes Control* **21**(2): 275–282.

Morrell S, Taylor R, Roder D, Dobson A (2012) Mammography screening and breast cancer mortality in Australia: an aggregate cohort study. *J Med Screen* **19**(1): 26–34.

Moser K, Patnick J, Beral V (2009) Inequalities in reported use of breast and cervical screening in Great Britain: analysis of cross sectional survey data. *BMJ* **338**: b2025.

Moses LE (1995) Measuring effects without randomized trials? Options, problems, challenges. *Med Care* **33**(4 Suppl): AS8–AS14.

Moss SM, Nyström L, Jonsson H, Paci E, Lynge E, Njor S, Broeders M (2012) The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. *J Med Screen* **19** (Suppl 1): 26–32.

Nagtegaal ID, Allgood PC, Duffy SW, Kearins O, Sullivan EO, Tappenden N, Wallis M, Lawrence G (2011) Prognosis and pathology of screen-detected carcinomas: how different are they? *Cancer* **117**(7): 1360–1368.

National Cancer Intelligence Network (2012) The second all breast cancer report, focusing on inequalities: variation in breast cancer outcomes with deprivation and age <http://www.ncin.org.uk/publications/reports/default.aspx> (accessed 24 August 2012).

Nederend J, Duijm LE, Voogd AC, Groenewoud JH, Jansen FH, Louwman MW (2012) Trends in incidence and detection of advanced breast cancer at biennial screening mammography in The Netherlands: a population based study. *Breast Cancer Res* **14**(1): R10.

Neta G, Anderson WF, Gilbert E, Berrington A (2012) Variation in the risk of radiation related contralateral breast cancer by histology and estrogen receptor expression in SEER. *Breast Cancer Res Treat* **131**(3): 1021–1027.

Nyström L, Larsson L-G, Rutqvist LE, Lindgren A, Lindqvist M, Rydén S, Andersson I, Bjurstam N, Fagerberg G, Frisell J, Tabár L (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**(2): 145–152.

O'Dowd A (2011) Women in their 40s should not be screened for breast cancer, new Canadian guideline says. *BMJ* **343**: d7625.

Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW (2006) Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast J* **12**(4): 338–342.

Olsen AH, Njor SH, Vejborg I, Schwartz W, Dalgaard P, Jensen MB, Tange UB, Blichert-Toft M, Rank F, Mouridsen H, Lynge E (2005) Breast cancer mortality in Copenhagen after introduction of mammography screening: cohort study. *BMJ* **330**(7485): 220.

Olsson A, Borgquist S, Butt S, Zackrisson S, Landberg G, Manjer J (2012) Tumour related factors and prognosis in breast cancer detected by screening. *Br J Surg* **99**(1): 78–87.

Ong G, Austoker J (1997) Recalling women for further investigation of breast screening: women's experiences at the clinic and afterwards. *J Public Health Med* **19**(1):29–36.

Otten JD, Broeders MJ, Fracheboud J, Otto SJ, de Koning HJ, Verbeek AL (2008) Impressive time-related influence of the Dutch screening programme on breast cancer incidence and mortality, 1975-2006. *Int J Cancer* **123**(8): 1929–1934.

Otto SJ, Fracheboud J, Looman CW, Broeders MJ, Boer R, Hendriks JH, Verbeek AL, de Koning HJ (2003) Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review. *Lancet* **361**(9367): 1411–1417.

Paci E, Miccinesi G, Puliti D, Baldazzi P, De Lisi V, Falcini F, Cirilli C, Ferretti S, Mangone L, Finarelli AC, Rosso S, Segnan N, Stracci F, Traina A, Tumino R, Zorzi M (2006) Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. *Breast Cancer Res* **8**(6): R68.

Paci E, Warwick J, Falini P, Duffy SW (2004) Overdiagnosis in screening: is the increase in breast cancer incidence rates a cause for concern? *J Med Screen* **11**(1): 23–27.

Peeters PH, Verbeek AL, Straatman H, Holland R, Hendriks JH, Mravunac M, Rothengatter C, Van Dijk-Milatz A, Werre JM (1989) Evaluation of overdiagnosis of breast cancer in screening with mammography: results of the Nijmegen programme. *Int J Epidemiol* **18**(2): 295–299.

Percy C, Dolman A (1978) Comparison of the coding of death certificates related to cancer in seven countries. *Public Health Rep* **93**(4): 335–350.

Peto R, Boreham J, Clarke M, Davies C, Beral V (2000) UK and USA breast cancer deaths down 25% in year 2000 at ages 20-69 years. *Lancet* **355**(9217): 1822.

Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E (2009) An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* **45**(18): 3166–3171.

Puliti D, Miccinesi G, Manneschi G, Buzzoni C, Crocetti E, Paci E, Zappa M (2012) Does an organised screening programme reduce the inequalities in breast cancer survival? *Ann Oncol* **23**(2): 319–323.

Puliti D, Miccinesi G, Zappa M, Manneschi G, Crocetti E, Paci E (2012) Balancing harms and benefits of service mammography screening programs: a cohort study. *Breast Cancer Res* **14**(1): R9.

Puliti D, Zappa M (2012) Breast cancer screening: are we seeing the benefit? *BMC Med* **10**:106.

Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. (2009) An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. *Eur J Cancer* **45**(18):3166–3171.

Quinn M, Babb P, Jones J, Allen E (1999) Effect of screening on incidence of and mortality from cancer of cervix in England: evaluation based on routinely collected statistics. *BMJ* **318**(7188): 904–908.

Raftery J, Chorozoglou M (2011) Possible net harms of breast cancer screening: updated modelling of Forrest report. *BMJ* **343**: d7627.

Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, Edwards BK, Berry DA (2007) The decrease in breast-cancer incidence in 2003 in the United States. *N Engl J Med* **356**(16): 1670–1674.

Reeves GK, Beral V, Green J, Gathani T, Bull D (2006) Hormonal therapy for menopause and breast-cancer risk by histological type: a cohort study and meta-analysis. *Lancet Oncol* **7**(11): 910–918.

Reeves GK, Pirie K, Beral V, Green J, Spencer E, Bull D (2007) Cancer incidence and mortality in relation to body mass index in the Million Women Study: cohort study. *BMJ* **335**(7630): 1134.

Reeves GK, Travis RC, Green J, Bull D, Tipper S, Baker K, Beral V, Peto R, Bell J, Zelenika D, Lathrop M (2010) Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. *JAMA* **304**(4): 426–434.

Retsky M, Demicheli R, Hrushesky W (2001) Premenopausal status accelerates relapse in node positive breast cancer: hypothesis links angiogenesis, screening controversy. *Breast Cancer Res Treat* **65**(3): 217–224.

Retsky MW, Demicheli R, Hrushesky WJ, Baum M, Gukas ID (2008) Dormancy and surgery-driven escape from dormancy help explain some clinical features of breast cancer. *APMIS* **116**(7–8): 730–741.

Retsky M, Demicheli R, Hrushesky W. (2001) Breast cancer screening for women aged 40-49 years: screening may not be the benign process usually thought. *J Natl Cancer Inst* **93**(20): 1572.

Rijnsburger AJ, van Oortmarssen GJ, Boer R, Draisma G, To T, Miller AB, de Koning HJ (2004) Mammography benefit in the Canadian National Breast Screening Study-2: a model evaluation. *Int J Cancer* **110**(5): 756–762.

Sanders ME, Schuyler PA, Dupont WD, Page DL (2005) The natural history of low grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer* **103**(12): 2481–2484.

Schoen RE, Pinsky PF, Weissfeld JL, Yokochi LA, Church T, Laiyemo AO, Bresalier R, Andriole GL, Buys SS, Crawford ED, Fouad MN, Isaacs C, Johnson CC, Reding DJ, O'Brien B, Carrick DM, Wright P, Riley TL, Purdue MP, Izmirlian G, Kramer BS, Miller AB, Gohagan JK, Prorok PC, Berg CD (2012) Colorectal cancer incidence and mortality with screening flexible sigmoidoscopy. *N Engl J Med* **366**(25): 2345–2357.

Schopper D, de Wolf C (2009) How effective are breast cancer screening programmes by mammography? Review of the current evidence. *Eur J Cancer* **45**(11): 1916–1923.

Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, Kwiatkowski M, Lujan M, Lilja H, Zappa M, Denis LJ, Recker F, Berenguer A, Maattanen L, Bangma CH, Aus G, Villers A, Rebillard X, van der Kwast T, Blijenberg BG, Moss SM, de Koning HJ, Auvinen A (2009) Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med* **360**(13): 1320–1328.

Smith RA, Duffy SW, Gabe R, Tabar L, Yen AM, Chen TH (2004) The randomized trials of breast cancer screening: what have we learned? *Radiol Clin North Am* **42**(5): 793–806.

Smith-Bindman R, Chu PW, Miglioretti DL, Sickles EA, Blanks R, Ballard-Barbash R, Bobo JK, Lee NC, Wallis MG, Patnick J, Kerlikowske K (2003) Comparison of screening mammography in the United States and the United kingdom. *JAMA* **290**(16): 2129–2137.

Stefanek ME.(2011) Uninformed compliance or informed choice? A needed shift in our approach to cancer screening. *J Natl Cancer Inst* **103**(24):1821–1826.

Stuelten CH, Barbul A, Busch JI, Sutton E, Katz R, Sato M, Wakefield LM, Roberts AB, Niederhuber JE (2008) Acute wounds accelerate tumorigenesis by a T cell dependent mechanism. *Cancer Res* **68**(18): 7278–7282.

Suhrke P, Maehlen J, Schlichting E, Jørgensen KJ, Gøtzsche PC, Zahl PH (2011) Effect of mammography screening on surgical treatment for breast cancer in Norway: comparative analysis of cancer registry data. *BMJ* **343**: d4692.

Taylor R, Morrell S, Estoesta J, Brassil A (2004) Mammography screening and breast cancer mortality in New South Wales, Australia. *Cancer Causes Control* **15**(6): 543–550.

Tengs TO, Adams ME, Pliskin JS, Safran DG, Siegel JE, Weinstein MC, Graham JD (1995) Five-hundred life-saving interventions and their cost-effectiveness. *Risk Anal* **15**(3): 369–390.

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

Travis RC, Reeves GK, Green J, Bull D, Tipper SJ, Baker K, Beral V, Peto R, Bell J, Zelenika D, Lathrop M (2010) Gene-environment interactions in 7610 women with breast cancer: prospective evidence from the Million Women Study. *Lancet* **375**(9732): 2143–2151.

van den Akker-van Marle E, de Koning H, Boer R, van der Maas P (1999) Reduction in breast cancer mortality due to the introduction of mass screening in the Netherlands: comparison with the United Kingdom. *J Med Screen* **6**(1): 30–34.

Viel JF, Rymzhanova R, Fournier E, Danzon A (2011) Trends in invasive breast cancer incidence among French women not exposed to organized mammography screening: an age-period-cohort analysis. *Cancer Epidemiol* **35**(6): 521–525.

Wald NJ, Murphy P, Major P, Parkes C, Townsend J, Frost C (1995) UKCCCR multicentre randomised controlled trial of one and two view mammography in breast cancer screening. *BMJ* **311**(7014): 1189–1193.

Waller M, Moss S, Watson J, Moller H (2007) The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. *Cancer Epidemiol Biomarkers Prev* **16**(11): 2257–2261.

Weedon-Fekjaer H, Bakken K, Vatten LJ, Tretli S (2012) Understanding recent trends in incidence of invasive breast cancer in Norway: age-period-cohort analysis based on registry data on mammography screening and hormone treatment use. *BMJ* **344**: e299.

Welch HG (2010) Screening mammography–a long run for a short slide? *N Engl J Med* **363**(13): 1276–1278.

Welch HG. How much overdiagnosis? (2006) < http://www.bmj.com/rapid-response/2011/10/31/how-much-overdiagnosis > (accessed 14 October 2012)

Welch HG, 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**(11): 1023–1028.

White E, Miglioretti DL, Yankaskas BC, Geller BM, Rosenberg RD, Kerlikowske K, Saba L, Vacek PM, Carney PA, Buist DS, Oestreicher N, Barlow W, Ballard Barbash R, Taplin SH (2004) Biennial versus annual mammography and the risk of late-stage breast cancer. *J Natl Cancer Inst* **96**(24): 1832–1839.

Willis L, Alarcon T, Elia G, Jones JL, Wright NA, Tomlinson IP, Graham TA, Page KM (2010) Breast cancer dormancy can be maintained by small numbers of micrometastases. *Cancer Res* **70**(11): 4310–4317.

Yen AM, Duffy SW, Chen TH, Chen LS, Chiu SY, Fann JC, Wu WY, Su CW, Smith RA, Tabar L (2012) Long-term incidence of breast cancer by trial arm in one county of the Swedish Two-County Trial of mammographic screening. *Cancer* **118**(23): 5728–5732.

Zackrisson S, Janzon L, Manjer J, Andersson I (2007) Improved survival rate for women with interval breast cancer—results from the breast cancer screening programme in Malmo, Sweden 1976-1999. *J Med Screen* **14**(3): 138–143.

Zahl PH, Gøtzsche PC, Maehlen J (2011) Natural history of breast cancers detected in the Swedish mammography screening programme: a cohort study. *Lancet Oncol* **12**(12): 1118–1124.

Zahl PH, Maehlen J (2012) Overdiagnosis of breast cancer after 14 years of mammography screening. *Tidsskr Nor Laegeforen* **132**(4): 414–417.

Zahl PH, Maehlen J, Welch HG (2008) The natural history of invasive breast cancers detected by screening mammography. *Arch Intern Med* **168**(21): 2311–2316.

Zahl PH, Strand BH, Maehlen J (2004) Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. *BMJ* **328**(7445): 921–924.

Zbuk K, Anand SS (2012) Declining incidence of breast cancer after decreased use of hormone-replacement therapy: magnitude and time lags in different countries. *J Epidemiol Community Health* **66**(1): 1–7.