# Quantifying Breast Cancer Over-diagnosis in an Organised Mammography Screening Program

KERRI BECKMANN

BSc (Hons) MPH

September 2014

Discipline of Public Health
School of Population Health
Faculty of Health Sciences
University of Adelaide
Australia

A thesis submitted for the degree of Doctor of Philosophy

# Contents

# List of Tables

# List of Figures

# Thesis Abstract

Mammography screening is effective in reducing breast cancer (BC) mortality; however there are widespread concerns that it may also lead to over-diagnosis, i.e. the detection of BC that would not have emerged clinically in a woman's lifetime had she not participated in screening. The extent of over-diagnosis due to mammography is contested, with published estimates varying from 0% to 54%. The principal aim of this research is to quantify the level of over-diagnosis of BC associated with population-based mammography screening in South Australia (SA).

The following questions are addressed: (1) Have BC incidence rates increased following the introduction of screening in SA and is the increase greater than expected based on projections of pre-screening trends?, (2) Has the prevalence of key breast cancer risk factors also increased?, (3) To what extent does hormone replacement therapy (HRT) use affect breast cancer risk and screening outcomes?, (4) Are there any differences in the underlying risk of BC among screening participants and non-participants?, and the central question (5) What is the level of over-diagnosis due to organised mammography screening in SA? Questions 2-4 relate to the potential for estimates of over-diagnosis to be confounded by risk factor differences/temporal changes. A review of previous studies of over-diagnosis due to mammography screening is included, which highlights methodological complexities relating to measurement of over-diagnosis and offers some explanations for why published estimates vary to such a great extent.

The first two questions were answered through descriptive analyses of BC incidence trends in SA from 1977-2009, as well as trends in the prevalence several key breast cancer risk factors collected via the SA Health Omnibus Surveys during 1991-2009 (alcohol use, body weight, HRT use) and Australian Bureau of Statistics (fertility rates and age at first birth). The effect of HRT on various screening outcomes (e.g. screen-detection rates, interval cancer rates, recall to assessment) was examined though multivariable Poisson regression modelling using individual person level data from BreastScreen SA, which included self-reported HRT use at the time of each screening episode. Differences in underlying risk of BC between screening participants and non-

participants were investigated using 2012 South Australian Health Omnibus Survey data.

Two different methods were used to quantify over-diagnosis. Method 1 used a case-control design to compare screening histories for women with and without BC. Odds ratios (OR) were determined across different time intervals after screening to allow for lead time effects and applied to background reference rates based on pre-screening incidence trends. Over-diagnosis estimates were obtained by comparing cumulative incidence with and without screening. Method 2 used a lead time modelling approach in which estimates of lead time duration and screening sensitivity, and screening participation data were used to adjust the background incidence rates (without screening). This was achieved by iteratively adding the number of cancers expected to be brought forward by screening each year, then subtracting this number from the pool of cancers in future years. Over-diagnosis was calculated by comparing the lead time adjusted cumulative incidence with the observed cumulative incidence.

Studies presented in this thesis demonstrate that: (1) screening led to an increase in breast cancer incidence that was sustained beyond what was expected, based on projection of pre-screening incidence, however age-specific patterns suggest changing prevalence of HRT use have also impacted on incidence trends, (2) the prevalence of key risk factors also increased over this period, potentially contributing to an increase in background incidence rates, (3) HRT use among South Australian women is causally associated with increased risk of breast cancer which complicates estimation of over-diagnosis due to the marked changes in prevalence of HRT use, (4) women who participated in screening had a higher prevalence of breast cancer risk factors (most notably HRT use), indicating the potential for estimates to be confounded by underlying risk differences, (5) mammography screening is likely to result in a modest level of over-diagnosis (8% for IBC and 12-14% for all BC among women eligible to participate in screening). Estimates were lower after adjustment for confounding. These results are comparable with findings from long-term follow-up of screening trials and with several recent cohort studies of European screening programs, but are lower than many other estimates.

# Thesis Declaration

This work contains no material that has been accepted for an award of any degree or diploma in any university or other tertiary institution and to the best of my belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent for this copy of my thesis, when deposited in the University Library being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

The author acknowledges that the copyright of the published works contained within this thesis (as listed on the next page) resides with the copyright holders of those works.

I also give permission for a digital version of this thesis to be made available on the web, via the University's digital research repository, the library catalogue and also through web engines, unless permission has been granted by the University to restrict access for a period of time.

Signed ……..

Kerri Beckmann (Candidate)

Date ……..

# Publications/manuscripts contributing to this thesis

**Beckmann KR**, Farshid G, Roder DM, Hiller JE and Lynch JW. Impact of hormone replacement therapy use on mammographic screening outcomes. *Cancer Causes Control*. 2013; 24: 1417-26.

**Beckmann KR**, Roder DM, Hiller JE, Farshid G, Lynch JW. Do breast cancer risk factors differ among those who do and do not undertake mammography screening? *Journal of Medical Screening* 2013; 20:208-19.

**Beckmann KR**, Roder DM, Hiller JE, Farshid G, Lynch JW. The influence of mammographic screening on breast cancer incidence trends in South Australia. *Asian Pacific Journal of Cancer Prevention* 2014; 15:3105-12.

**Beckmann KR**, Lynch JW, Hiller JE, Farshid G, Houssami N, Duffy Sw, Roder DM. A novel case-control design to estimate the extent of over-diagnosis of breast cancer due to organised mammography screening. *International Journal of Cancer* 2014. Aug 5. doi: 10.1002/ijc.29124. [Epub ahead of print]

**Beckmann K**, Duffy SW, Lynch JW, Hiller JE, Farshid G, Roder DM. Estimates of over-diagnosis of breast cancer due to population-based mammography screening in South Australia after adjustment for lead time effects. [Submitted to *Journal of Medical Screening*]

# Presentations arising from this thesis

**Beckmann KR.** Over-diagnosis of breast cancer. Higher Degree Research Student Seminar Series, School of Population Health, Adelaide, Australia, August 2011. (Oral presentation)

**Beckmann KR**, Farshid G, Roder DM, Hiller JE, Lynch JW. Impact of hormone replacement therapy use on mammographic screening outcomes. 2012 Faculty of Health Sciences Postgraduate Conference. University of Adelaide, Adelaide, Australia, August 2012. (Poster presentation)

**Beckmann KR**, Farshid G, Roder DM, Hiller JE, Lynch JW Impact of hormone replacement therapy use on mammographic screening outcomes. Sydney International Breast Cancer Congress 2012, Sydney, Australia, October 2012. (E-poster)

**Beckmann KR**, Roder DM, Hiller JE, Lynch JW. Over-diagnosis of breast cancer due to organised mammography screening. Centre for Cancer Prevention, Wolfson Institute, Queen Mary University of London, London, UK, October 2013. (Invited oral presentation)

**Beckmann KR**. Over-diagnosis of breast cancer through mammography screening: Fact or Fiction? Higher Degree Research Student Seminar Series, School of Population Health, Adelaide, Australia, October 2013. (Oral presentation)

**Beckmann KR**, Hiller JE, Lynch JW, Houssami N, Farshid G, Roder DM. Over-diagnosis due to mammography screening programs: Estimates from South Australia using two different methods. Breast Screen Australia Conference, Melbourne, October 2014 (Oral presentation)

**Beckmann KR**. Does mammography screening lead to over-diagnosis of breast cancer? Research Seminar Series, School of Population Health, Adelaide, Australia, October 2014. (Oral presentation)

**Beckmann KR**, Hiller JE, Lynch JW, Farshid G, Houssami N, Roder DM. Over-diagnosis due to mammography screening programs: Evidence from South Australia World Cancer Congress, Melbourne, December 2014 (E-poster presentation)

# Acknowledgements

I would like to sincerely thank all who have helped me in undertaking this thesis, and supported me throughout this journey.

A very special thank you goes to my supervisors, David Roder, Janet Hiller and John Lynch. John is an inspirational leader and mentor who instils in those around him the passion to continually learn and grow, and I have gained more than I imagined from being part of his team. I am grateful for the support and advice that everyone in his group has generously offered me. Janet continually provided invaluable support and encouragement throughout my candidature. I always felt I could rely on her advice to help when I felt stuck and keep me focussed and on track. I am especially thankful to David for giving me the opportunity to tackle such an important topic and for believing in my capacity to undertake this challenging research successfully. It has been invaluable to have had a supervisor that was so passionate and interested in the research question, with whom I could discuss any and all issues, from detailed study design through to broader policy implications.

Our research team received financial supported from the National Breast Cancer Foundation, through their Novel Concept Award, and we are grateful for their decision to support research into this important question.

It would not have been possible to have undertaken this work had it not been for the confidence and trust placed in me and my supervisory team by the Manager of BreastScreen SA, Lou Williams. It was extremely courageous for BreastScreen to be open to us addressing the question of over-diagnosis and to facilitate the research process, given the potential impact the findings may have on the screening program. I also am extremely grateful to Gelareh Farshid, the Clinical Director of BreastScreen SA, for taking great interest in my thesis and providing such valuable feedback, particularly in relation to the interpretation and clinical implications of the research findings.

A big thank you goes to Gillian Rogers for helping me to understand the ins and outs of the Breast Screen SA's record systems and for ensuring we received the highest quality data from the screening service. Special mention must also go to Minh Nguyen, from the Epidemiology Branch of SA Health, for the scores of hours she put into undertaking

the enormous task of record linkage (not once but twice) which was required to compile the datasets needed for studies described in this thesis. I would also like to thank Tommy Tucker from the Epidemiology Branch who spent many hours ensuring he understood the exact nature of our data needs along with the numerous and complex linkage processes required.

I am most appreciative of the generosity of staff at the Centre for Cancer Prevention at the Wolfson Institute, Queen Mary University of London, who hosted me for four weeks during my final year. Learning from some of the most eminent researchers in the field such as Stephen Duffy and Sue Moss was both a privilege and a pleasure.

Many thanks go to my fellow PhD buddies for their support, encouragement and camaraderie along the way. Their willingness to listen, make suggestions and offer kind words of encouragement when things were not going according to plan have helped me push through those occasional periods of self-doubt and frustration. I have developed some lovely friendships during this period which I will cherish for many years to come.

Lastly, I could not have achieved this goal without the love and support of my family and special friends, who all believed in my ability to successfully complete my PhD. I thank them dearly.

## ABBREVIATIONS

| | |
|---|---|
| ABS: | Australian Bureau of Statistics |
| APC: | Annual percent change |
| ARIA: | Accessibility and Remoteness Index for Australia |
| ASR: | Age-standardised rate |
| BC: | Breast cancer |
| BMI: | Body mass index |
| BSSA: | BreastScreen SA |
| CI: | Confidence interval |
| CISNET: | Cancer Intervention and Surveillance Modelling Network |
| DCIS: | Ductal carcinoma in-situ |
| ER: | Electoral roll |
| ERP: | Estimated residential population |
| FNA: | Fine needle aspiration |
| GP: | General Practitioner |
| HRT: | Hormone replacement therapy |
| IARC: | International Agency for Research on Cancer |
| IBC: | Invasive breast cancer |
| IDC: | Invasive ductal cancer |
| IRR: | Incidence rate ratio |
| IRSAD: | Index of Socioeconomic Advantage and Disadvantage |
| MET: | Metabolic equivalent time |
| MI: | Multiply imputed |
| MST: | Mean sojourn time |
| NSW: | New South Wales |
| OD: | Over-diagnosis |
| OR: | Odds ratio |
| PCNB: | Percutaneous needle biopsy |
| PPV: | Positive predictive value |
| RCT: | Randomised controlled trial |
| RR: | Relative risk |
| SA: | South Australia |
| SACR: | South Australian Cancer Registry |
| SEP: | Socioeconomic position |
| UK: | United Kingdom |
| US: | United States of America |
| WHI: | Women's Health Initiative |

# Chapter 1: Introduction

## BACKGROUND

Over the past decade, concerns have been raised about the potential for mammography screening to lead to over-diagnosis of breast cancers, i.e. the diagnosis of breast cancer that would not have been detected during a woman's lifetime had she not participated in breast screening[1]. The extent to which over-diagnosis occurs through mammography screening has been hotly contested, with estimates of over-diagnosis ranging from none or negligible to one in three breast cancer cases diagnosed through screening. Variability in the methods used to measure and report the extent of over-diagnosis, as well as the potential for multiple sources of bias due to the reliance on ecological study designs, has contributed to this inconsistency[2, 3].

The potential for harm through over-detecting breast cancer and hence over-treating women who participate in mammography screening has serious implications for well-established, publicly-funded screening programs that are operating across many western countries, including Australia. More reliable study designs are required to equip service providers with evidence to make future policy decisions about mammography services and to allow the millions of women who choose to undergo mammography screening to be informed about associated risks.

## THE EPIDEMIOLOGY AND NATURAL HISTORY OF BREAST CANCER

Breast cancer is the most commonly diagnosed cancer (with the exception of skin cancers) and the leading cause of cancer death among women in Australia[4], and across the world[5]. Even though breast cancer incidence is higher in high-income countries, it is also the leading female cancer in many low and middle-income countries as well. Global patterns reflect differences in the prevalence of known risk factors for the disease – for example early age of menarche, later age at menopause, later age at first birth, lower parity, higher exogenous hormone exposure, obesity and reduced physical activity[5]. Over the past few decades incidence rates have been increasing in most regions of the world including in both developing and developed countries, with the

most rapid rises observed in developing countries. Such trends are attributed to the adoption of a more "western" lifestyle in developing countries and to changing patterns of childbearing, obesity, hormone use and other reproductive factors in developed countries. The introduction of mammography screening (discussed below) is also believed to have contributed to an increase in cancer incidence[6, 7]. In contrast, stable or decreasing trends in breast cancer mortality have been observed in most developed countries, and are likely to be the result of earlier detection through screening and improvements in treatment[8].

Australia has a relatively high rate of breast cancer by world standards[9]. Approximately 1,100 women are diagnosed with invasive breast cancer each year in South Australia and 12,500 are diagnosed across the whole of Australia, while 250 women die from breast cancer each year in South Australia and 2700 women die nationally[4, 10]. Breast cancer incidence rates (for Australia as a whole) increased steadily between 1982 and 1995 but appeared to stabilise from 1996 onwards. In contrast, breast cancer death rates were relatively stable between 1982 and 1994, but have steadily declined since 1995. The reduction in breast cancer moratlity between 1995 and 2007 was around 30%[4].

Prognosis for women diagnosed with breast cancer in Australia is generally very good, with 88% of women surviving five years or more after their diagnosis[11]. However, survival is strongly associated with the stage of disease at diagnosis. Over 93% of Australian women diagnosed with early breast cancer (small tumours with no nodal involvement) survive for five or more years after diagnosis, whereas those with more advanced disease have poorer outcomes – less than 50% if the cancer has spread to distant sites[12].

Despite extensive research into the various aspects of the disease process, the natural history of breast cancer is not well understood. This lack of understanding stems from the difficulty observing disease progression over time, since tumours are rarely left untreated once detected[13]. The predominant view has been that breast cancer develops though a multistep process with increasing genetic alterations occurring along the pathway. Early models proposed a linear transition from normal breast epithelial cells to atypical hyperplasia to carcinoma in-situ (proliferating malignant cells

that have not penetrated the basement membrane of mammary ducts/lobes) to invasive carcinoma[14]. Evidence from epidemiologic, pathologic and genetic studies all support the notion that ductal carcinoma in-situ (DCIS) is a precursor to invasive ductal cancer (IDC), the most common form of invasive breast cancer[15], and, while less extensively studied, lobular carcinoma in-situ is likely to be a direct precursor for invasive lobular carcinoma, the next most commonly diagnosed form of breast cancer[16].

Recent evidence from genetic studies suggests that this linear model is too simplistic[14-17]. Genetic profiling of pre-invasive and invasive breast cancer indicate two distinct pathways to invasive disease. Similarities in the genetic signatures of pre-invasive and invasive ductal cancer suggest these pathways diverge at a very early stage of tumour development and that these early genetic changes pre-determine the propensity for how aggressively the tumour progresses. According to this model, low grade DCIS progresses to low grade IDC and high grade DCIS progresses to high grade IDC (which is a much more aggressive form of breast cancer). Histological studies showing high correlation between the morphologic features of DCIS and IDC when they coexist in the same lesion provide further evidence of dual pathways[14-16].

Progression from pre-invasive to invasive cancer is believed to occur at very different rates for each of these pathways (taking decades in the case of low grade DCIS and possibly as little as 5 years for high grade DCIS[15]). Follow-up studies of patients who only received a diagnostic biopsy indicate that up to 50% of DCIS lesions will progress to invasive disease if left untreated[13]. Autopsy studies of women indicate a moderate level of undiagnosed DCIS and low level of undiagnosed invasive cancer among women who had not been diagnosed with breast cancer (pooled estimates of 9% and 1% respectively)[18]. These observations also support the concept of heterogeneity in breast cancer growth rates.

There is also uncertainty surrounding the mediating processes and timing of the transition from invasive to metastatic disease. Ninety percent of deaths from solid cancers result from metastases - the spread of the primary cancer to distant organ sites, where the tumour is able to grow and disrupt physiological functioning[19]. In the case of breast cancer, the process of metastases can go undetected, with distant

metastatic lesions only coming to light many years after treatment of the primary tumour, suggesting a period of tumour cell dormancy[20]. Genetic signatures associated with the transition to metastatic disease, suggest that this pathway is distinct from the transition from pre-invasive to invasive disease[19]. Certain subtypes of breast cancer are associated with higher risk of recurrence and metastases, but whether this reflects different growth rates or dormancy remains unclear[20].

Currently none of the proposed classification systems for DCIS consistently predict disease progression[14]. Nor have any molecular biomarkers been identified that accurately predict clinical behaviour of invasive cancers[16]. Hence it is not possible, as yet, to distinguish the more indolent cancers from the more aggressive cancers at the time of diagnosis.

This heterogeneity in tumour types and disease progression rates poses challenges in relation to the detection and treatment of pre-invasive and (early stage) invasive breast cancer. Mammography has enabled the detection of quite small in-situ lesions, whereas prior to the advent of screening, DCIS lesions were rarely detected except where palpable. Consequently the incidence of ductal carcinoma in-situ has risen markedly across many countries since the introduction of screening[21], with a 700% increase reported for South Australia [22].

While most pre-invasive cases are treated in a similar manner to early stage invasive disease (surgical removal with radiotherapy and hormonal therapy where appropriate), appropriate management of DCIS remains a contested issue[23]. The rationale for treating pre-invasive in-situ cancers is to prevent the development of invasive disease. However, it is likely that there is considerable overtreatment of DCIS, since not all tumours will progress to invasive disease during a woman's lifetime[18]. Nor is it certain whether all invasive cancers progress to be clinically significant.

## MAMMOGRAPHY SCREENING TO REDUCE BREAST CANCER MORTALITY

Opportunities for primary prevention of breast cancer are limited because many of the risk factors for breast cancer are difficult to modify (particularly those relating to reproductive factors)[8]. Early detection has therefore been the main preventive

strategy employed to reduce mortality and morbidity associated with breast cancer, across most countries with well-resourced health care sectors[24].

The aim of screening mammography is to reduce breast cancer mortality through detection of cancers at an earlier stage, when they are more amenable to treatment. Mammographic screening was specifically developed as a radiographic method for viewing abnormalities in breast tissue. It is based on differential absorption of X-rays across different types of tissue. Areas of calcification (which are associated with in-situ cancers) or dense cellular masses (e.g. tumours) will be highlighted on X-ray film because of impedance of the X-ray beam, whereas areas of fatty tissue will remain dark on the film[24]. Women with suspicious lesions on X-ray film would be recalled for further investigation using other diagnostic techniques to establish whether the identified abnormality was in fact breast cancer. Mammography therefore allows for detection of very small tumours, that would otherwise not have been clinically detectable (i.e. palpable), as well as pre-invasive forms of the disease such as DCIS.

Screening mammography has been extensively evaluated as a tool for early detection of breast cancer, originally through randomised trials[25-32], and more recently through observational studies of population-based screening programs[33, 34]. Seven of the eight original randomised trials (which recruited women from the 1960's through to 1980's) have shown a reduction in mortality from breast cancer among women invited to participant in screening. Most, but not all, of the subsequent meta-analyses of the data from these trials have supported the efficacy of mammographic screening[24, 35-38]. The one dissenting review was that of the Cochrane Collaboration (published in 2001) which concluded that 'screening for breast cancer with mammography was unjustified'[37]. The authors of this review excluded six of eight previous trials, judged to have significant flaws, from their analysis. Their conclusion (that there was no reduction in mortality) was based on the findings of the two remaining trials. Several peak health bodies, including the International Agency for Research on Cancer (IARC)[24], the US Preventive Services Task Force[38] and the Independent UK Review Panel on Breast Cancer Screening[39] have subsequently re-examined the evidence from the original randomised trials. The consensus among these expert review panels is that screening mammography reduces the risk of breast cancer mortality for women aged

50-69 years. (IARC have estimated the reduction in this age group to be 25% among women invited to screen and around 35% among those who actually attended screening[24].) Evidence for reduced mortality is less clear for women aged 40-49 years and inclusion of this age group in screening programs remains a controversial issue[40-42]. The effectiveness of established screening programs has also been evaluated across a number of countries (along with Australia) using observational studies including case-control, cohort and non-randomised trial designs[43-49]. Despite the diversity in study methods and variations in the implementation of different screening programs, these studies have shown a consistent trend toward reduced breast cancer mortality due to screening. Reductions are similar to or better than those observed in trials, with reduced mortality ranging from 16-36% among women invited to participate and 24-48% among those who actually attended screening[50].

Australia introduced a nation-wide publicly-funded breast cancer screening program (operated and administered through State Health Departments) in the early 1990s, following positive findings from the international trials of mammography screening. The South Australian screening program (BreastScreen SA) began as a pilot program in 1989, and became fully operational from 1991. It has now been operating for over 20 years[51]. BreastScreen SA invites women to participate in biennial screening from 50 through to 69 years of age. While women aged 40-49 years and 70 years and older are also eligible to attend the service, they are not actively recruited. (The upper screening target age has now been extended to 74 years, however this only became effective from 2013 so is not applicable to the study period for the body of work reported in this thesis.) As of June 2011, BreastScreen SA had provided over 1,200,000 rounds of mammography to approximately 300,000 South Australian women (Personal communication, BreastScreen SA). Mammography screening is provided through both fixed and mobile units covering the whole of South Australia and involves two-view radiology of both breasts, as well as double reading of X-ray films. Digital technology was only introduced very recently. BreastScreen SA has successfully achieved national accreditation throughout its operation[51].

Mammography screening programs have also been introduced in over 20 other countries (including many European countries, Canada, Japan, and Uruguay). However

there is considerable variation in relation to the target age groups, methods of invitation, screening intervals, participation rates, coverage (regional or national), whether clinical examinations are also undertaken, the number of views taken of each breast and whether double reading is employed[24]. Variation in the way mammography programs have been implemented, along with differences in population characteristics across different countries, has implications for the generalisability of research findings between one setting and another, particularly in relation to the balance between benefits and potential harms.

## WHAT IS OVER-DIAGNOSIS AND WHY SHOULD WE BE CONCERNED ABOUT IT?

Over-diagnosis of cancer is defined as "the diagnosis of cancer as a result of screening that would not have arisen clinically in a person's lifetime had screening not taken place"[52]. While sometimes referred to as 'pseudo-cancers', over-diagnosed cancers are not the same as false-positive cases arising from screening, in that they have been histologically confirmed and classified as cancerous (rather than benign) tumours.

Screening programs aim to detect cancer at an earlier stage when the tumour is more amenable to treatment and/or when more conservative treatment is possible. Effective screening programs depend on there being a preclinical disease state that can be identified through the screening procedure, which precedes the symptomatic disease state. The period from the time that screening detected the preclinical lesion to the onset of clinical symptoms is known as the lead time. Lead time reflects how much earlier cancers are diagnosed within a screening program compared with when they would have been diagnosed through presentation of symptoms. Assuming screening detects cancers that would have presented 'naturally' over the next few years, screening programs will result in higher rates of diagnosis initially, which should be compensated for by lower rates in the period following cessation of screening. Theoretically, after the 'lead time' has passed, cancer incidence rates should return to a similar level to that expected had screening not occurred[53]. If the initially high incidence is not balanced by lower incidence (over a follow-up period sufficiently long enough to cover the lead time) then screening may have resulted in over-diagnosis.

In any screening program some over-diagnosis is likely to be inevitable simply because of competing causes of death (people die from other causes before the cancer would have become symptomatic). However, it is also possible that screening detects some cancers that are very slow growing or 'indolent' which would not have become clinically significant during a person's lifetime, and hence would not have resulted in any health consequences or required treatment. This implies that some subtypes of screen-detected cancers do not progress to symptomatic disease within the average life span. Alternatively, as proposed by some researchers, it may be possible that screening detects a proportion of cancers that 'spontaneously' regress[54], though this concept is difficult to test since the process of regression cannot be observed in any 'natural way' in human breast tissue.

Over-diagnosis of cancer is considered to be a potential harmful effect of screening, primarily because it leads to unnecessary (often quite invasive) treatment and increased physical and psychological morbidity associated with a cancer diagnosis[55]. However, despite the considerable research efforts being directed toward identifying biomarkers that predict prognosis, currently for most types of cancer it is not possible to distinguish cancers with little or no potential to progress from those that will be life threatening. As a consequence most screen-detected cancers will be treated as if they were life threatening, including the proportion which may be over-diagnosed[55].

Given that there is currently no way of identifying specifically which cancers are over-diagnosed and which are not, over-diagnosis has been described predominantly as an 'epidemiological concept'[56] which can only be measured at a population level. Even so, concerns about over-diagnosis of cancers (and consequent overtreatment) need to be addressed through rigorous research, to ensure that policy decisions about screening services are evidence-based, and that information to consumers about potential risks is accurate.

## ESTIMATES OF OVER-DIAGNOSIS DUE TO MAMMOGRAPHY

Estimates of the percentage of breast cancers 'over-diagnosed' among women in mammography screening programs vary widely, with reported estimates ranging from 0% to 50% or more of diagnosed cancers, depending on the way measures are expressed[3, 57]. Differences in methodologies and assumptions used to derive various

estimates have also contributed to this variability. Methodological factors include: which populations are being assessed (e.g. populations where screening is offered, people invited for screening, screening participants or screen-detected cases); what adjustments are made for other risk factors and how these are determined (e.g. using ecological or person-level data); whether in-situ cancer (e.g. DCIS) is included in the analyses; the length of time women are followed after screening cessation; estimates of lead time effects; differences in the protocols followed within the screening programs/trials which have been examined; and differences in the way results are reported, particularly in relation to what denominator population is used[3]. Due to the variability in methodology and results, a consensus about the extent of over-diagnosis has not been reached[58].

Assessment of over-diagnosis is complicated by changes in the prevalence of risk factors for breast cancer during and after the introduction of breast screening programs, including increasing bodyweight, earlier age of menarche, later age of menopause, increasing age at first birth, decreasing number of children and increasing use of hormone replacement therapy (HRT)[2, 59]. Among these factors, the use of HRT is one risk factor that has changed substantially over the past 20-30 years[60]. During the late 1980's and 1990's use of HRT for the relief of menopausal symptoms steadily increased in a number of developed countries. HRT use then declined abruptly following publication of results of the Women's Health Initiative trial of combined HRT in 2003, which indicated increased risk of breast cancer (among other adverse outcomes) [61]. Australian women were among the highest users of HRT and were also quick to discontinue use following the publication of the trial findings[62, 63]. According to a number of researchers, the changing patterns of HRT use have had a marked impact on breast cancer incidence rates[60, 64-68]. However, as yet there is little published research which directly examines the impact of HRT use when estimating over-diagnosis in mammography screening programs.

Generally methods for estimating over-diagnosis have involved either follow-up studies of the original randomised trials of mammography screening[69-71] or ecological study designs which have compared breast cancer incidence in screened and unscreened populations[72-75] or in populations before and after screening was

introduced[54, 69, 76-82]. The former approach can yield results with limited generalisability, since findings are based on outdated mammographic technologies and procedures. The latter approach is limited by lack of data at the individual person level and is dependent on debatable assumptions about temporal trends and/or assumptions of equivalence of risk factor profiles in comparison populations. More recently several cohort studies using individual level data on screening exposure and breast cancer outcomes have been published which are likely to provide more reliable estimates of over-diagnosis[83-85]. However, confounding due to different levels of risk for breast cancer between attendees and non-attendees is a potential source of bias in these studies. In addition to experimental and observational study designs, several researchers have developed micro-simulation multistate models in which parameters derived from trial data or from screening programs (e.g., screening sensitivity, lead time, screening participation rates, tumour progression rates) are applied to predict cancer outcomes in the presence or absence of screening[3, 52, 86-88]. Again these methods rely on questionable assumptions, for example, constant tumour progression rates.

To date, only two studies have examined over-diagnosis of breast cancer in the context of Australian mammography screening programs[77, 82]. Both used ecological designs based on breast cancer incidence trend data for the state of New South Wales (NSW). Each found relatively higher levels of over-diagnosis (53% including invasive and DCIS in the first study[82] and 30-42% for invasive breast cancers in the second study[77]).

Rather than providing a more definitive answer, recent reviews of breast cancer over-diagnosis have served to further highlight the extreme variation in methodological approaches and reported estimates of previous studies[2, 3, 39, 57, 89]. The US Preventive Services Task Force concluded that estimates could not be pooled or summarised due to the considerable variation in methodologies[90]. A number of these reviews have called for more rigorously designed studies to be undertaken to provide reliable estimates of breast cancer over-diagnosis due to mammography.

Chapter 5 provides a more detailed appraisal of the various methodological approaches used to estimate the extent of over-diagnosis due to mammography screening.

## SIGNIFICANCE

Without reliable and consistent evidence about the extent of over-diagnosis due to mammography screening, consensus on the scale of over-diagnosis has not been achieved, despite vigorous debate.

Negative publicity about mammography screening has the potential to undermine women's confidence in publicly-funded breast screening programs, despite proven benefits[91]. Such publicity can also trigger service providers to re-examine the value of providing screening programs. The issue of over-diagnosis of breast cancer is increasingly being presented in the research literature and in the public media as a significant (and proven) harm arising from mammography screening programs, despite considerable uncertainty about the degree to which it is problematic.

Even so, it is important that women are well informed about the risks, as well as the benefits, of mammography, including the potential for over-diagnosis. Hence there is a pressing need to undertake a range of novel study designs, including studies using person-level data, to complement the ecological data on over-diagnosis in contemporary mammography screening programs to better understand the extent of the over-diagnosis arising from mammography screening. Furthermore, since screening protocols and practices vary considerably across mammography programs, it is important that locally relevant data are available for making policy decisions in the Australian context.

## RESEARCH AIMS AND QUESTIONS

The aim of the research outlined in this thesis is to quantify the level of over-diagnosis of breast cancer associated with population-based mammography screening in South Australian.

The primary research questions are:

- What impact has population-based mammography screening had on breast cancer incidence patterns and trends in South Australia?
- Have there been any significant changes in the prevalence of breast cancer risk factors that may have also contributed to changes in breast cancer incidence?

- What impact has hormone replacement therapy (HRT) use had on breast cancer screening outcomes (i.e. screening detection rates, interval cancer rates) in South Australia?
- Do women who participate in screening have a higher risk of breast cancer than those who do not?
- To what extent has the introduction of an organised, publicly-funded breast screening program led to 'over-diagnosis' of breast cancer in South Australia (i.e. the diagnosis of breast cancer cases that would not have presented clinically during a woman's lifetime had she women not participated in screening)?

## THESIS OUTLINE

Chapter 2 (Publication 1) addresses the first two questions through descriptive analyses of breast cancer incidence rates in South Australia prior to and after the introduction of the breast screening program in relation to BreastScreen SA participation rates and trends in the prevalence of key breast cancer risk factors (body weight, alcohol use, HRT use and fertility rates). While predominantly descriptive in nature, the findings of this chapter will establish whether breast cancer incidence has increased as a result of the introduction of a population-based mammography screening program, beyond what would have been predicted had screening not been implemented. It also explores whether any other factors could possibly explain observed patterns of breast cancer incidence in South Australia.

Chapter 3 (Publication 2) examines the impact of HRT use on screening outcomes in South Australia. Changing patterns of HRT use are likely to have impacted on breast cancer risk and screening outcomes (i.e. screen-detection rates, interval cancers) and thus may have independently contributed to higher than expected breast cancer incidence, in parallel with screening. Previous research indicates a positive association between HRT use and increased risk of breast cancer. However, the effects of HRT can vary by the type and duration of use, and detailed information on specific patterns of use in South Australia over the past two to three decades is scant. It is therefore important to establish whether HRT, as used by South Australian women, is actually associated with increased risk of breast cancer. HRT use has been recorded for women

attending BreastScreen, providing a reliable and consistent source of data to examine the effects of HRT in the South Australian population. This study specifically investigates the impact of 'current' HRT use on rates of screen-detected breast cancers, interval cancers, recall for assessment and false-positive rates among women attending BreastScreen SA. Findings will be directly applied in later work relating to modelling lead time effects due to mammography screening, which is one of the approaches used to measure over-diagnosis (detailed in Chapter 7).

Chapter 4 (Publication 3) addresses the question of whether the risk of breast cancer is equivalent for women who do and don't attend publically funded mammography screening. Differences in the background level of risk can potentially bias estimates of over-diagnosis in studies that compare breast cancer incidence among women who have and have not participated in screening. The study presented in Chapter 4 examines the differences in the prevalence of key breast cancer risk factors among those who had ever and never attended BreastScreen for a mammogram using cross-sectional population-based survey data for SA women aged 40 to 84 years. Overall levels of risk are compared using risk prediction models. These findings will inform the sensitivity analyses undertaken to investigate the potential confounding effect of risk differences on estimates of over-diagnosis derived from the case-control study presented in Chapter 6. Collectively, Chapters 2 to 4 provide important contextual information for interpreting results from the final two studies examining the extent of over-diagnosis.

Chapter 5 reviews previously published studies that have attempted to measure over-diagnosis due to mammography screening. An important aim of this chapter is to better understand why the findings of previous studies have varied so widely and which among these may provide more reliable estimates. Methodological complexities associated with the various approaches used to estimate over-diagnosis are discussed and the findings from previous studies critically appraised.

Chapter 6 (Publication 4) and Chapter 7 (manuscript under review) are both directed at estimating the extent of breast cancer over-diagnosed due to mammography screening in South Australia. Each uses a different methodological approach. The first study uses a case-control design, with individual level measures of screening exposure

and breast cancer occurrence. Sensitivity analyses include adjustment for risk differences between those who have and have not attended screening based on the findings presented in Chapter 4. The second study uses a lead time modelling approach whereby the background incidence in the absence of screening is adjusted for lead time effects and then compared with the observed incidence. The model makes adjustments based on the number of women undergoing breast screening by age and year, using estimates of lead time length and screening sensitivity derived from South Australian data on incidence and interval cancer rates.

Chapter 8 will provide a summary of the results of this work, along with discussion of the limitations and difficulties related to measuring the extent of over-diagnosis measuring and concluding remarks.

# Chapter 2: The influence of mammographic screening on breast cancer incidence trends in South Australia

Paper published in Asian Pacific Journal of Cancer Prevention 2014; 15: 3015-22.
Authors: Kerri Beckmann, David Roder, Janet Hiller, Gelarah Farshid, John Lynch

## PREFACE

Concerns about over-diagnosis of breast cancer are, in part, due to the marked increases in breast cancer incidence rates that have coincided with the implementation of organised mammography screening programs in many countries where they have been introduced. In many cases, rates have remained higher than the expected background incidence even after several decades of screening, which has been taken by some to indicate that screening results in over-diagnosis. Whether there has been any persistent increase in breast cancer incidence in South Australia, at what points in time changes occurred and what factors may have contributed, is largely unknown.

The following study examines trends in breast cancer incidence over the past three decades among SA women aged 40-84 years, to determine whether incidence rates for breast cancer have increased beyond what would have been predicted based on projections of pre-screening incidence trends. It also aims to identify the specific time points when breast cancer incidence trends began to increase, stabilise and, or decrease and qualitatively relate these changes to changing patterns of participation in mammography screening. Trends in the prevalence of breast cancer risk factors are also examined to determine whether there were any significant changes in key risk factors that might also have influenced incidence trends. The discussion section of this paper provides a narrative commentary on factors that have potentially influenced incidence trends in South Australia over the past two to three decades based on the descriptive analysis of trends.

## AUTHORSHIP STATEMENT

**Influence of mammographic screening on breast cancer incidence trends in South Australia.**
*Asian Pacific Journal of Cancer Prevention* 2014; 15: 3105-12.

**Kerri Beckmann (Candidate)**
I conceived and designed the study, managed data collation, analysed and interpreted the data, drafted and critically revised the manuscript, coordinated contributions from co-authors and handled revisions prior to print.

25/09/14
Signed …… Date……

**David Roder**
My contribution to this publication involved advising on research design and methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

22/09/14
Signed …… Date……

**Janet Hiller**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**Gelareh Farshid**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**John Lynch**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

23/09/14
Signed …… Date……

Beckmann, K., Roder, D., Hiller, J., Farshid, G. & Lynch, J. (2014) Influence of mammographic screening on breast cancer incidence trends in South Australia.
*Asian Pacific Journal of Cancer Prevention, v. 15(7), pp. 3105-3112*

# Chapter 3: Impact of hormone replacement therapy use on mammographic screening outcomes

Paper published in Cancer Causes and Control 2013; 24: 1417-26.

Authors: Kerri Beckmann, Gelareh Farshid, David Roder, Janet Hiller, John Lynch

## PREFACE

Results of the previous chapter indicate that, aside from the uptake of mammography screening, other factors are likely to have impacted on breast cancer incidence trends, with the most obvious candidate being changing patterns of HRT use. However, because the previous paper was essentially descriptive, it does not offer any direct evidence of a causal link between HRT use and increasing or decreasing incidence. Also, since the level of risk associated with HRT varies according to the type of hormone therapy and duration of use, it is difficult to assess the extent to which changes in incidence can be attributed to HRT use in the local context, without detailed information at a population level about hormone therapy modalities most frequently used. The following study examines the effect of HRT use on mammography screening outcomes among South Australian women attending BreastScreen SA. The study considers the impact on rates of screen-detected cancers, interval cancers, recall and biopsy rates, in order to determine whether the high prevalence of HRT use in South Australia led to increases in breast cancer incidence, independently of screening effects, and to shed light on pathways through which this may have occurred. Any effect of HRT use on interval cancer rates will impact on estimates of screening lead times. Accurate estimates of mammography lead time are fundamental to modelling the effects of screening in the population, which is the underpinning methodology used in the second approach applied in this thesis to estimate over-diagnosis due to mammography screening (presented in Chapter 7).

## AUTHORSHIP STATEMENT

**Impact of hormone replacement therapy use on mammographic screening outcomes.**

*Cancer Causes and Control* 2013;24:1417-26.

**Kerri Beckmann (Candidate)**
I conceived and designed the study, managed data collation, analysed and interpreted the data, drafted and critically revised the manuscript, coordinated contributions from co-authors and acted as corresponding author for peer review and preparation for publication.

25/09/14
Signed …… Date……

**Gelareh Farshid**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**David Roder**
My contribution to this publication involved advising on research design and methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

22/09/14
Signed …… Date……

**Janet Hiller**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**John Lynch**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

23/09/14
Signed …… Date……

## ABSTRACT

**Purpose**: This study aims to measure the impact of hormone replacement therapy (HRT) use at the time of screening on rates of screen-detected invasive breast cancer (IBC) and ductal carcinoma in-situ (DCIS), interval cancers and investigative procedures, within a well-established population-based mammography screening program.

**Methods**: Using South Australian BreastScreen data from 1998-2009 pertaining to 819722 screening episodes, Poisson regression models were undertaken to estimate incidence risk ratios (IRR) for various screening outcomes at both the first and subsequent screening rounds, among women who had been using HRT in the 6 months prior to screening compared with those who had not.

**Results:** Current HRT use was associated with increased risk of recall for assessment, biopsy procedures and breast cancer diagnosis among BreastScreen participants. Risk of screen-detected breast cancer was increased at subsequent screening rounds [IRR=1.30, 95% confidence interval 1.18-1.34] but not at women's first screening round [1.05, 0.88-1.25]. This increased risk applied to IBC [1.35, 1.27-1.45] but not DCIS [1.04, 0.89-1.23]. Interval cancer risk was elevated among HRT users following both the first screen [1.77, 1.33-2.37] and subsequent screening episodes [1.92, 1.72-2.15].

**Conclusions**: Increased risks of recall, biopsy rates, screen-detected and interval cancers among HRT users have important implications for population-based breast cancer screening programs. Our findings support the concept that HRT use may increase the growth of pre-existing cancers. Lack of effect on DCIS could imply different aetiology or time frames for DCIS and IBC development or increased transition from pre-invasive to invasive disease due to HRT use.

## INTRODUCTION

Prior studies have described an increased risk of breast cancer among hormone replacement therapy (HRT) users[102, 121-128], and have indicated decreased sensitivity and specificity of mammography screening with current HRT use[127-135]. With the exception of studies focusing on screening accuracy there has been little research on the impact of HRT use within mammography screening programs, and less still looking at effects by screening round (first/prevalent verses subsequent/incident screening rounds) or by age group. Uncertainty also remains about whether HRT use is associated with increased risk of in-situ (pre-invasive) cancers[136, 137]. Confounding due to different rates of uptake and/or frequency of mammography screening may explain some of the inconsistencies, since little consideration has been given to potential biases due to differences in screening behaviour among HRT users. Restricting analyses to women known to have participated in screening programs is one way to limit the potential for biased results.

South Australian women have been among the most enthusiastic users of HRT, with nearly 50% of women in their 50's using HRT during the 1990's[138]. Oestrogen-progestin combinations, which confer greatest risk[110], were most commonly prescribed and long term use (>7 years) was not uncommon[138]. Globally, the prevalence of HRT use and prescribing patterns have changed substantially since the publication of results from the Women's Health Initiative (WHI) trial in 2002, showing increased breast cancer risk with combined HRT[101]. As elsewhere, South Australian women were also quick to cease using HRT[62]. Of the few studies that have examined the effects of HRT use within a screening population, only two have included analysis for screening periods beyond 2002, with contradictory results for breast cancer risk[127, 139].

South Australia has had a long standing, high quality population-based mammography program with participation rates around 60% for most of its 20 years of operation[51]. The high rate of participation in screening, and the high prevalence of HRT use provide an ideal opportunity to study the effects of HRT in a steady-state screening population.

The aim of this study was to comprehensively examine the effect of HRT use on mammography screening outcomes within a mature population-based screening

program, including the impact on rate of screen-detected cancers, interval cancers, recall and biopsy rates.

## METHODS

This project was reviewed and approved by the Human Research Ethics Committees of SA Health and the University of Adelaide. Approval was obtained from BreastScreen SA's Medical Executive Committee.

### Study population

The study population consisted of all screening episodes undertaken by women aged 40 years or older who attended BSSA between January 1998 and December 2009 (n=234,370 women; 819,719 screening episodes). During this period the proportion and average age of women attending first round screening was relatively stable. Our sample included 123,517 first and 696,205 subsequent screening rounds, with an average of 4 screening episodes per individual.

### Screening program

BreastScreen SA has operated since 1991 as a single centralised service with several fixed and mobile units providing state-wide coverage. Until 2010 the service only operated analogue screening units. Women in the target age group (50-69 years) receive a written invitation to participate in mammography screening, while women aged 40-49 years and 70 years or older are eligible to attend but are not actively recruited. Symptomatic women are referred to diagnostic screening services. Screening is undertaken biennially, except among high risk groups (e.g. those with strong family history of breast cancer) who may undertake annual mammography. Participation and rescreen rates across the study period were 60% and 80% respectively [51].

Routine screening involves two-view standard mammography of each breast (medio-lateral-oblique and cranio-caudal views) read independently by two specialist radiologists, with a third reader engaged if discordance occurs. Women are recalled for further assessment based on the classification of mammographic findings, graded on a 5-point scale according to level of suspicion of malignancy [140]. Further assessment involves clinical examination, additional mammographic imaging and ultrasonography,

if appropriate. Women with lesions considered indeterminate, suspicious or potentially malignant on imaging undergo percutaneous needle biopsy (PCNB) methods by fine needle aspiration biopsy, ultrasound-guided core biopsy or vacuum assisted core biopsy to determine the nature of the lesion. Women with malignancy are referred for treatment while those with inconclusive results are referred for open surgical biopsy.

**Measures/Definitions**

Using BSSA system records for each screening episode, de-identified data were collected regarding age, date, round and postcode, mammographic findings, additional imaging, assessment procedures and biopsy results for women with screen-detected abnormalities. Pathology data was retrieved for participants treated for breast cancer. Interval cancers (to 24 months) were identified through routine linkage with South Australian Cancer Registry.

At each screening appointment women are asked to provide information about family history of breast cancer, prior breast procedures and use of hormone replacement therapy through a written questionnaire. Women were asked to report on whether they had used HRT in the 6 months prior to their screening appointment. No other details (e.g. type or duration of HRT use) were collected by BSSA. We classified 'current use of HRT' as any use of HRT in the 6 months prior to each specific screening episode. 'Past use' was defined as self-reported HRT use at any prior screening appointment. Dichotomous variables were created for family history of breast cancer and prior breast disease. Family history status was determined via an algorithm within the BSSA database and applied prospectively from the screening episode when first reported 'Family history of breast cancer' applied to women with one first-degree relative diagnosed with breast cancer before age 50, or two or more first-degree relatives diagnosed at any age. Women were classified as having a 'prior breast problem' if they reported having had any breast problems or procedures in the past, including removal of a lump or cystic fluid, breast abscess, breast reconstruction, reduction or other surgery. Socioeconomic position (SEP) was assigned on the basis of women's postcode at the time of screening, using the Australian Bureau of Statistic's Index of Socioeconomic Advantage and Disadvantage (IRSAD), which is a measure of the

average level of access to resources and ability to participate in society at the neighbourhood level[141]. SEP was categorised into quintiles corresponding to the distribution within the SA population.

Outcome measures included:

*Screen-detected breast cancer* - All invasive breast cancers (IBC) or ductal carcinoma in-situ (DCIS) diagnosed as a result of mammography screening (i.e. ICD 10 C50). According to BreastScreen Australia's data dictionary, lobular carcinoma in-situ is considered a pre-malignant condition and was excluded;

*Interval breast cancer* - All IBC diagnosed with 24 months of the last screening mammogram following either a negative mammogram or negative findings at assessment identified through cancer registry follow-up. Interval cancers were further classified as occurring within the first year (0-12 months) or the second year (13-24 months) of the screening mammogram;

*Recall for assessment* - All episodes where women were recalled for further tests following a screening mammogram. This includes further imaging workup, with or without biopsy procedures;

*Percutaneous needle biopsy (PCNB)* - All fine needle aspiration or core biopsy procedures undertaken during assessment following a screening mammogram. Biopsy methods were combined to simplify analysis;

*Open biopsy* - All surgical biopsy procedures following recall for assessment;

*Clinically palpable lesion* - All cases where clinical examination at assessment detected a palpable lesion;

*Dominant radiological feature* – The dominant radiological abnormality, as prospectively designated by the assessment radiologist. Categories include microcalcifications, stellate mass, discrete mass with or without calcifications, architectural distortions and non-specific/asymmetric density. Only one dominant category is assigned per lesion.

Recall, dominant radiological features, PCNB and open biopsy were further classified according to whether or not these investigations ultimately resulted in a malignant diagnosis.

**Statistical methods**

The overall risk of breast cancer, including all screen-detected or interval IBC and DCIS, associated with HRT use at current and prior screening rounds, compared with no reported use at screening, was examined using Poisson regression modelling with all screening episodes included. Further separate regression models were undertaken to estimate incidence risk ratios (IRR) and determine 95% confidence intervals (CI) for various screening outcomes, for current HRT users compared with past or never users of HRT. All models were adjusted for age at screening (5 year age groups), period of screening (3 year periods), SEP (quintiles), family history of breast cancer and prior breast problems. Time since last screening episode was also included in our models due the potential association between HRT use and screening interval, given that women seeking or renewing HRT prescriptions were often prompted to attend screening by their GP. To determine whether effects differed for prevalent and incident screening rounds, analyses were stratified according to whether screening was a first or subsequent screening episode. Similarly, to examine age-specific effects, models were stratified by age group. Likelihood ratio tests were used to test for interaction in full models which included interaction terms for HRT with age and HRT with screening round. In the analyses of radiological features at assessment, cases with multiple lesions (n=2189 screening episodes) were excluded since it was not possible to specify which lesion was ultimately malignant due to the design of the database. Since models included data for each screening episode, the resulting IRRs represent the risks for current HRT use at screening (rather than cumulative risk for an individual over time). The potential for non-independence of outcomes across screening episodes was accounted for using robust variance estimation, clustering at the person level, where appropriate[142]. Models showed no evidence of over-dispersion. All analyses were undertaken using Stata 12.0 statistical software[143].

## RESULTS

Characteristics of screening participants at first and subsequent screening episodes are presented in Table 3.1. Current HRT use was reported by 20% of first round participants and 30% of subsequent round participants. Screen-detection, recall and biopsy rates were higher at the first screening round than at subsequent screening rounds.

Overall risk of breast cancer (IBC or DCIS diagnosed at or within 24 months of a screening appointment) was increased among current HRT users (IRR=1.38; 1.30-1.46) but not past users of HRT (IRR=0.95; 0.88-1.02). As seen in Table 3.2, risk of breast cancer was also increased with age, higher SEP, family history of breast cancer and prior breast problems. The risk associated with current HRT use compared with never use was similar for the earlier and latter period of this study ($IRR_{[1998-2003]}$=1.33, 1.23-1.43; $IRR_{[2004-2009]}$=1.41, 1.30-1.54).

Risk of screen-detected breast cancer among current HRT users compared with past or never users was increased at subsequent screening rounds (IRR=1.31; 1.22-1.40) but there was no evidence of increased risk at the first screening round (IRR=1.05; 0.89-1.25), as shown in Table 3.3. The increased risk at subsequent screening rounds related to invasive breast cancer (IRR=1.37; 1.27-1.48), with no increased risk of DCIS (IRR=1.05; 0.89-1.23). Risk of interval cancer was increased among HRT users following both first (IRR=1.77; 1.33-2.37) and subsequent screening rounds (IRR=1.92; 1.72-2.15), and during both the first and second year after the negative screen. Risk of recall for further assessment was increased among current HRT users at subsequent screening rounds (IRR= 1.36; 1.31-1.40) but not at the first screen. Similarly the risks of PCNB and open biopsy were increased following subsequent screening rounds but not during the first screen. Likelihood ratio tests confirmed that effects of HRT differed according to screening round for screen-detected cancer (p=0.049), recall to assessment (p<0.001) and need for biopsy (p<0.001) but not for interval cancer (p=0.801) or screen-detected DCIS (p=0.596). Tests for interaction were borderline with respect to screen-detected IBC (p=0.058)

**Table 3.1 Characteristics and outcomes for first, subsequent and all screening rounds**

| | First screen | | Subsequent screens | | All screens | |
|---|---|---|---|---|---|---|
| **Characteristics at screening** | **No.** | **%** | **No.** | **%** | **No.** | **%** |
| **Total** | 123517 | 100.0 | 696205 | 100.0 | 819722 | 100.0 |
| **Age group (yrs)** | | | | | | |
| <50 | 50289 | 40.7 | 65400 | 9.4 | 115689 | 14.1 |
| 50-54 | 50720 | 41.1 | 141286 | 20.3 | 192006 | 23.4 |
| 55-59 | 9483 | 7.7 | 165169 | 23.7 | 174652 | 21.3 |
| 60-64 | 5424 | 4.4 | 144594 | 20.8 | 150018 | 18.3 |
| 65-69 | 3747 | 3.0 | 116615 | 16.8 | 120362 | 14.7 |
| 70+ | 3854 | 3.1 | 63141 | 9.1 | 66995 | 8.2 |
| **Screening period** | | | | | | |
| 1998-1999 | 27644 | 22.4 | 96590 | 13.9 | 124234 | 11.4 |
| 2000-2001 | 23761 | 19.2 | 111465 | 16.0 | 135226 | 12.4 |
| 2002-2003 | 18886 | 15.3 | 118791 | 17.1 | 137677 | 12.7 |
| 2004-2005 | 18379 | 14.9 | 122360 | 17.6 | 140739 | 12.9 |
| 2006-2007 | 15358 | 12.4 | 119557 | 17.2 | 134915 | 12.4 |
| 2008-2009 | 19489 | 15.8 | 127442 | 18.3 | 146931 | 12.5 |
| **SEP quintile** | | | | | | |
| Lowest | 23641 | 19.1 | 142137 | 20.4 | 165778 | 20.2 |
| Mid-low | 23259 | 18.8 | 136230 | 19.6 | 159489 | 19.5 |
| Mid | 21467 | 17.4 | 126118 | 18.1 | 147585 | 18.0 |
| Mid-high | 25052 | 20.3 | 135888 | 19.5 | 160940 | 19.6 |
| Highest | 30095 | 24.4 | 155828 | 22.4 | 185923 | 22.7 |
| **HRT use at screen** | 24547 | 19.9 | 206267 | 29.6 | 230814 | 28.2 |
| **Strong Family history BC** | 5354 | 4.3 | 57422 | 8.3 | 62776 | 7.7 |
| **Prior breast problems** | 26551 | 21.5 | 154927 | 22.3 | 181478 | 22.1 |
| **Outcomes of screening** | **No.** | **n/1000** | **No.** | **n/1000** | **No.** | **n/1000** |
| Screen-detected BC | 790 | 6.4 | 3979 | 5.7 | 4769 | 5.8 |
| Screen-detected IBC | 633 | 5.1 | 3246 | 4.7 | 3879 | 4.7 |
| Screen-detected DCIS | 156 | 1.3 | 733 | 1.1 | 889 | 1.1 |
| Interval BC (0-24mths) | 229 | 1.9 | 128 | 1.8 | 1516 | 1.8 |
| Recall for assessment | 6131 | 49.6 | 16043 | 23.0 | 22174 | 27.1 |
| Percutaneous biopsy | 2474 | 20.0 | 6711 | 9.6 | 9185 | 11.2 |
| Open (surgical) biopsy | 385 | 3.1 | 991 | 1.3 | 1296 | 1.6 |

SEP=socio-economic position; BC=breast cancer; HRT=hormone replacement therapy

**Table 3.2 Risk of breast cancer (IBC + DCIS) detected at screening or diagnosed within 24 months of screening among BreastScreen SA participants 1998-2009[a] (n= 819715)**

| Risk factors | IRR | 95% CI. | p-value |
|---|---|---|---|
| **HRT use at screening[b]** | | | |
| None (reference) | 1.00 | - | - |
| Past[#] | 0.95 | 0.88-1.02 | 0.162 |
| Current | 1.38 | 1.30-1.46 | <0.001 |
| **Age group (yrs)** | | | |
| <50 | 0.64 | 0.58-0.71 | <0.001 |
| 50-54 | 0.81 | 0.75-0.88 | <0.001 |
| 55-59 (reference) | 1.00 | - | - |
| 60-64 | 1.24 | 1.15-1.34 | <0.001 |
| 65-69 | 1.38 | 1.27-1.49 | <0.001 |
| 70+ | 1.57 | 1.43-1.72 | <0.001 |
| **SEP quintile** | | | |
| Lowest (reference) | 1.00 | - | - |
| Low-mid | 1.06 | 0.98-1.15 | 0.164 |
| Mid | 1.09 | 1.00-1.18 | 0.042 |
| Mid-high | 1.14 | 1.05-1.23 | 0.001 |
| Highest | 1.17 | 1.08-1.26 | <0.001 |
| **Screening period** | | | |
| 1998-2000 (reference) | 1.00 | - | - |
| 2001-2003 | 1.02 | 0.95-1.10 | 0.518 |
| 2004-2006 | 0.99 | 0.93-1.07 | 0.880 |
| 2007-2009 | 1.02 | 0.95-1.10 | 0.572 |
| **Strong family history BC** | 1.38 | 1.21-1.57 | <0.001 |
| **Prior breast problems** | 1.29 | 1.22-1.36 | <0.001 |
| **Screening interval** | | | |
| First screen | 1.53 | 1.42-1.66 | <0.001 |
| <18 months | 0.82 | 0.70-0.96 | 0.013 |
| 18-29 months (reference) | 1.00 | - | - |
| 30+ months | 1.26 | 1.17-1.36 | <0.001 |

[a]Incidence risk ratios (IRR) and 95% confidence intervals (CI) from Poisson regression modelling with all factors entered simultaneously.

[b]HRT use was determined from women's self-reported use in the 6 months prior to each screening episode. Past use may therefore be under-ascertained due to cases where HRT was used between screening events but not in the 6 months prior to screening, or before commencing the screening program.

Increased risk of recall to assessment applied to both recall resulting in malignant findings (IRR= 1.27; 1.19-1.35) and recall where no malignancy was found (IRR= 1.27; 1.23-1.32) as shown in Table 3.4. Similarly current HRT users were more likely to have clinically palpable masses or multiple screen-detected lesions, regardless of the final outcome of assessment. The risk of recall was increased significantly across all categories of radiological abnormalities with the exception of architectural distortions. We note the positive predictive value (PPV) of the radiologic categories are widely variable, with stellate lesions and calcifications having the highest PPVs (85% and 32% respectively) and nonspecific density and discrete masses having the lowest PPVs (11% and 9% respectively). In relation to lesions leading to malignant findings, risk with current HRT uses was only increased for stellate lesions (IRR= 1.49; 1.35-1.64). In addition, HRT users were at greater risk of having needle biopsy procedures with both malignant (IRR= 1.27; 1.19-1.35) and non-malignant findings (IRR= 1.18; 1.11-1.26). The slightly increased risk for open biopsy rates was not statistically significant.

In age stratified analyses there was no increased risk of screen-detected IBC with current HRT use among women aged under 50 years and 50-54 years, whereas risk was significantly increased for all other age groups (Figure 3.1). Likewise there was no increased risk of being recalled or undergoing biopsy procedures among younger women, but risk was increased for women aged 55 years and older at screening. Risk of interval cancers following subsequent screening rounds was not increased in women aged under 50 years but was increased among HRT users aged 50-54 years and older. Risk of screen-detected DCIS was not significantly increased in any age group. Statistically significant interactions by age group were found for screen-detected cancers ($p=0.049$), recall to assessment ($p<0.001$) and the need for biopsy ($p<0.001$) but not for interval cancers ($p=0.122$).

.

**Table 3.3 Risk of various screening outcomes at first and subsequent screening episodes for current compared with past or never HRT use, among BreastScreen SA participants 1998-2009[a]**

| Screening outcomes: | Total No. Events | First screening round n=123514 | | | | | Subsequent screening round n=696201 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | No HRT n=98970 | Current HRT n=24547 | IRR | 95% CI | p-value | No/past HRT n=489938 | Current HRT n=489938 | IRR | 95% CI | p-value |
| Screen-detected BC | 4769 ( 0.58) | 610 (0.62) | 180 (0.73) | 1.05 | 0.89-1.25 | 0.573 | 2594 (0.53) | 1385 (0.67) | 1.31 | 1.22-1.40 | <0.001 |
| -IBC | 3879 (0.47) | 485 (0.49) | 148 (0.60) | 1.08 | 0.89-1.30 | 0.444 | 2089 (0.43) | 1157 (0.56) | 1.37 | 1.27-1.48 | <0.001 |
| -DCIS | 889 (0.11) | 124 (0.13) | 32 (0.13) | 0.96 | 0.64-1.43 | 0.832 | 505 (0.10) | 228 (0.11) | 1.05 | 0.89-1.23 | 0.563 |
| Interval cancers (IBC only) | 1516 (0.18) | 158 (0.16) | 71 (0.29) | 1.77 | 1.33-2.37 | <0.001 | 708 (0.14) | 579 (0.28) | 1.92 | 1.72-2.15 | <0.001 |
| -first 12 months | 556 (0.07) | 63 (0.06) | 24 (0.10) | 1.54 | 0.94-2.50 | 0.084 | 259 (0.05) | 210 (0.10) | 1.91 | 1.59-2.31 | <0.001 |
| -13-24 months | 960 (0.12) | 95 (0.10) | 47 (0.19) | 1.93 | 1.34-2.77 | <0.001 | 449 (0.09) | 369 (0.18) | 1.93 | 1.67-2.22 | <0.001 |
| Recall for assessment | 22174 (2.71) | 4886 (4.94) | 1245 (5.07) | 1.02 | 0.96-1.09 | 0.522 | 10326 (2.11) | 5717 (2.77) | 1.36 | 1.31-1.40 | <0.001 |
| FNA/core biopsy performed | 9185 (1.12) | 1968(1.99) | 506 (2.06) | 0.97 | 0.87-1.07 | 0.498 | 4352(0.89) | 2359 (1.14) | 1.31 | 1.25-1.38 | <0.001 |
| Open biopsy performed | 1296 (0.16) | 303 (0.31) | 82 (0.33) | 0.94 | 0.73-1.20 | 0.604 | 595(0.12) | 316 (0.15) | 1.17 | 1.01-1.34 | 0.033 |

[a]Poisson regression models adjusted for age group (5yr), SEP (quintiles), screening period (3yr), family history (Y/N), prior breast problems (Y/N), and screening interval

IRR = incidence rate ratio, CI = confidence interval, FNA= fine needle aspiration; BC=breast cancer

**Table 3.4 Risk of various assessment outcomes for current compared with past or never HRT use, among BreastScreen SA participants (all screening episodes) 1998-2009[a]**

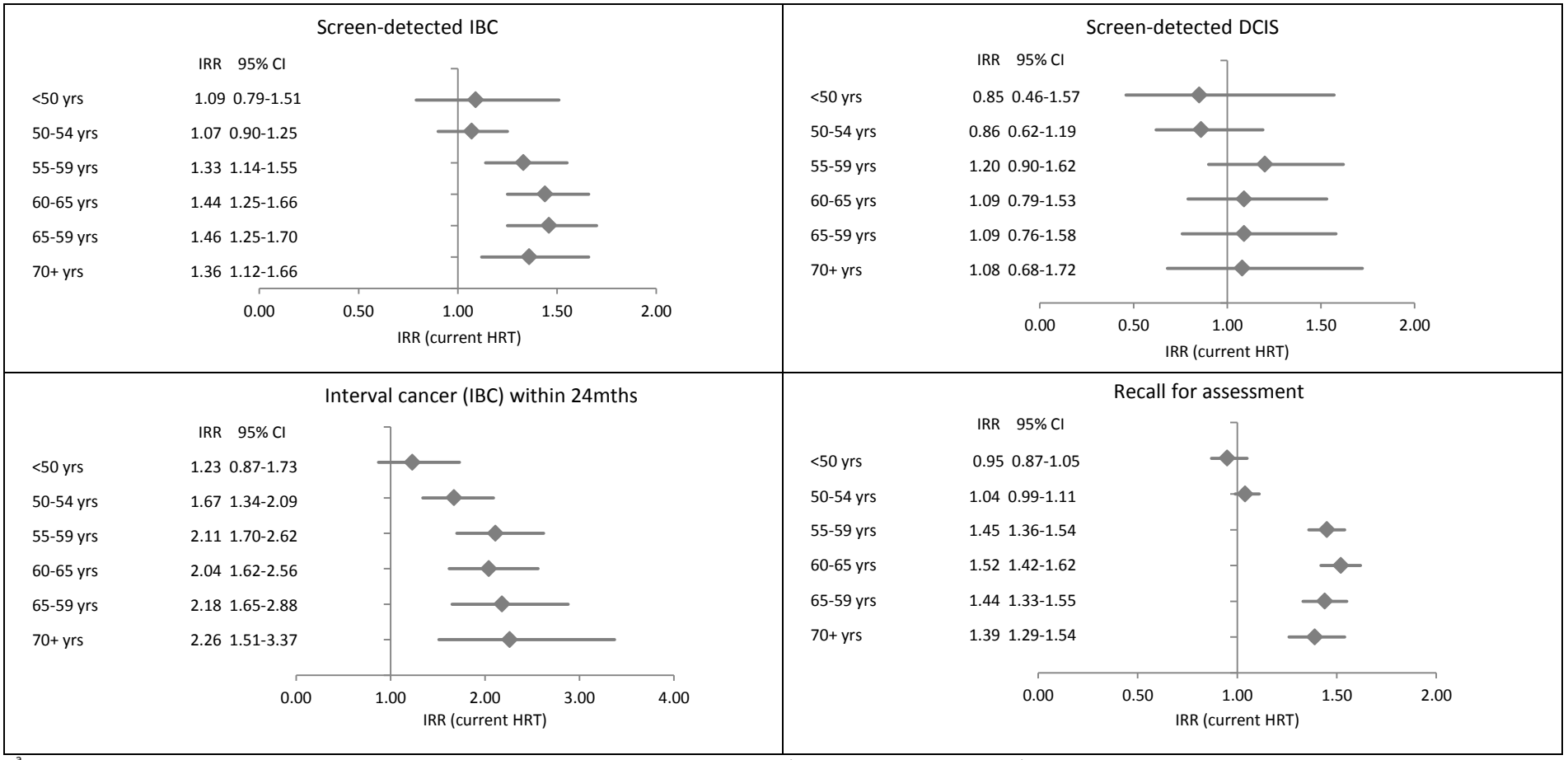
| Assessment findings | All events | | | | Non-malignant outcome | | | | Malignant outcome | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| N=819715 | n (%) | IRR | 95%CI | p-value | n (%) | IRR | 95%CI | p-value | n (%) | IRR | 95%CI | p-value |
| Recall for assessment | 22174 (2.71) | 1.27 | 1.24-1.31 | <0.001 | 17405 (2.12) | 1.27 | 1.23-1.32 | <0.001 | 4769 (0.58) | 1.27 | 1.19-1.35 | <0.001 |
| Palpable mass | 2201 (0.27) | 1.33 | 1.21-1.45 | <0.001 | 522 (0.06) | 1.22 | 1.01-1.46 | 0.037 | 1679 (0.20) | 1.36 | 1.22-1.51 | <0.001 |
| Multiple lesions | 2189 (0.27) | 1.33 | 1.21-1.46 | <0.001 | 1245 (0.15) | 1.27 | 1.13-1.44 | <0.001 | 944 (0.12) | 1.40 | 1.22-1.61 | <0.001 |
| Percutaneous biopsy | 9185 (1.12) | 1.23 | 1.17-1.28 | <0.001 | 4425 (0.54) | 1.18 | 1.11-1.26 | <0.001 | 4760 (0.58) | 1.27 | 1.19-1.35 | <0.001 |
| Open biopsy | 1296 (0.16) | 1.10 | 0.98-1.25 | 0.110 | 828 (0.10) | 1.11 | 0.96-1.30 | 0.166 | 468 (0.06) | 1.09 | 0.89-1.33 | 0.414 |
| *Radiological findings*[b] (N=817526) | | | | | | | | | | | | |
| Calcifications | 2851 (0.35) | 1.18 | 1.09-1.28 | <0.001 | 1900 (0.23) | 1.24 | 1.12-1.37 | <0.001 | 951 (0.12) | 1.08 | 0.94-1.25 | 0.264 |
| Stellate lesion | 2122 (0.26) | 1.46 | 1.33-1.60 | <0.001 | 322 (0.04) | 1.30 | 1.02-1.65 | 0.033 | 1800 (0.22) | 1.49 | 1.35-1.64 | <0.001 |
| Discrete mass | 6711 (0.82) | 1.28 | 1.21-1.34 | <0.001 | 6006 (0.73) | 1.32 | 1.25-1.39 | <0.001 | 705 (0.09) | 0.95 | 0.80-1.13 | 0.558 |
| Architectural distortion | 244 (0.03) | 1.09 | 0.81-1.46 | 0.579 | 178 (0.02) | 1.10 | 0.77-1.57 | 0.5930 | 66 (0.01) | 1.04 | 0.60-1.80 | 0.884 |
| Non-specific density | 1717 (0.21) | 1.38 | 1.24-1.53 | <0.001 | 1475 (0.18) | 1.45 | 1.30-1.61 | <0.001 | 242 (0.03) | 1.01 | 0.74-1.37 | 0.956 |

[a]Poisson regression models adjusting for age group (5yr), SEP (quintiles), screening period (3yr), family history (Y/N), prior breast problems (Y/N) and screening interval

n = number with outcome, % = percent of screening events with outcome, IRR = incidence rate ratio for current HRT use, CI = confidence interval

[b]Excludes cases with multiple lesions in analyses for radiologic findings

**Figure 3.1 Age stratified incidence rate ratios and 95% confidence intervals for screening outcomes among current compared with past or never HRT use at screening, among BreastScreen SA participants (all screening episodes) 1998-2009[a]**




[a]Poisson regression models adjusting for SEP (quintiles), screening period (3yr), family history (Y/N), prior breast problems (Y/N) and screening interval , stratified by age group.

**DISCUSSION**

Current but not past HRT use was associated with moderately increased risk of breast cancer at or within 2 years of screening. While information is lacking on the specific types and duration of HRT use at the individual level, our results indicate that, at a population level, recent patterns of HRT use among South Australian women have led to increased breast cancer risk. It is possible this association is confounded by unmeasured risk factors (e.g. reproductive history), however risk profiles between users and non-users of HRT would have to differ substantially for confounding to explain these findings, given that the study sample is restricted to BSSA participants and analyses are adjusted for screening interval, as well as SEP, family history and prior breast problems.

The lack of effect among past HRT users is consistent with the WHI trial results showing a rapid decrease in risk within 2 years of ceasing hormone therapy [109], suggesting that the effect of HRT is transient rather than permanent in nature.

While current HRT was associated with increased risk of screen-detected breast cancer, interval cancers and investigative procedures, as reported in most other screening cohorts[126-129, 133, 144-147], no effect was observed among younger women (<50 years), or at first round screening (except for interval cancers) or in relation to pre-invasive disease.

Most previous studies have either excluded first round screening from analysis to intentionally remove prevalent cancer cases[126, 127, 134, 145] or have combined first and subsequent screening rounds[128, 144]. The only other study to examine first round screening separately also found no increased risk of screen-detected cancer at the first round[148]. One explanation may be that HRT's effect is 'hidden' within the large pool of prevalent cancer cases detected at the first round. Another explanation may be that duration of use at first screen is too short for any impact to be evident. Several studies have shown a minimal effect in the first 2-3 years of HRT use with increased risk with increasing duration thereafter[109, 124], which is a plausible explanation of our data. Alternatively, if HRT acts to increase the growth rate of pre-existing breast cancers as has been proposed[111], HRT users would disproportionately develop clinical symptoms leading to diagnosis before commencing screening. They will thus be removed from

the eligible screening population, leading to fewer screen-detected cancers. If inflow into the screen-detectable pool is balanced by outflow (through clinical diagnosis) then no increased risk will be evident at the first round of screening. At subsequent screening rounds this balance would be disrupted through earlier diagnosis, leading to increased risk of screen-detectable breast cancer.

The 2-fold increased risk of interval cancer (over 24 months) observed in this study is consistent with findings in other studies[127, 129, 131, 135]. Most but not all studies[149, 150] found similar effects following first and subsequent screening rounds, and in the first and second year after screening. The equivalent effect of HRT use on interval cancers in the first and second year after screening suggests that the impact of HRT is relatively constant over the 2 year interval. HRT use has been associated with increased breast density in some women which is thought to lead to reduced sensitivity of mammography and greater risk of 'missed' cancers at screening[151]. Recent modelling studies examining growth rates using mammography screening data, have estimated a 25% increased growth rate with HRT use, which could also account for both increased screen-detected cancers and an increased interval cancer rate[152, 153].

Our findings suggest that current HRT use does not increase the risk of DCIS, even though the risk of IBC is increased. Evidence for increased risk of DCIS with HRT use is much less clear than for IBC[136]. Several early studies found no increased risk of DCIS with HRT use, though these are likely to have been underpowered[144, 154, 155]. The WHI trials found a non-significant increased risk of DCIS with use of oestrogen-progestin HRT compared with placebo[102] but no evidence of increased risk with oestrogen only [156]. In contrast other observational studies have found increased risk of DCIS with oestrogen only preparations but no or little effect with oestrogen-progestin use[121, 157], though one has found the reverse[123]. A recent study by Calvocoressiet al.[137] found no increased risk with either oestrogen alone or oestrogen and progestin combined, for any duration of use. While IBC and DCIS are thought to share similar risk factors[158] it is possible that they actually have different aetiologies and hence could respond differently to HRT. Certainly the rapid and transient nature of the effect of HRT on invasive breast cancers indicates that it acts at a different time scale to that of DCIS. An

alternative explanation of the data is that HRT use increases transition from pre-invasive to invasive disease.

No association was found between HRT use and screen-detected IBC or interval cancer in younger women (aged <55 years and <50 years, respectively). One explanation for the lack of association at younger ages is that comparison groups were not equivalent in terms of menopausal status, with non-users who had not reached menopause compared with HRT users who had. If this is the case, the effect of exogenous hormone therapy in younger menopausal women is likely to be equivalent to natural hormone levels in premenopausal women of similar age. This may also contribute to null findings at first round screening, since 80% were aged under 55 yearss at first screen.

This study found increased risk of recall and investigative procedures with current HRT use, as observed by others[128, 129, 133, 146, 147], as well as increased risk of clinically palpable masses and various classes of mammographic lesions. Some lesions such as stellate masses have high PPV for malignancy, while others, such as discrete masses and non-specific densities are more likely to be associated with non-malignant findings [159]. Since HRT use increases breast density in some women[151], it plausibly leads to an increased mammary fibroglandular tissue in the breast, some of which will produce non-specific densities without an underlying malignancy. The increased risk of discrete masses, which also have low PPV for malignancy, may indicate increased risk of other benign proliferative lesions. In contrast, stellate lesions have a high PPV for malignancy[160], and the detection of more stellate lesions would be attributable to the demonstrated increased risk of breast cancer among HRT users. This diversity of radiological features suggests that HRT may simultaneously exert effects on both benign and malignant breast tissue, which result in a range of radiologic abnormalities. Evidence that HRT increases the risk of benign proliferative disease[161, 162] supports this hypothesis, though further investigation is required.

**Strengths and limitations**

The large size and population-based nature of this study provide distinct advantages over many previous studies. Findings, particularly of non-association, are unlikely to be due to lack of power in a population of this size, with sufficient number of outcome

events. Quality of outcome measures would be high given prospective, routine and mandatory case ascertainment via the central cancer registry and complete monitoring of recall and biopsy rates for accreditation purposes. While measures of exposure (HRT use) and several covariates such as family history and prior breast procedures were based on self-reported data, questions were relatively easy for women to answer, hence data are close to complete and unlikely to be inaccurately reported. Furthermore, information was collected prior to screening, minimising the potential for recall bias. Since HRT users are more likely to attend screening[48, 163], restricting analyses to screening participants reduces the potential for outcomes to be confounded by differences in screening uptake.

Lack of information about the type of HRT used and use prior to commencing screening are limitations, since risks vary with type and duration of HRT used. Another limitation is the lack of data on other risk factors such as reproductive history, BMI and alcohol use, which may confound the relationship between HRT and screening outcomes. In addition we did not have information about menopausal status, so could not exclude premenopausal women from analyses. However, because the study was examining effects within the context of a screening program, inclusion of all screening participants can be justified. Age stratified analysis sheds some light on the effects among menopausal women in older age groups (>55 years), as most would have reached menopause by this age.

**Conclusion**

Our findings demonstrate important associations between current HRT use and the outcomes of mammographic screening at the population level. HRT users have higher recall rates, are more likely to undergo biopsy procedures and more likely to be diagnosed with invasive breast cancer both as a result of screening and after negative screening results. Awareness of these facts is important for the operations of breast cancer screening at the population level. Our results are consistent with a possible role for HRT in enhancing the growth of pre-existing tumours rather than initiating nascent cancers. Lack of effect among past users suggest the effect of HRT is transient, while increased risk of both screen-detected and interval cancers, suggests a faster growth rate to reach a screen-detectable threshold (or clinically-detectable threshold in the

case of interval cancers) with HRT use, as suggested by modelling studies. Furthermore, HRT may also be having a more general proliferative effect on hormonally responsive non-malignant breast tissue, giving rise to a range of other non-malignant breast abnormalities and increasing mammographic density. This latter effect would lead to reduced mammographic sensitivity and increased interval cancers. The lack of effect on DCIS is somewhat incongruous, however, in addition to enhancing tumour growth, HRT may also increase the invasive potential of in-situ cancers. Alternatively, the effects of HRT on DCIS may be acting on a longer time horizon that is not detectable given the relatively short period of HRT use among study participants.

**Acknowledgements**

This study was funded by the National Breast Cancer Foundation. The authors wish to acknowledge Ms Jill Rogers from BreastScreen SA who assembled the dataset used for this research project. We also wish to acknowledge the contributions of past and current staff including data managers, radiologists, pathologists and surgeons working at BSSA, as well as the support of the General Manager, Ms Lou Williamson.

# Chapter 4: Do breast cancer risk factors differ among those who do and do not undertake mammography screening?

Paper published in Journal of Medical Screening 2013;20:208-19.

**Authors:** Kerri Beckmann, David Roder, Janet Hiller, Gelareh Farshid, John Lynch

## PREFACE

In observational studies that attempt to measure over-diagnosis by comparing breast cancer incidence in cohorts which have and have not participated in screening, it is important to consider whether the women in each of the comparison groups have equivalent risk of breast cancer. Any imbalance may result in biased estimates of over-diagnosis, that is, some or all of the observed difference may be due to screening participants having a higher (or lower) underlying risk of breast cancer. This is a key consideration in relation to the case-control study presented in Chapter 6, which uses individual level data on screening participation to estimate over-diagnosis. The study presented in this chapter examines differences in breast cancer risk factor profiles among South Australian women who have ever participated in BreastScreen with those who have not. Risk factor profiles of women who have ever and never participated in private screening mammography outside of BreastScreen are also compared.

## AUTHORSHIP STATEMENT

**Do breast cancer risk factors differ among those who do and do not undertake mammography screening?**
*Journal of Medical Screening* 2013;20:208-19.

**Kerri Beckmann (Candidate)**
I conceived and designed the study, managed data collection, analysed and interpreted the data, drafted and critically revised the manuscript, coordinated contributions from co-authors and acted as corresponding author for peer review and preparation for publication.

25/09/14
Signed …… Date……

**David Roder**
My contribution to this publication involved advising on research design and methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

22/09/14
Signed …… Date……

**Janet Hiller**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**Gelareh Farshid**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**John Lynch**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

23/09/14
Signed …… Date……

## ABSTRACT

**Objectives:** There is considerable interest in whether mammography screening leads to over-diagnosis of breast cancer. However, self-selection into screening programs may lead to risk differences that affect estimates of over-diagnosis. This study compares the breast cancer risk profiles of participants and non-participants of population-based mammography screening. Risk profiles are also compared between those who have and have not used private screening services.

**Setting:** This study involved 1162 women aged 40-84 years who participated in the 2012 Health Omnibus, an annual face-to-face interview-based survey of a representative sample of South Australians.

**Methods:** Data were collected on participation in mammography screening, demographic characteristics and breast cancer risk factors (including reproductive, familial and lifestyle factors). Missing data were multiply imputed. Factors independently associated with ever having screened were identified using multivariable logistic regression for population-based and ad hoc, private screening mammography separately.

**Results:** Compared with non-participants, participants of population-based screening were more likely to have used HRT (OR=3.72), experienced breast biopsy or surgery (OR=2.22), and be overweight or obese (OR=1.57). They were less likely to be sufficiently active (OR=0.57) or be born in a non-English speaking country (OR=0.50) or aged under 50 years (OR=0.09). Women who screened privately were more likely to have a family history of breast cancer (OR=1.66) and have experienced breast biopsy or surgery (OR=3.17) than those who had not.

**Conclusions:** South Australian women who participated in the population-based mammography screening have a slightly higher prevalence of breast cancer risk factors. This also applies to those who undertook private screening.

## INTRODUCTION

While numerous studies have examined socioeconomic and psychosocial predictors of participation in mammography screening[164-168], few have focused on whether breast cancer risk differs between participants and non-participants. Given the current debate about the extent of 'over-diagnosis' due to mammography screening (i.e. the detection of cancers that would not have been diagnosed in a woman's lifetime had she not participated in screening[56]), it is important to determine whether women who attend mammography screening differ in their level of breast cancer risk from those who do not. Self-selection into screening programs may affect estimates of over-diagnosis if breast cancer risk is different among screening participants compared with non-participants.

The extent to which breast cancer risk factor profiles differ between breast screening participants and non-participants is largely unknown and may vary by population. Data from the original mammography screening trials suggest that women who did not take up the invitation to attend screening had a higher risk of death from breast cancer[169]. However, the underlying risk of breast cancer among non-attenders was not investigated, and little information is available in current non-trial settings. Heightened concern about cancer among women with a family history of breast cancer or a personal history of benign breast disease may lead to greater participation, and consequently, higher breast cancer risk among screening participants[170, 171]. Several studies have identified socioeconomic differences in mammography screening uptake[172-174]. Many established reproductive risk factors (e.g. nulliparity, later age at first birth, later menopause) are more prevalent among affluent women[175], who are more likely to screen, again leading to higher breast cancer risk among screening participants. A recent Swedish study, however, found little evidence of associations between screening attendance and reproductive histories[176]. Conversely, other lifestyle risk factors associated with lower socioeconomic position (e.g. high alcohol consumption, physical inactivity, obesity) may be more prevalent among non-participants[177, 178], leading to lower breast cancer risk among screening participants.

Local context is also important. Many studies on mammography participation have been conducted in the United States, where factors associated with socioeconomic

differences may be more important due to the lack of organised population-based screening[179]. Socioeconomic disparities, however, were not evident in most other countries where organized population-wide screening programs have been established[180, 181].

Australia introduced population-based mammography screening, BreastScreen Australia, in 1991, which offers free biennial mammograms targeting women aged 50-69 years. Women aged 40-49 years and over 69 years are eligible to attend but are not been actively recruited. Participation has been reasonably stable at around 55-60% of women in the target age range[51]. However, BreastScreen records indicate that approximately 20% of women 50 years or older have never participated in the screening program. Women may choose to access mammography screening through private imaging services, but little is known about the extent of screening outside the BreastScreen program.

In this study we examine differences in breast cancer risk factor profiles among South Australian (SA) women who have ever participated in BreastScreen with those who have not. We also compare the risk factor profiles of women who have ever and never participated in screening mammography outside of BreastScreen.

## METHODS

### Study population

Data on screening participation, breast cancer risk factors and demographic characteristics were collected from 1162 women aged 40-84 years, via the 2012 Health Omnibus Survey. The Health Omnibus is an annual face-to-face interview survey which uses a random stratified cluster sampling technique to recruit a representative sample of 3000 South Australian men and women from selected households[107]. The participation rate in 2012 was 70.4%. Survey methodology and research questions were approved by the University of Adelaide Human Research Ethics Committee.

### Measures

Demographic factors included age, highest education level, household income, country of birth and place of residence. Socioeconomic position (SEP) was derived from

participant's residential postcode using the Index of Relative Socioeconomic Advantage and Disadvantage for 2006[141], which represents the average socioeconomic status of people living within a specified postcode district. SEP was categorised into quintiles according to the distribution for SA.

Data on screening attendance and breast cancer risk factors were only asked of women aged 40 years and over. Women were asked whether they had "ever had a screening mammogram at a BreastScreen clinic or in a BreastScreen mobile van" and whether they had "ever had a screening mammogram at a private clinic or other facility that was not run by BreastScreen". Screening was defined as "a routine check for breast cancer when there were no specific problems or symptoms".

Personal and behavioural risk factors included parity, age at menarche, menopause and first birth, past or current hormone replacement therapy (HRT) use, height, weight, level of physical activity, alcohol consumption, family history of breast cancer and personal experience of breast biopsy or surgery. Body mass index (BMI) was categorised as normal or underweight (<=24.9kg/$m^2$), overweight (25-29.9kg/$m^2$) and obese (30+kg/$m^2$), based on self-reported height and weight. Physical activity was assessed as the time per week spent walking, undertaking moderate activities (e.g. lawn bowls, golf, gentle swimming), and undertaking vigorous activities (e.g. tennis, jogging, cycling, keep fit classes), converted to metabolic equivalent time (MET-mins/week) according to IPAQ scoring protocol (short form)[182]. Insufficient physical activity was classified as <600MET-mins/week. Alcohol consumption was assessed using two questions, "How often do you usually drink alcohol?" and "On a day that you drink how many standard drinks do you have?" with a visual cue-card for standard serving sizes. 'At-risk consumption' was classified as >7 standard drinks/week, based on recommendations for cancer prevention[183]. Family history of breast cancer was classified as one or more close female relatives ever having been diagnosed with breast cancer (including mother, sisters, daughters and genetically related aunties). Women were also asked whether they had "ever had a breast biopsy or other surgical procedure involving the breast for any reason". This question was intended as a proxy measure for precancerous conditions (e.g. hyperplasia) hence women with breast cancer excluded from the positive response category.

**Analysis**

Analysis was restricted to women aged 40-84 years. Data were initially weighted by individuals' inverse probability of selection according to area and household size and reweighted to reflect the age profile for SA females, using the 2011 mid-year estimated residential population[105].

Due to the high rate of case-wise missingness (cases with missing data in one or more variables) we conducted analysis using both complete case and multiply imputed (MI) data. Imputation was undertaken using multiple imputation chain equations to impute 20 datasets, with 5 cycles of regression switching[184]. Our imputation model included all outcomes and covariates in our analytic models, as well as marital status and employment status which both predicted missingness. Logistic regression was used in imputation modelling for binary variables and truncated regression for continuous variables, with upper and lower bounds taking the highest and lowest values of survey responses. Education and income levels were modelled using ordinal logistic regression and age at menopause modelled using multinomial logistic regression (to include 'pre-menopausal'). Validity of imputed data was checked using diagnostic plots and frequency tabulations. Analyses using MI data were restricted to cases with complete data on screening attendance[185].

Weighted cross tabulations were performed for each of the selected characteristics and risk factors, comparing women who'd ever participated in BreastScreen with those who had never participated. Multivariable logistic regression was used to identify risk factors that were independently associated with BreastScreen participation, with all variables modelled simultaneously. Reproductive and lifestyle risk factors were modelled as dichotomous variables indicating high versus low risk categories. Similar analyses were undertaken to identify risk factors associated with participation in private mammography screening.

Three summary measures of breast cancer risk were also developed to allow comparison of overall risk among screening and non-screening groups. The first of these measures was the total number of risk factors from twelve possible factors, including insufficient physical activity, overweight or obese, tall stature, at-risk alcohol consumption, HRT use, prior biopsy/surgery, nulliparity, late age at first birth, one or

more family members with breast cancer, early menarche, late menopause and being in the highest SEP quintile. The second measure was a risk score based the Gail model[186] with the following modifications: hyperplasia was not included in the model (since data were not collected); women reporting any breast biopsy or surgery were categorised as having had one biopsy, with no option for multiple biopsies in the model. The third of these measures was a risk score derived from the recently published risk prediction model for breast cancer developed by Pfeiffer et al[231]. This prediction model includes parity, age at first birth, menopausal status, age at menopause, length of use of estrogen-progestin HRT, other HRT use, BMI, alcohol consumption, benign breast disease and family history of breast and ovarian cancer. Risk factors were categorised as described by the authors expect that any biopsy/breast procedure was substituted for benign breast disease and family history of breast cancer was substituted for family history of breast and ovarian cancer. Risk scores were calculated for each individual based on published relative risks for each risk prediction model. We used linear regression with imputed data to compare both the risk scores and mean number of risk factors, according to screening participation.

All analyses were undertaken using Stata v12.0[143].

**Table 4.1 Demographic and risk factor profile of participants in the 2012 South Australian Health Omnibus Survey (females 40-84 years)**

| Factors | Total sample | | Complete case | | Multiply imputed | | % with complete data |
|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | |
| **All respondents** | 1148 | - | 705 | - | 1148 | - | **61.0** |
| (weighted no.) | (1150) | | (700) | | (1148) | | |
| **Age** | | | | | | | |
| 40-49yrs | 333 | 29.0 | 238 | 34.1 | 333 | 29.0 | 71.5 |
| 50-59yrs | 323 | 28.1 | 202 | 28.8 | 323 | 28.1 | 62.5 |
| 60-69yrs | 263 | 22.9 | 159 | 22.7 | 263 | 22.9 | 60.5 |
| 70-84yrs | 229 | 19.9 | 101 | 14.4 | 229 | 19.9 | 44.1 |
| **Education** | | | | | | | |
| Less than high school | 170 | 14.8 | 72 | 10.4 | 170 | 14.8 | 42.4 |
| Completed high school | 418 | 36.4 | 260 | 37.2 | 418 | 36.4 | 62.2 |
| Certificate or diploma | 348 | 30.3 | 211 | 30.1 | 348 | 30.3 | 60.6 |
| Bachelor degree | 212 | 18.5 | 156 | 22.3 | 212 | 18.5 | 73.6 |
| *Missing* | *2* | | | | | | |
| **Income** | | | | | | | |
| <=$30,000 | 241 | 28.1 | 179 | 25.6 | 334 | 29.1 | 53.6 |
| $30,001-60,000 | 176 | 20.6 | 143 | 20.4 | 247 | 21.5 | 57.9 |
| $60,001-100,000 | 174 | 20.2 | 149 | 21.2 | 225 | 19.6 | 66.2 |
| >$100,000 | 267 | 31.0 | 230 | 32.8 | 341 | 29.7 | 67.6 |
| *Not stated/missing* | *270* | | | | | | |
| **Birth place** | | | | | | | |
| Australia | 837 | 72.8 | 516 | 73.7 | 836 | 72.8 | 61.7 |
| Other: English speaking | 193 | 16.8 | 128 | 18.3 | 193 | 16.8 | 66.3 |
| Non-English speaking | 120 | 10.4 | 56 | 8.0 | 119 | 10.4 | 47.1 |
| **Residence** | | | | | | | |
| Metropolitan | 890 | 77.4 | 548 | 78.2 | 890 | 77.4 | 61.6 |
| Rural | 259 | 22.6 | 152 | 21.8 | 259 | 22.6 | 58.7 |
| **SEP (area)** | | | | | | | |
| Low | 263 | 22.9 | 162 | 23.2 | 263 | 22.9 | 61.6 |
| Low-mid | 206 | 17.9 | 116 | 16.5 | 205 | 17.9 | 56.6 |
| Mid | 180 | 15.6 | 112 | 16.1 | 180 | 15.7 | 62.2 |
| Mid-high | 233 | 20.3 | 159 | 22.8 | 233 | 20.3 | 68.2 |
| High | 267 | 23.3 | 150 | 21.5 | 267 | 23.3 | 56.2 |
| **Height** | | | | | | | |
| < 1.6m | 324 | 29.7 | 213 | 30.5 | 345 | 30.0 | 61.7 |
| 1.6-1.69m | 576 | 52.8 | 360 | 51.5 | 603 | 52.5 | 59.7 |
| 1.7+m | 191 | 17.5 | 126 | 18.0 | 200 | 17.4 | 63.0 |
| *Missing* | *57* | | | | | | |
| **Weight** | | | | | | | |
| Under or normal | 402 | 39.8 | 271 | 38.7 | 448 | 39.0 | 60.5 |
| Overweight | 330 | 32.6 | 245 | 34.9 | 371 | 32.3 | 66.0 |
| Obese | 227 | 27.5 | 184 | 26.4 | 319 | 28.7 | 57.7 |
| *Missing* | *139* | | | | | | |

***Table 4.1 continued***

| Factors | Total sample | | Complete case | | Multiply imputed | | % with complete data |
|---|---|---|---|---|---|---|---|
| | No. | % | No. | % | No. | % | |
| **Activity level (METs)** | | | | | | | |
| Low (<600METmin/wk) | 494 | 43.0 | 275 | 39.3 | 494 | 43.0 | 55.7 |
| Medium | 335 | 29.1 | 212 | 30.2 | 334 | 29.1 | 63.7 |
| High (>3000METmin/wk) | 321 | 27.9 | 213 | 30.4 | 320 | 27.9 | 66.6 |
| **Alcohol cons** | | | | | | | |
| Low (none) | 247 | 21.8 | 131 | 18.7 | 252 | 21.9 | 52.0 |
| Medium (<=7drks/wk) | 642 | 56.7 | 410 | 58.6 | 646 | 56.3 | 63.5 |
| High (>7drks/wk) | 244 | 21.6 | 159 | 22.7 | 250 | 21.8 | 63.6 |
| *Missing* | *15* | | | | | | |
| **Parity** | | | | | | | |
| No births | 90 | 7.8 | 55 | 7.8 | 90 | 7.8 | 61.1 |
| 1-2 birth | 614 | 53.6 | 397 | 56.8 | 614 | 53.5 | 64.7 |
| 3+ births | 443 | 38.6 | 248 | 35.4 | 444 | 38.7 | 55.9 |
| *Missing* | *4* | | | | | | |
| **Age first birth** | | | | | | | |
| <20yrs | 159 | 13.9 | 84 | 12.0 | 160 | 13.9 | 52.5 |
| 20-24yrs | 377 | 32.9 | 220 | 31.4 | 377 | 32.8 | 58.3 |
| 25-29yrs | 296 | 25.8 | 177 | 25.3 | 295 | 25.7 | 60.0 |
| 30+yrs | 225 | 19.6 | 165 | 23.5 | 226 | 19.7 | 73.0 |
| No births | 90 | 7.8 | 55 | 7.8 | 91 | 7.9 | 60.4 |
| *Missing* | *4* | | | | | | |
| **Family history BC** | | | | | | | |
| No relatives | 783 | 69.8 | 477 | 68.2 | 788 | 68.6 | 60.5 |
| 1 relative | 246 | 21.9 | 156 | 22.3 | 256 | 22.3 | 60.9 |
| 2+ relatives | 94 | 8.3 | 67 | 9.5 | 104 | 9.1 | 64.4 |
| *Missing* | *26* | | | | | | |
| **Prior biopsy** | | | | | | | |
| No | 887 | 77.2 | 520 | 74.3 | 884 | 77.0 | 58.8 |
| Yes | 262 | 22.8 | 180 | 25.7 | 264 | 23.0 | 68.2 |
| **Menopause age** | | | | | | | |
| <45yrs | 136 | 12.8 | 77 | 11.0 | 150 | 13.1 | 51.3 |
| 45-49yrs | 172 | 16.2 | 124 | 17.7 | 188 | 16.4 | 66.0 |
| 50-54yrs | 287 | 27.0 | 176 | 25.1 | 315 | 27.4 | 55.9 |
| 55+yrs | 109 | 10.2 | 67 | 9.6 | 124 | 10.8 | 54.0 |
| Not reached meno | 359 | 33.8 | 256 | 36.5 | 372 | 32.4 | 68.8 |
| *Missing* | *95* | | | | | | |
| **Menarche age** | | | | | | | |
| <12yrs | 155 | 14.0 | 96 | 14.0 | 161 | 14.0 | 59.6 |
| 12-14yrs | 753 | 68.2 | 474 | 63.9 | 782 | 68.1 | 60.6 |
| 15+yrs | 197 | 17.8 | 130 | 15.4 | 205 | 17.9 | 63.4 |
| *Missing* | *50* | | | | | | |

No. and per cent in each subcategory derived from survey-weighted data. Counts may not total due to rounding.

**RESULTS**

Demographic characteristics and risk factor profiles of survey participants are shown in Table 1 for complete case and multiply imputed data. On a case-wise basis, data were missing for 39% of our survey participants. Case-wise missingness was highest among older women, those with least education and those born in non-English speaking countries. Missing responses were particularly high for questions on household income, height and weight, age at menopause and age at menarche. Profiles did not differ substantially between the complete-case and MI datasets.

Sixty-five percent of women aged 40-84 years had undergone mammography screening at BreastScreen. Based on univariate analysis with MI data (Table2), participation varied according to age ($p<0.001$), education ($p=0.018$), household income ($p<0.001$), height ($p=0.021$), BMI ($p=0.014$), age at first birth ($p=0.001$) and menopausal age ($p<0.001$). Women who had participated in BreastScreen were also much more likely to have used HRT ($p<0.001$) and had a breast biopsy or surgery ($p<0.001$). In multivariable analyses (Table 4.2), BreastScreen participants had lower odds of being born in a non-English speaking country (OR=0.50; 0.29-0.85), having a sufficient level of physical activity (OR=0.57; 0.39-0.83) and being 40-49 years (OR=0.09; 0.06-0.15), and higher odds of being overweight and obese (OR=1.57; 1.08-2.29). The strongest associations were for ever using HRT (OR=3.72; 2.26-6.13) and having had a breast biopsy or surgery (OR=2.22; 1.23-3.99). Results showed a curvilinear association with SEP, indicating lower participation among the lowest and highest SEP groups. No significant differences were found for household income, education level, height, age at first birth, and age at menopause, after simultaneous adjustment. Results from complete case analysis (not shown) differed only slightly from multiply imputed data, in that associations with body weight and country of birth were statistically significant in the latter but not the former analysis. In addition a larger effect size was observed for non-English speaking country of birth in the MI analysis.

**Table 4.2 BreastScreen participation according to demographic characteristics and breast cancer risk factors (South Australian women aged 40-84 years)**

| Factors | Total[a] | No. ever screened at BreastScreen | % ever screened at BreastScreen | p-value[b] | OR (adjusted)[c] | 95% CI | p-value |
|---|---|---|---|---|---|---|---|
| **All respondents** | 1148 | 748 | 65.1 | - | - | - | - |
| **Age group** | | | | | | | |
| 40-49yrs | 333 | 78 | 23.5 | | 0.09 | 0.06-0.15 | <0.001 |
| 50-59yrs | 323 | 248 | 76.8 | <0.001 | 1.00 | - | - |
| 60-69yrs | 263 | 225 | 85.5 | | 1.39 | 0.78-2.47 | 0.260 |
| 70-84yrs | 229 | 196 | 85.8 | | 1.69 | 0.92-3.08 | 0.088 |
| **Education** | | | | | | | |
| Less than high school | 170 | 127 | 74.5 | | 1.00 | - | - |
| Completed high school | 418 | 269 | 64.4 | 0.018 | 1.14 | 0.66-1.97 | 0.631 |
| Certificate/diploma | 348 | 226 | 64.8 | | 1.57 | 0.86-2.88 | 0.143 |
| Bachelor degree + | 212 | 126 | 59.5 | | 1.54 | 0.75-3.16 | 0.243 |
| **Income** | | | | | | | |
| <$30,000 | 335 | 262 | 78.2 | | 1.00 | - | |
| $30,001-60,000 | 247 | 170 | 68.9 | <0.001 | 0.98 | 0.52-1.86 | 0.955 |
| $60,001-100,000 | 225 | 123 | 54.5 | | 0.64 | 0.34-1.19 | 0.159 |
| $100,001+ | 341 | 193 | 56.7 | | 0.96 | 0.49-1.86 | 0.896 |
| Residence | | | | | | | |
| Metropolitan | 889 | 582 | 65.5 | 0.708 | 1.00 | - | - |
| Rural | 259 | 166 | 64.0 | | 0.74 | 0.46-1.20 | 0.223 |
| **Birth place** | | | | | | | |
| Australia | 836 | 536 | 64.1 | | 1.00 | - | - |
| Other: English Speaking | 193 | 135 | 69.7 | 0.515 | 1.03 | 0.62-1.70 | 0.920 |
| Non-English Speaking | 119 | 77 | 64.8 | | 0.50 | 0.29-0.85 | 0.011 |
| **SEP** | | | | | | | |
| Low | 263 | 161 | 61.1 | | 1.00 | - | - |
| Low-mid | 205 | 143 | 69.7 | | 1.99 | 1.09-3.66 | 0.026 |
| Mid | 180 | 119 | 66.3 | 0.606 | 1.47 | 0.86-2.53 | 0.161 |
| Mid-high | 233 | 147 | 63.1 | | 1.33 | 0.75-2.38 | 0.329 |
| High | 267 | 178 | 66.5 | | 1.19 | 0.69-2.07 | 0.533 |
| **Tall stature** | | | | | | | |
| Short/average | 948 | 634 | 66.9 | 0.021 | 1.00 | - | - |
| Tall (>1.7m) | 200 | 114 | 56.8 | | 0.88 | 0.54-1.42 | 0.588 |
| **BMI category** | | | | | | | |
| Underweight/normal | 448 | 268 | 59.9 | 0.014 | 1.00 | - | - |
| Overweight/obese | 700 | 479 | 68.4 | | 1.57 | 1.08-2.29 | 0.018 |
| **Physical activity** | | | | | | | |
| Adequate | 654 | 436 | 66.6 | 0.290 | 1.00 | - | - |
| Insufficient (<600 MET-min/week) | 494 | 311 | 63.2 | | 0.57 | 0.39-0.83 | 0.003 |
| **Alcohol risk** | | | | | | | |
| Low | 898 | 590 | 65.7 | 0.484 | 1.00 | - | - |
| At risk (>7drks/week) | 250 | 157 | 63.0 | | 1.06 | 0.68-1.64 | 0.810 |

*Table 4.2 continued*

| Factors | Total[a] | No. ever screened at BreastScreen | % ever screened at BreastScreen | p-value[b] | OR (adjusted)[c] | 95% CI | p-value |
|---|---|---|---|---|---|---|---|
| **HRT use** | | | | | | | |
| Never | 825 | 452 | 54.8 | <0.001 | 1.00 | - | - |
| Ever | 323 | 295 | 91.4 | | 3.72 | 2.26-6.13 | <0.001 |
| **Family history** | | | | | | | |
| No | 788 | 506 | 64.2 | 0.444 | 1.00 | - | - |
| Yes (1+ female relative) | 360 | 242 | 67.1 | | 1.35 | 0.92-1.98 | 0.124 |
| **Breast biopsy** | | | | | | | |
| No | 943 | 584 | 61.9 | <0.001 | 1.00 | - | - |
| Yes | 205 | 164 | 79.9 | | 2.22 | 1.23-3.99 | 0.008 |
| **Menopausal age** | | | | | | | |
| <55yrs[d] | 1019 | 634 | 62.2 | <0.001 | 1.00 | - | - |
| 55+yrs | 129 | 113 | 87.2 | | 1.57 | 0.78-3.15 | 0.201 |
| **Menarche age** | | | | | | | |
| 12+yrs | 987 | 637 | 64.5 | 0.348 | 1.00 | - | - |
| <12yrs | 161 | 111 | 69.1 | | 1.00 | 0.55-1.85 | 0.991 |
| **Nulliparous** | | | | | | | |
| No | 1058 | 692 | 65.4 | 0.514 | 1.00 | - | - |
| Yes | 90 | 56 | 61.9 | | 0.77 | 0.44-1.37 | 0.378 |
| **1st Birth age** | | | | | | | |
| <30yrs (or no birth) | 964 | 651 | 67.5 | 0.001 | 1.00 | - | - |
| >=30yrs | 184 | 97 | 52.7 | | 0.99 | 0.60-1.63 | 0.975 |

[a] counts may not total due to rounding

[b] p-values for differences across subcategories derived from separate unadjusted logistic regression models with individual variables using MI data

[c] adjusted odds ratio (OR) derived from weighted multivariate logistic regression analysis with MI data with all variables modelled simultaneously

[d] includes premenopausal if <55yrs old

**Table 4.3 Participation in private screening (outside of BreastScreen) according to demographic characteristics and breast cancer risk factors (SA women 40-84yrs)**

| Factors | Total[a] | No. ever screened privately | % ever screened privately | p-value[b] | OR (adjusted)[c] | 95% CI | p-value |
|---|---|---|---|---|---|---|---|
| **All respondents** | 1130 | 270 | 23.9 | - | - | - | - |
| **Breast Screen** | | | | | | | |
| Non-participants | 398 | 100 | 25.2 | 0.505 | - | - | - |
| Participants | 732 | 170 | 23.2 | | - | - | - |
| **Age** | | | | | | | |
| 40-49yrs | 330 | 71 | 21.4 | | 0.86 | 0.55-1.36 | 0.502 |
| 50-59yrs | 320 | 80 | 25.1 | 0.648 | 1.00 | - | - |
| 60-69yrs | 260 | 71 | 27.2 | | 1.07 | 0.66-1.73 | 0.786 |
| 70-84yrs | 220 | 48 | 21.8 | | 0.90 | 0.53-1.54 | 0.710 |
| **Education** | | | | | | | |
| < Secondary school | 167 | 32 | 18.9 | | 1.00 | - | - |
| Secondary school | 408 | 82 | 20.2 | <0.001 | 1.11 | 0.65-1.87 | 0.710 |
| Certificate or diploma | 345 | 86 | 24.9 | | 1.36 | 0.78-2.40 | 0.279 |
| Bachelor degree | 211 | 70 | 33.1 | | 2.08 | 1.10-3.92 | 0.024 |
| **Income** | | | | | | | |
| <=$30,000 | 323 | 74 | 23.0 | | 1.00 | | |
| $30,001-60,000 | 245 | 50 | 20.6 | 0.168 | 0.79 | 0.47-1.33 | 0.378 |
| $60,001-100,000 | 225 | 51 | 22.5 | | 0.85 | 0.44-1.62 | 0.613 |
| >$100,000 | 338 | 95 | 28.0 | | 0.90 | 0.50-1.62 | 0.716 |
| **Birth place** | | | | | | | |
| Australia | 824 | 191 | 23.2 | | 1.00 | - | - |
| Other: English speaking | 188 | 45 | 23.8 | 0.279 | 1.04 | 0.67-1.62 | 0.846 |
| Non-English speaking | 119 | 34 | 28.8 | | 1.27 | 0.76-2.12 | 0.352 |
| **Residence** | | | | | | | |
| Metropolitan | 878 | 219 | 24.9 | 0.169 | 1.00 | - | - |
| Rural | 252 | 51 | 20.4 | | 0.96 | 0.62-1.48 | 0.840 |
| **SEP (area)** | | | | | | | |
| Low | 256 | 43 | 16.8 | | 1.00 | - | - |
| Low-mid | 201 | 42 | 20.8 | | 1.28 | 0.75-2.16 | 0.362 |
| Mid | 178 | 43 | 23.9 | <0.001 | 1.53 | 0.89-2.64 | 0.124 |
| Mid-high | 230 | 64 | 28.0 | | 1.73 | 1.02-2.92 | 0.042 |
| High | 266 | 78 | 29.4 | | 1.70 | 1.01-2.88 | 0.047 |
| **Tall stature** | | | | | | | |
| Short/average | 993 | 242 | 24.4 | 0.374 | 1.00 | - | - |
| Tall (>1.7m) | 197 | 42 | 21.1 | | 0.78 | 0.49-1.26 | 0.318 |
| **BMI category** | | | | | | | |
| Underweight/normal | 437 | 111 | 25.5 | 0.375 | 1.00 | - | - |
| Overweight/obese | 693 | 159 | 22.9 | | 0.92 | 0.71-1.39 | 0.969 |
| **Physical activity** | | | | | | | |
| Adequate | 647 | 161 | 24.9 | 0.410 | 1.00 | - | - |
| Insufficient (<600 MET-min/week) | 483 | 109 | 22.5 | | 0.92 | 0.66-1.28 | 0.623 |

*Table 4.3 continued*

| Factors | Total[a] | No. ever screened privately | % ever screened privately | p-value[b] | OR (adjusted)[c] | 95% CI | p-value |
|---|---|---|---|---|---|---|---|
| **Alcohol risk** | | | | | | | |
| Low | 884 | 209 | 23.6 | 0.698 | 1.00 | - | - |
| At risk (>7drks/week) | 246 | 61 | 24.9 | | 1.03 | 0.70-1.52 | 0.865 |
| **HRT use** | | | | | | | |
| Never | 815 | 181 | 22.2 | 0.048 | 1.00 | - | - |
| Ever | 315 | 89 | 28.2 | | 1.24 | 0.87-1.79 | 0.237 |
| **Family history** | | | | | | | |
| No | 780 | 164 | 21.0 | 0.003 | 1.00 | - | - |
| Yes (1+ female relative) | 350 | 105 | 30.1 | | 1.66 | 1.19-2.31 | 0.003 |
| **Breast biopsy** | | | | | | | |
| No | 942 | 180 | 19.4 | <0.001 | 1.00 | - | - |
| Yes | 203 | 90 | 44.3 | | 3.17 | 2.20-4.55 | 0.001 |
| **Menopausal age** | | | | | | | |
| <55yrs[d] | 1003 | 240 | 23.9 | 0.917 | 1.00 | - | - |
| 55+yrs | 127 | 30 | 23.5 | | 0.82 | 0.49-1.38 | 0.460 |
| **Menarche age** | | | | | | | |
| <15yrs | 972 | 236 | 24.3 | 0.462 | 1.00 | - | - |
| 15+yrs | 158 | 33 | 21.3 | | 0.84 | 0.51-1.39 | 0.492 |
| **Nulliparous** | | | | | | | |
| No | 1041 | 253 | 24.3 | 0.215 | 1.00 | - | - |
| Yes | 89 | 17 | 18.5 | | 0.64 | 0.36-1.13 | 0.124 |
| **1st Birth age** | | | | | | | |
| <30 (+ no birth) | 946 | 219 | 23.1 | 0.235 | 1.00 | - | - |
| >=30yrs | 183 | 51 | 27.7 | | 1.12 | 0.73-1.73 | 0.611 |

[a] counts may not total due to rounding

[b] p-values for differences across subcategories derived from separate unadjusted logistic regression models with individual variables using MI data

[c] adjusted odds ratio (OR) derived from weighted multivariable logistic regression analysis with MI data

[d] includes premenopausal if <55yrs old

**Table 4.4 Comparison of summary measures of breast cancer risk by participation in BreastScreen or other private screening service**

| Summary risk measures | BreastScreen | | | Other private screening | | |
|---|---|---|---|---|---|---|
| | Participant | Non-participant | | Participant | Non-participant | |
| | Mean (95%CI) | Mean (95%CI) | p-value[d] | Mean (95%CI) | Mean (95%CI) | p-value[d] |
| **Total number of risk factors**[a] (count of: At-risk alcohol, overweight/obese, tall stature, inactive, HRT use, nulliparious, late first birth, early menarche, late menopause, family history breast cancer, breast biopsy, high socioeconomic position) | | | | | | |
| 40-84yrs | 3.10 (2.99-3.21) | 2.61 (2.46-2.76) | <0.001 | 3.20 (3.02-3.38) | 2.85 (2.74-2.95) | 0.001 |
| 50-69yrs | 3.21 (3.07-3.35) | 2.70 (2.44-2.97) | 0.004 | 3.37 (3.12-3.60) | 3.04 (2.89-3.19) | 0.022 |
| **Gail risk prediction model**[b] Age at first birth, age at menarche, breast biopsy, family history breast cancer | | | | | | |
| 40-84yrs | 2.22 (2.06-2.37) | 2.17 (2.00-2.35) | 0.728 | 2.47 (2.25-2.69) | 2.12 (1.98-2.26) | 0.009 |
| 50-69yrs | 2.20 (2.04-2.35) | 2.30 (1.96-2.64) | 0.603 | 2.50 (2.21-2.80) | 2.12 (1.97-2.28) | 0.025 |
| **Pfeiffer risk prediction model**[c] BMI, Alcohol use, HRT, age at first birth, parity, menopausal age, breast biopsy, family history breast cancer | | | | | | |
| 40-84yrs | 2.21 (2.15-2.27) | 1.95 (1.88-2.01) | <0.001 | 2.43 (2.32-2.54) | 2.02 (1.97-2.07) | <0.001 |
| 50-69yrs | 2.25 (2.17-2.32) | 1.91 (1.79-2.04) | <0.001 | 2.50 (2.36-2.64) | 2.08 (2.01-2.15) | <0.001 |

[a] mean total number of listed risk factors

[b] mean risk score based on adapted Gail breast cancer risk prediction model – excluding hyperplasia and combines 1 or more biopsy rather than specifying the number of biopsies

[c] mean risk score based on Pfeiffer breast cancer risk prediction model, replacing family history of breast cancer for family history of breast or ovarian cancer

(risk scores above assume all women have the same baseline risk of breast cancer irrespective of age)

[d] p-values derived from regression models of mean score (or total number of risk factors) by screening attendance, adjusted for age, using multiply imputed datasets

No. in models for Breast Screen participation: (40-84yrs) = 1148; (50-69yrs) = 587

No. in models for participation in private screening; (40-84yrs) = 1130; (50-69yrs) = 581

Similar analyses were undertaken to examine differences in risk factor profiles in relation to use of private mammography screening services (Table 4.3). Screening outside of BreastScreen was reported by 24% of women aged 40-84 years. The proportion screening privately was similar for BreastScreen participants and non-participants (25.2% vs 23.2%, p=0.505). From multivariable analyses of MI data, those who screened privately had higher odds of having a bachelor degree (OR=2.08; 1.10-3.92), being from the most socioeconomically advantaged neighbourhoods (OR=1.70, 1.01-2.88), having a family history of breast cancer (OR=1.66; 1.19-2.31) and having had breast biopsy or surgery (OR=3.17; 2.20-4.55). No differences were observed by age, income level, place of residence, country of birth, or any lifestyle or reproductive risk factors. Findings were similar for complete case analysis.

Differences in summary measures of breast cancer risk according to screening participation are shown in Table 4. Both and the mean number of risk factors and the mean risk score, based on the Pfeiffer model, differed significantly between BreastScreen participants and non-participants, as well as between women who had attended screening outside BreastScreen compared with those who had not. Risk scores, based on the Gail model, did not differ significantly between BreastScreen participants and non-participants, but were higher for women who participated in private screening compared with those who had not. These results applied to all women (40-84yrs) as well as those within the target age range (50-69yrs).

## DISCUSSION

This study indicates that, for South Australian women, breast cancer risk factor profiles differ between those who have and have not participated in population-based mammography screening. Participation was less likely among women younger than 50 who are not actively targeted by BreastScreen, those born in non-English speaking countries and those who were not sufficiently active, while the likelihood of having used HRT and of having experienced breast biopsy or surgery was much more likely among BreastScreen participants. Contrary to expectations, BreastScreen participants were more likely to have been overweight or obese, and no statistically significant differences were noted for family history of breast cancer.

The association between HRT use and BreastScreen participation may be due to increased contact with medical practitioners (to obtain or renew prescriptions). Due to concerns about increased risk of breast cancer many doctors may have recommended screening to patients using HRT. There is consistent evidence that a doctor's recommendation is a strong predicator of screening attendance[164].

As observed in other studies[171], a history of breast biopsy or surgery was associated with screening participation. However, determining the causal direction is difficult given the cross-sectional design of this study. Women who attend BreastScreen would have a greater chance of biopsy due to investigation of abnormal mammographic lesions.

While our finding of increased participation among overweight and obese women contradicts other studies showing lower participation[177, 187, 188], it is consistent with findings from a recent US study[189]. One explanation may be that lighter women with smaller breasts feel more confident in detecting breast abnormalities through self-examination and therefore see less benefit from attending mammography screening.

Also in contrast to other studies[171], including a previous South Australian study[48], we did not observe any association between family history of breast cancer and participation in population-based screening. Additional analyses using different post-hoc definitions of family history (e.g. two or more close female relatives diagnosed at any age and/or one relative diagnosed before age 50) also failed to show any statistically significant differences, though odds were consistently elevated.

Undergoing mammography screening outside of BreastScreen, however, was associated with family history of breast cancer, as well as higher education and higher SEP. There was also a strong association between private screening and breast biopsy/surgery. It is possible that some women confused screening and diagnostic mammography, which may explain the greater odds of biopsy/surgery among those who reported screening privately. Alternatively those with prior breast problems may seek additional security by screening more frequently using publically and private services. This may also apply to women with a family history of breast cancer. Additionally women with some premalignant diagnoses, such as atypical ductal hyperplasia, may be discharged from population-based screening programs as a

matter of policy since their risk exceeds population level risks. The association with higher levels of education and SEP suggests those with financial resources may also seek additional security through private screening. We did not find evidence that BreastScreen non-participants were more (or less) likely undergo mammography elsewhere.

Taken together our findings suggest participants of population-based mammography screening have a slightly higher risk of developing breast cancer than non-participants. Both HRT use and being overweight are associated with a moderately increased risk of breast cancer[110, 190]. Likewise, increased likelihood of breast procedures may indicate a higher prevalence of premalignant abnormalities, and potentially greater risk of breast cancer[191]. Furthermore, women from non-English speaking countries (who participate less) tend to have lower risk of breast cancer[192]. Only the observed differences in physical activity levels would potentially lead to reduced risk among screening participants[193].

Comparison of the summary risk measures supports this conclusion. While we did not find statistically significant differences in risk scores based on the Gail model, this is not unexpected given there was no difference between BreastScreen participants and non-participants for family history and age at first birth, on which the Gail model is largely based. However, the total count of risk factors did differ, given it allows for a broader number of risk factors to be considered.

This study has a number of limitations. Measures of both screening attendance and risk factors may be inaccurate, due to our reliance on self-reported data. Recall and response bias may also have led to erroneous findings. Since BreastScreen is a well-recognised service we believe that recall of attendance at BreastScreen would have been reasonably accurate. However, participants may have over-reported attendance at private screening services if they did not distinguish between screening and diagnostic mammography. This study did not include all probable risk factors (e.g. breastfeeding, oral contraceptive use) or explore potential risk factors where evidence is less clear (e.g. tobacco smoking, fat intake). Nor did it account for interaction between certain risk factors (e.g. effect modification by BMI on HRT effects[194]). Also the cross-sectional design does not allow causal inferences to be determined. This is

particularly pertinent to disentangling whether increased breast procedures result from, or are a motivation for, screening participation. Finally, our summary measures of risk were relatively crude. We used a modified version of the Gail model which did not included hyperplasia or multiple biopsy procedures, and has not been validated in our population. Therefore, results should be considered indicative rather than definitive. We were unable to apply more sophisticated risk models that incorporated a broader range risk factors[195], due to lack of data (e.g. regarding genetic mutations, detailed family history, dietary patterns) or because some measures were inappropriate for this study (e.g. breast density).

The strengths of this study are that we collected a range of risk factor data from a randomly selected sample of women who are representative of the eligible screening population for screening and used multiple imputation to increase precision and reduce potential biases due to the exclusion of a large proportion of cases with missing data.

**CONCLUSIONS**

Our findings suggest that South Australian women who participate in population-based screening have a slightly worse risk factor profile than those who do not. For the most part this can be attributed to greater HRT use among BreastScreen participants. These findings may not be generalizable to other settings, given variations in risk factor prevalence and/or organisation of screening services in different countries. This highlights the need to understand differences in breast cancer risk profiles within the local context, especially in relation to evaluating benefits and risks of mammography screening including over-diagnosis.

**ACKNOWLEDGEMENTS**

We wish to thank Ms Lora Dal Grande and Mr Graeme Tucker for their assistance in relation to survey weighting. We also wish to acknowledge the contribution of women who took part in the 2012 HOS survey. This study was funded by the National Breast Cancer Foundation through a Novel Concept Award (#NC-11-53). Ms Beckmann is the recipient of a NHMRC PhD scholarship.

Authors have no conflict of interest.

# Chapter 5: Measuring over-diagnosis due to mammography screening: A review of study designs and findings

## INTRODUCTION

Over-diagnosis is difficult to measure because essentially it is "unobservable". For the most part, there is conceptual agreement about what is meant by over-diagnosis[196]. Over-diagnosis is commonly defined as the detection of cancer at screening (histologically confirmed IBC or DCIS), that would never have surfaced clinically in a woman's life time in the absence of screening[57]. Since currently there are no definitive tumour markers or clinical tests to determine which screen-detected cancers would and would not have progressed, or progressed sufficiently for clinical diagnosis within a woman's life time, it is not possible to directly observe which individual cases are 'over-diagnosed'. Therefore, over-diagnosis is only inferred at a population level by comparing cancer incidence in the presence and absence of screening[196]. Even then, outside of randomised controlled trial setting of screening, over-diagnosis is not easily measured, since the number of cancers that would occur in the absence of screening must be estimated from rates either in a different population or at a different period in time.

As outlined in Chapter 1, estimates of over-diagnosis due to mammography screening have varied considerably across studies (with reported levels ranging from 0-54%). Differences in estimates may reflect, to some extent, differences in screening and population characteristics across regions[197]. For example, greater levels of participation in screening and shorter screening intervals would increase the chances of detection of irrelevant cancers, while less inclusive diagnostic criteria would lead to less frequent recall and fewer non-progressive cancers being detected. Likewise the age at which screening starts and finishes may affect the level of over-diagnosis since tumour growth rates differ across age groups[197]. However, the variation across studies cannot be adequately explained by local factors, given that substantial disparities are seen across studies conducted within the same countries at similar times. Differences

in how over-diagnosis is measured and reported may offer more likely explanations for the extremely variable findings across studies.

This chapter will focus on critically reviewing evidence for breast cancer over-diagnosis, with emphasis on original studies and systematic reviews which explicitly aimed to measure the extent of over-diagnosis due to population-based mammography screening. The first section describes the three general approaches used to estimate over-diagnosis. The second section discusses the key methodological concepts that need to be considered in relation to study design and interpretation of findings. The final section examines and critiques previously published studies reporting estimates of breast cancer over-diagnosis. The underlying questions being addressed in this chapter are: **why do estimates of over-diagnosis vary to such an extent, and which estimates are more likely to be valid?**

This review has been heavily influenced by previous methodological reviews by Biesheuval et al. 2007[198], Duffy et al. 2008[59], Puliti et al. 2011[57], and de Gelder et al. 2011[197], Puliti et al. 2012[89]. Original studies that report measures of over-diagnosis were sourced through searching medical and health related databases including PubMed, Scopus and Google Scholar, using the following search terms: over-diagnosis, over-detection, mammograph*, mass screening, breast cancer, breast neoplasm, mammary. Search restrictions included studies involving humans and English only publications.

Reference lists from identified studies were also searched for other potential studies not identified through online database searches. This review includes studies published to June 2014.

## STUDY DESIGNS

Authors of the recent review of the UK Mammography Screening Program[39] have stated that the ideal study design to measure over-diagnosis has not and cannot be undertaken and therefore estimates need to be inferred using other, less ideal study designs.

The 'ideal' study would be a large randomised controlled trial of current mammography screening technologies and protocols, with perfect randomisation,

100% compliance with screening protocols, no uptake of screening among participants in the control arm, and complete follow-up of all subjects until death. The difference in total numbers of cancers in the two arms of the trial would provide a valid estimate of the level of over-diagnosis due to screening. In this 'ideal study', no adjustment for risk differences would be necessary since randomisation would have ensured that apart from chance, background risks of breast cancer were equivalent in the screened and unscreened groups. Even then, there could be bias in the absence of "double blinding", which is likely to be unachievable in the context of mammography screening. Likewise adjustment for lead time (i.e. the time that cancer diagnoses are advanced due to screening) would not be necessary, providing that both arms were followed throughout women's lifetimes. Such a trial is neither practical nor feasible, especially as mammography screening is already so widely available, nor would it provide a timely answer to the question of over-diagnosis since follow-up would need to continue for extensive periods (i.e. up to 30 years or more), given life expectancies of greater than 80 years in most economically developed countries with mammography screening programs.

Long term follow-up of the original randomised screening trials, comparing cumulative incidence among those invited to screen and the control group who were not invited, would appear to offer the best alternative option available from a research design perspective. The obvious advantage of this approach is that no adjustment would be required for differences in breast cancer risk or lead time effects, provided satisfactory randomisation, no biasing effects due to lack of double blinding, and sufficiently long follow-up. However, not all of the original trials were considered to be adequately randomised, due to inadequacies in the randomisation process or as a result of exclusion of selected participants due to prior breast cancer diagnoses[39, 70, 199]. Furthermore, several of the trials offered screening to the control group at the end of the trial and therefore do not allow for direct estimates of over-diagnosis[196]. Even in the trials that did not offer screening to controls, it is possible that some participants undertook screening through other avenues after the close of the trials, which could maintain a lead time effect and bias results of follow-up assessment.

Questions have also been raised about the generalisability of findings from the original screening trials to current screening programs[89]. Most of the trials were conducted in the 1970's under experimental conditions, which do not mirror screening protocols in most countries (i.e. biennial screening over a 20 year period). Furthermore, screening sensitivity has improved considerably over the past few decades, which may impact on the extent of over-diagnosis.

An alternative to long term follow-up of screening trials is to measure over-diagnosis by comparing the observed breast cancer incidence in a cohort or population that has been offered screening with incidence in a cohort/population that has not had the opportunity to participate in screening. These 'observational' studies may use either aggregated population data or, where it is possible to link different data sources, individual level data pertaining to screening participation and breast cancer outcomes. In a cohort approach with individual level data, adjustment needs to be made for any differences in the risk of developing breast cancer between screening participants and non-participants. The length of time that a cohort has been followed after they stop screening will affect the estimates. Longer follow-up periods allow for lead time effects of screening to dissipate while shorter periods would result in inflated estimates of over-diagnosis due to lead time bias. (Lead time is the time between when a cancer is detected at screening and when it would have been diagnosed on the basis of clinical symptoms had no screening occurred, and is a necessary feature of screening to realise the goal of early detection.) In studies that use population data, there is usually no obvious 'unexposed' comparison group, so incidence without screening must be estimated from a different population or from another time period. Adjustment is again required for any differences in background incidence between regions or across time so that the population groups are equivalent in terms of underlying risk of breast cancer. Failure to adequately adjust for the confounding by risk differences or for lead time effects in observational studies will lead to biased estimates of over-diagnosis.

In addition to observational and experimental studies, various modelling approaches have also been used to estimate the level of over-diagnosis due to mammography. Most of the modelling approaches rely on estimating the number of cancers likely to have been diagnosed at an earlier time through screening, based on estimated

measures of lead time length and its assumed distribution, along with other parameters such as screening sensitivity, participation rates, and tumour progression rates. Lead time and other screening parameters are usually estimated from trial or screening program data based on which values give the best fit with observed data. The various screening parameters are then applied to simulated or observed incidence data to predict cancer outcomes in the presence or absence of screening. While the complexity of models can vary considerably, all modelling studies rely on assumptions about the screening parameters being measured or the natural history of breast cancer, which cannot be directly verified.

## METHOLOGICAL CONSIDERATIONS

There are three key factors that need to be taken into account in relation to evaluating study designs to measure over-diagnosis. These include:

- whether lead time effects have been taken into account and adequate adjustment made for these effects,
- whether the comparison groups or populations have equivalent underlying risks of breast cancer,
- and how measures are calculated and reported.

Other design issues that can affect study findings include whether the measures were derived using aggregate or individual level data on exposure and outcome, whether the study was undertaken within a dynamic population or "closed" cohort of women, and when the study was conducted relative to the start of the screening program.

Each of the above mentioned factors will be discussed in the following section in relation to the various research approaches used to estimate over-diagnosis due to mammography screening (i.e. experimental trials, observational studies and modelling approaches).

### Lead time effects

All study designs need to account for, or adjust for, any excess in breast cancer incidence due to lead time effects[56]. Incidence will be elevated among women who screen because screening shifts the date of diagnosis forward for screen-detected cases. This elevation will continue in the screening cohort until screening ceases, at

which point incidence should drop to levels lower than those among women who have not screened, since a large proportion of cancers has already been detected during the screening period[89]. Screening essentially shifts the age distribution with regard to cancer diagnoses, toward higher incidence in the younger (screening) age groups and lower incidence in the older (post-screening) age groups, among those who participate[57].

If no over-diagnosis occurs the total number of cancers over the long term ought not to vary in comparable groups of women who have and have not participated in screening[198]. Lead time effects can therefore be accounted for by following populations for long enough after screening has stopped so that the reference (non-screened) population can catch-up in terms of the total number of breast cancers diagnosed. With no over-diagnosis, the cumulative incidence should be equal in both the screened and comparison populations. Any excess in cancers in the screened group would indicate the extent of over diagnosis (provided the groups are comparable in all other ways). This is often referred to as the cumulative incidence approach[198]. No additional lead time adjustment is necessary provided the follow-up period is adequate. What constitutes adequate follow-up time in both experimental and observational studies remains unclear, and is dependent on lead time length and its distribution in the population. Some commentators believe 5 years is sufficient[89, 198] while others have indicated periods of longer than 10 years are required[117, 197]. Based on an estimated average lead time of 3.7 years with an exponential distribution, de Gelder et al[197] maintain that 20% of cancers detected at screening have a lead time greater than 5 years, and 5% have a lead time greater than 10 years. Thus studies with less than 10 years follow-up after completion of screening may be overstating the extent of over-diagnosis since they have not accounted fully for lead time effects.

Another commonly used approach to adjust for lead time is the compensatory drop method[57, 198], sometimes referred to as the deficit method[56] or excess incidence method[80]. In this approach over-diagnosis is estimated by subtracting any observed deficit in cancer incidence in the age group who have completed screening from the observed excess in incidence among screening-aged women. This method can only be used in populations or cohorts where most, if not all, of the older age group has had

the opportunity to participate in screening[56]. Thus the timing of the study relative to the implementation of the screening program, as well as the extent of uptake among women before reaching the maximum age, will impact on estimates[57]. In essence, the compensatory drop method is similar to the cumulative incidence method in that it is based on the total incidence during the active screening period and the post-screening period. The main difference is that in the compensatory drop method over-diagnosis is usually calculated cross-sectionally in a dynamic population by summing age-specific incidence rates at a particular point in time rather than by following a defined cohort over time.

An alternative method to adjust for lead time in observational studies, which has been used to a lesser extent, is the incidence rate shift method[198]. This approach supposes that, in a 'stable' screening program, the incidence rate in a screened population will be equivalent to the rate that would be expected without screening, shifted by the average lead time length. For example, if the average screening lead time is 3 years (i.e. screening advances the diagnosis of breast cancer by 3 years on average) then incidence rates among those who are screened should be the same as those in three years' time, among women three years older, who hadn't been screened. It should be noted that the rate shift method ought to include adjustment for the advancement of diagnosis both temporally and by age, which has not always been the case[59, 89]. One of the criticisms of this approach is that the first few screening rounds following the introduction of a screening program are deliberately excluded from the analysis. Critics have contended that estimates of over-diagnosis using this method understate the extent of over-diagnosis because cancers detected at the 'prevalence' screening round are not included[200]. However, this criticism is not valid in relation to studies undertaken in dynamic populations with mature screening programs, since such studies will include women who have undergone or are undergoing their first mammogram.

A more important question relates to the validity of the claim that the shift in incidence rates during the stable phase will be equivalent to the average lead time length. Although there have been no empirical studies to support this claim, the work of Duffy and Parmar[117] provides indirect evidence that the excess incidence due to

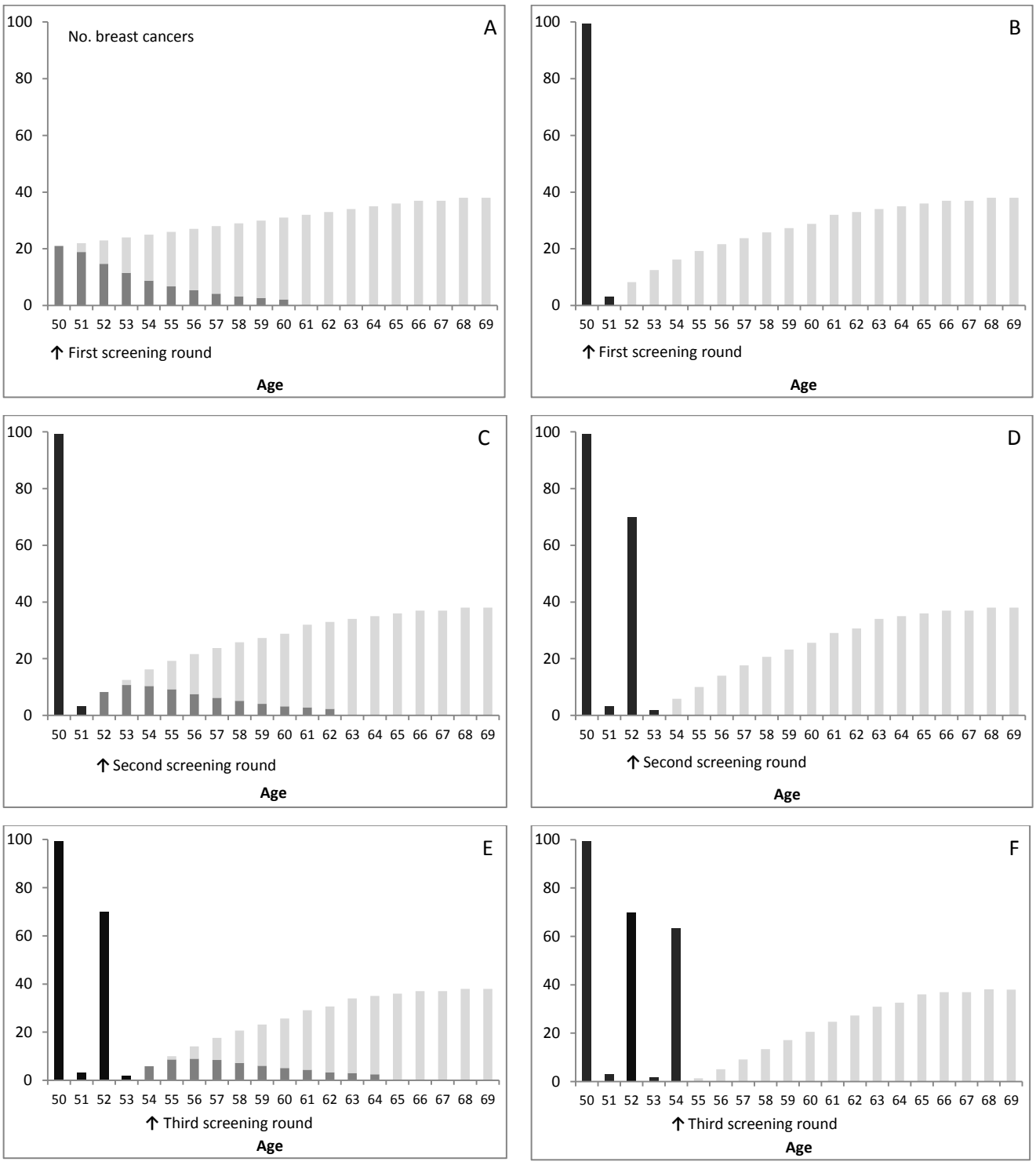
lead time effects is likely to be higher than would be achieved through the rate shift method even after programs are stable. This is due mainly to the dynamic, ongoing nature of screening whereby first time screening participants are continually being recruited into the program. From data provided in Tables 2 and 3 in the paper by Duffy and Parmar[117] it is possible to calculate the incidence rate ratio due to lead time effects during the stable phase of screening for women aged 50-69 years. In the screening scenario presented, the incidence rate ratio (IRR) during the stable screening period compared with background rates is 1.22. By comparison, shifting incidence rates forward by 3 years results in an IRR of 1.06 compared with background incidence rates. (IRR=1.10 for a 5-year shift). The rate shift method is therefore likely to lead to inflated estimates of over-diagnosis due to inadequate adjustment for lead time.

Figure 5.1 shows how lead time effects alone can substantially change incidence rates in a hypothetical cohort of 10,000 women attending biennial screening from age 50 with an average lead time of 40 months (3.7 years), assuming an exponential distribution. (The validity of this assumption is discussed later.) Graph A shows the expected incidence over time in the cohort prior to any screening. The mid-grey shading represents the cancers that would be diagnosed at the first round of screening (if all women participated and sensitivity of screening was 100%). In addition to the cancers expected to occur in the index year, a large proportion of cancers that were destined to have presented clinically in the next year would have their diagnosis date advanced through screening, along with a substantial proportion the following year and diminishing proportions in subsequent years. The additive effect of bringing forward cancers that would have presented over several consecutive years results in a greatly elevated incidence rate at the time of screening (shown in black shading [Graph B]). Since cancers that have their diagnosis date advanced by screening are removed from the pool of future cancers within the population, subsequent rounds of screening will detect fewer cancers than in the first round. Graph C shows the redistribution of breast cancer incidence by age after one round of screening, with mid-grey shading indicating the proportion of cancers in future years that are likely to have their diagnosis advanced during the second round of screening. The resulting redistribution after the second round of screening is shown in Graph D. Graphs E and F illustrate the effect at the third round of screening. The cumulative effect of advancing the diagnosis

of remaining future cancers results in substantially higher incidence rates compared with the original background rate had no screening occurred at the first round and during subsequent rounds as illustrated. Such a pattern would continue with ongoing screening rounds. In a dynamic population, lead time effects would result in continually elevated rates as new cohorts of women commenced screening. This figure serves to illustrate how the relative increase in incidence would be considerably higher than the background rate for women three or even five years older in the absence of screening.

Statistical modelling approaches have also been used to adjust for lead time effects[57, 197, 198]. The complexity of these models can vary considerably. Generally, this approach uses multistate models to test which combination of screening parameters gives the best fit with the observed data. Several reviews have noted that, in many cases, methods and assumptions are not transparent, and hence it is very difficult to assess the validity of their findings[198].

**Figure 5.1 Impact of lead time effects on age-specific breast cancer incidence in a hypothetical cohort of 10,000 women undergoing biennial mammography screening over the first 3 screening rounds, assuming 40 months lead time, 100% participation and 100% screening sensitivity**




Mid-grey shading represents the number of cancers that would be brought forward at each round of screening due to lead time effects; dark shading represents the number of cancers diagnosed in each year of screening after adjustment for lead time effects (number expected in the index year plus future cancers brought forward though screening); light-grey shading represents the numbers of remaining cancers that would be expected to be diagnosed each year as the cohort ages from ages 50-69yrs with the impact of previous screening taken into account.

Another common criticism is that modelling methods rely on assumptions about the lead time length and its distribution, which cannot be verified empirically since lead time (like over-diagnosis) is unobservable[89, 198, 200]. Most modelling studies have assumed a negative exponential distribution for lead time. This assumption is consistent with the notion that tumour growth in the early stages is determined by the rate of cell division which results in an exponential growth curve over the preclinical period. This assumption was validated in the early work by Day and Walter[201] comparing different lead time distributions (which included an exponential, a log normal and a stepwise distribution). They showed that an exponential distribution gave the best fit when modelling lead time effects for breast cancer screening data. More recently a paper by Weedon-Frekjaer et al. (2008) demonstrated a better fit with a logistic tumour growth model[202]. A logistic growth model is consistent with the work of Spratt et al[203] who showed that, while growth rates follow an exponential growth curve in the early stages of tumour development, they tend to slow as the tumour gets larger, probably in response to increased nutritional requirements of the tumour. However, in relation to tumours that are reaching the size when they would become screen-detectable (5-10mm), the logistic distribution approximates an exponential distribution. Therefore, exponential distribution of lead time is likely to be a reasonable assumption.

De Gelder et al.[197] have noted that modelling approaches may be limited because they derive estimates for multiple screening variables simultaneously. Interchanging some of these estimates may still provide a good fit with observed data, for example high rates of tumour progression coupled with low levels of tumour regression may simulate observed data equally as well as lower rates of progression and higher levels of regression. Zahl et al. 2014[200] claim that estimates of the lead time applied in these models are inflated by the inclusion of over-diagnosed cases. They base this assertion on the belief that lead time estimates are derived from screen-detection rates, which include over-diagnosed cases which have, by definition, very long lead times. This in turn leads to underestimation of over-diagnosis due to over-adjustment for lead time. However, in a number of modelling approaches allowance appears to have been made for some portion of non-progressive breast cancer cases[52, 88, 204, 205]. Hence this criticism is not likely to apply to all modelling approaches.

Consistency across the various models may also be problematic. Zahl et al. 2013[206] give the example of widely differing estimates of over-diagnosis, ranging from 23% to 43%, resulting from the application of different models to the same data set. Similarly, authors of the CISNET study[87], which examined the benefits and harms from screening using 5 different, independently developed models, declined to report precise figures for over-diagnosis due to variability and uncertainties regarding the natural history of breast cancer.

On the other hand, De Gelder et al.[197] believe that estimates for many of the screening parameters are reliable because they are derived from long standing screening programs or trials, where outcome data are accurately recorded. Puliti et al. 2012[89] also maintain that modelling methods are the most appropriate methods when there is insufficient post-screening follow-up to allow the use of cumulative incidence or compensatory drop methods.

**Comparable populations**

It is essential that estimates of over-diagnosis are based on comparison between groups that have an equivalent level of breast cancer risk, so that the only difference affecting incidence is exposure to screening[89, 198]. In the ideal experimental trial, the risk of breast cancer should be equivalent due to random allocation into the screening or non-screening arms of the trial. However, imbalances may be introduced if exclusion criteria, such as having had a prior breast cancer diagnosis, are applied after the randomisation process[207].

Several of the modelling studies have simulated the background incidence of breast cancer in an artificial reference population (reflecting realistic population data), then modelled the impact of screening based on estimates of screening parameters (e.g. sensitivity and lead time estimates)[86, 197, 205]. In such studies the reference and screening populations have the same background incidence by design so no adjustment for risk difference is necessary[89]. Risk adjustment is however important in modelling studies that compare lead time adjusted incidence with observed incidence (or with expected incidence in the absence of screening) based on actual population data [69, 78, 79].

Self-selection into screening programs can lead to bias due to risk differences and these need to be accounted for in observational studies that use individual level data on exposures and outcomes[89]. If for example there is a higher prevalence of family history of breast cancer or prior benign breast disease among women who participate in screening, comparisons between women who have and have not attended screening will not be based on equivalent risk. Estimates of over-diagnosis will therefore be biased without adequate adjustment for breast cancer risk factors such as HRT use, family history, parity etc. More often than not this is not possible since such data are rarely available for large cohorts of women.

In observational studies based on population level data there is no natural comparison group with identical risk which was not exposed to screening, so background incidence rates need to be inferred from other populations or time periods and adjustments made for any observed or assumed differences in underlying risk. Background incidence is often estimated from incidence rates in the study population prior to screening[56, 198]. Given that breast cancer rates were increasing prior to the implementation of screening, adjustment should be made for the anticipated temporal trends in the intervening period[198]. Failure to adjust for temporal trends, is likely to result in inflated estimates of over-diagnosis. Projection of temporal trends aims to capture the influence of factors other than screening (i.e. changing risk factor prevalence) and assumes trends in breast cancer incidence have continued to increase at the same rate as during the pre-screening period. The alternative approach is to base comparisons on incidence in neighbouring regions where screening has not been implemented. While adjustment for any underlying differences in incidence rates between regions is essential to obtaining valid estimates, it is difficult to verify whether pre-screening differences would have persisted over time, so estimates based on pre-screening geographical comparisons may not be reliable.

A variety of methods have been used to predict the background incidence rates in the absence of screening including simple linear regression, Poisson regression by age and period, and age-period-cohort modelling, with linear regression being the least appropriate and most imprecise[89]. Since background incidence cannot be directly observed, uncertainties remain as to whether trends (or regional differences) before

the advent of screening would have continued unchanged. It is possible that changes in risk factor prevalence have had a greater or lesser influence on background incidence rates than predicted through modelling pre-screening incidence trends[208]. One obvious example that may have altered background rates is the rapid uptake and subsequent decline in use of hormone replacement therapy.

Zahl et al. 2014 has argued that many researchers have assumed a stronger background incidence trend than is likely to be the case[200]. They believe that opportunistic screening is likely to be the cause of increasing incidence prior to the introduction of screening programs as well as in non-screening regions and in age groups outside the screening target age range. They also believe that the influence of changing patterns of HRT use was insubstantial, arguing that, while HRT use has decreased dramatically in recent years, there was no corresponding decrease in BC incidence and therefore initial increases in prevalence would not have led to increases in background incidence. Several studies have actually presented evidence for a decline in breast cancer incidence since the sudden decline in 2002[100, 101, 209], so this latter criticism is likely to be incorrect.

The UK independent panel review of breast screening[39] also highlighted problems in estimating background incidence and were of the opinion that neither projection of pre-screening incidence rates, nor geographical comparisons, were robust methods for determining background incidence. However, they offered no alternative approach for estimating background incidence rates in observational studies.

**Dynamic vs closed populations / aggregate vs individual measures**

Generally, study designs aimed at quantifying over-diagnosis have been applied to either dynamic populations (where women are continually entering and leaving the screening program) or closed cohorts (as in a trial situation or within a specific cohort invited to a screening program)[57]. Studies have also differed in relation to whether aggregate or individual level measures of screening exposure and breast cancer outcome were used. Both these factors can impact on the precision and/or reliability of estimates of over-diagnosis.

In relation to observational studies, the cohort approach is preferable since it allows follow-up of all women throughout and after the screening period, so that lead time effects can be fully compensated[89]. It also allows for individual measures of exposure to mammography screening and outcome to be incorporated into the analyses, which should provide for more precise estimates of over-diagnosis. However as noted previously, results from observational studies that have used individual measures of screening may be biased due to self-selection into screening programs. Therefore adjustment for risk factor differences between participants and non-participants is required to ensure unbiased estimates[57].

The cohort approach can only be applied to those groups with sufficient post-screening follow-up to allow for lead time effects to dissipate, which in turn is dependent on the timing of the study in relation to when screening was first implemented[57]. Thus far, analysis is not possible for any cohort which has completed the recommended screening protocol (e.g. screening at regular intervals from age 50-69 years) since programs have not been in place for a long enough period. Ideally, this would be 30 years or more since screening was first established, to allow for 10 years follow-up after the completion of screening. To date, estimates from observational studies using a cohort approach have been restricted to women who started screening from age 56 years or older.

Most studies prior to 2011 were undertaken using aggregate data for dynamic populations[57]. In many of these studies exposure was based solely on the woman's age (i.e. age-related eligibility for screening), while outcome measures were based on age-specific incidence rates, resulting in the less accurate measures of over-diagnosis. Duffy et al. 2014 have recently shown that reliance on aggregate data to estimate background incidence in the absence of screening based on projections of trends during the pre-screening era can potentially produce erroneous estimates[208]. They found that in some of the screening age groups the excess in observed incidence compared with expected incidence was greater than the total number of screen-detected cancers. They also found deficits in incidence in some of the older age groups which were not eligible to attend screening when it first began. These discrepancies indicate the influence of factors other than screening that could not be predicted from

projection of regression models using aggregate level data (despite a good fit with data in the pre-screening era). They conclude that many of the estimates of over-diagnosis derived from aggregate level data may be incorrect, and advocated for the use of individual level data in future studies aimed at measuring over-diagnosis.

**Timing of the study**

De Gelder et al[197] contend that it is only possible to derive unbiased estimates after a screening program has reached its steady state. Using a modelling approach, they demonstrated that estimates of over-diagnosis are strongly dependent on the timing of the study relative to the start of screening. In their model, estimates of over-diagnosis were highest during the implementation period and lowest 17 years out from the program start up.

Incidence rates increase quite dramatically during the implementation phase, due to the large numbers of women participating in screening for the first time (e.g. 100% in the first year). First time screening yields the largest number of cancers, since there is a large pool of cancers in the preclinical phase in a population that has not been previously screened. Detection rates at the first screen are generally equivalent to three times the expected annual incidence without screening[59]. Also the age range for first-time screening would be much broader during the implementation phase and would include a large number of women at the upper limit of the targeted age range. Hence the yield of future cancers would be further elevated due to the advancement of cancers in older women who have higher incidence. The higher incidence in the early phases of screening implementation therefore constitutes a somewhat 'artificial' situation which does not reflect the screening protocols as they are intended. De Gelder et al. suggest that the implementation phase should be considered to extend to at least 8 years after a program is first introduced[197]. This of course can vary depending on how screening programs were actually implemented in various jurisdictions.

The timing of studies in relation to program implementation also impacts on the length of follow-up time available when applying either the compensatory drop method or a cumulative incidence approach to adjust for lead time[117]. In a dynamic population it is important to consider the proportion of women in the older age group who had the opportunity to participate in screening. Estimates of over-diagnosis will decrease as

the number of women contributing to the deficit increases and are only accurate when all women in the age group above the screening limit have had the opportunity to be screened[57]. In closed populations, timing of the study will determine which cohorts can be considered in the analyses. Currently this would not include any cohorts which have received the full complement of screening according to conventional protocols, as discussed above.

**How measures of over-diagnosis are calculated and reported**

Over-diagnosis has been calculated and reported differently across studies[89, 197]. This is one of the major reasons why estimates have varied so widely[197]. Most studies have defined the screening population as those who have had the opportunity to participate based on the eligible screening age range[89]. A smaller number of studies have refined the screening population to those who were invited to participate in screening (e.g. women in the screening arm of a trial), and a smaller number still have restricted this group to women who actually participated in screening. Differences in estimates can therefore arise because not all eligible or invited women will participate in screening.

Adding to the basic issue of whether estimates apply to screened, eligible or invited women, is the wide variation in denominators used in the calculation and reporting of over-diagnosis measures[39, 197]. Differences exist in relation to what is actually being referred to (e.g. cancers diagnosed, screening episodes, screen-detected cancers) and the age range included (e.g. the target screening age group [50-69 years], the screening age group with 10 years follow-up [50-79 years], screening age with extended follow-up [>50 years], the whole life span [0-100 years]). Denominators relating to the number of cancers can be further subdivided in to whether they refer to the expected cases in the absence of screening or the observed number of cases in the presence of screening[39]. Over-diagnosis has been most frequently expressed as the excess number of breast cancers in the screened population as a percentage of the expected number of cancers in the absence of screening[197]. However a number of researchers have also reported over-diagnosis in relation to screening-aged women only (often expressed as a relative risk) while others[1, 210] have used only screen-detected cases as the referent population, which tends to inflate the excess.

De Gelder et al[197] have examined the extent to which the methods of calculation affects the estimate of over-diagnosis by comparing seven different methods of calculation identified across published literature. Each was applied to the same modelled dataset. They found that expressing over-diagnosis as the excess proportion among screen-detected cancers resulted in the highest estimates. As a proportion among expected cancers in the screening target age range, estimates were also relatively high in comparison to expressing the excess as the proportion of cancers with 10 year follow-up after screening or over the lifetime. As the length of follow-up after screening increases, estimates tend to become much smaller due to the larger denominator relative to the excess. They found that over-diagnosis estimates can vary up to 4 fold based on the choice of denominator.

The UK independent review panel argued that expressing excess cancers as the percent of observed cancers in the screened population is preferable as a measure of the extent of over-diagnosis from a population perspective, as it is in line with common notions of risk[39]. This would result in lower estimates of over-diagnosis than are generally reported when the expected number in the absence of screening is used as the denominator. [This can be illustrated with a simple mathematical example. If the observed incidence with screening was 125/100000 and the expected incidence without screening was 100/100000, over-diagnosis would be 25% if expressed as a percentage of the expected incidence (25/100). However, if expressed as the percentage of the observed incidence it would be reported as 20% (25/125)].

With respect to the age range that should be included, de Gelder et al.[197] also argued that denominators should include the entire life span since the effects of screening continue beyond the upper screening age limit (i.e. incidence drops). This approach is consistent with the commonly argued definition of over-diagnosis. Limiting the denominator to the screening age group, as reported in some studies, does not measure differences in the remaining lifetime risk of breast cancer.

Estimates may also differ in relation to whether DCIS is included in addition to invasive disease. Not all previous studies have included pre-invasive breast cancer in the estimation of over-diagnosis. Mammography screening has been directly linked to a marked increase in the detection of DCIS[22]. While there is considerable uncertainty

about the extent and duration of transition to invasive disease, studies of women who declined treatment suggest that not all DCIS will progress to invasive disease[197]. Estimates of progression of untreated DCIS range from 11-60% in 10-20 years[211]. If DCIS is a precursor to invasive disease, as is often contended, screening may actually lead to a deficit of invasive cases over the long term, which has been observed in some of the screening trials[59]. While the relative increase is large, the number of in-situ cancers detected at screening is still relatively small (15-20% in SA[51]) as a percentage of all breast cancers diagnosed. If a substantial proportion of in-situ cancers do not progress, the increased detection of DCIS may contribute quite substantially to over-diagnosis when considering all breast cancers. It is important therefore that DCIS be included in the estimation of over-diagnosis.

What constitutes the "right" measure of over-diagnosis is unclear and consensus is unlikely due to the polarity in opinions. What is clear is that the way in which over-diagnosis is measured and reported can generate very different impressions of the extent of the problem[39]. Estimates based on the screening age group only do not convey the level of risk of over-diagnosis over the remaining life time, which is likely to be in excess of 10 years after screening cessation for the majority of women in most populations studied. They therefore tend to inflate the perceived risk of over-diagnosis due to the denominator being restricted. Careful attention to what measures are being used is required when interpreting published findings. This is not always easy given that definitions are not always provided and measures are not always transparent.

## FINDINGS FROM PREVIOUS STUDIES

The following section of this chapter will focus on critically reviewing evidence relating to the extent of over-diagnosis due to mammography screening, with emphasis on original studies and systematic reviews which explicitly aimed to measure over-diagnosis.

### Follow-up of mammography screening trials

Several published studies and reviews have presented estimates of over-diagnosis using data from screening trials: Moss 2005[70], Zackrisson et al. 2006[71], Miller et al. 2014[210], The UK Independent Panel Review 2012[39] and Cochrane Reviews (2006, 2009,

2011, 2013)[199, 207, 212, 213]. Findings from these studies, for the separate trials, are shown in Table 5.1.

Moss[70] reviewed all 8 major screening trials using incidence data extracted from published studies but acknowledged that valid estimates could only be obtained from trials that did not offer screening to the control arm at the completion of the trial. These included the Health Insurance Plan (HIP) study (initiated 1963), Malmo (1976) excluding the youngest cohort who was offered screening, the Edinburgh trial (1978) and the two Canadian trials, National Breast Screening Study (NBSS) I and II (1981). The estimate from the HIP trial (3.6% excess) was not considered generalisable to current screening programs, since screening sensitivity has improved dramatically since the 1960's when this trial was undertaken. Also, data were only available for a very short follow-up period after the trial was completed (1.5 years) so would not have provided valid estimates. Likewise, the 31% excess incidence among those invited to screen in the Malmo trial (Sweden) was not considered to provide a reliable estimate of over-diagnosis since it was based on incidence data from before the trial had been completed which does not allow for any lead time adjustment (catch up in the control arm). Estimates derived from the Canadian trials, which involved 4 rounds of annual screening with physical examination for 40-49 year olds (NBSS I) and 50-59 year olds (NBSS II), were 11% and 14% respectively (for IBC and DCIS) after an average of 13 years follow-up from the start of screening. Moss considered these to be the most valid estimates of over-diagnosis[70]. Interestingly, comparison of cumulative incidence in the trials where screening was offered to controls (The Swedish Two County, 1977; Stockholm, 1981; and Goteborg trials, 1982) indicate lower incidence of invasive cancers and slightly higher incidence of in-situ cancers in the screened groups after 5-11 years follow-up, suggesting screening may have resulted in a shift from invasive to in-situ disease. Similar overall incidence suggests no over-diagnosis in relation to "incident" screening rounds.

Subsequent to this review, Zackrisson et al.[71] published estimates of over-diagnosis based on 15 years of follow-up of the Malmo screening trial. They reported that incidence was 32% higher during the screening period, and 8% lower than controls after screening ceased. The excess cumulative incidence for the whole period, while

not statistically significant, was 10% for those invited to screen compared with controls who had not receive screening at the end of the trial (i.e. 55-69 year olds). They concluded that the high excess incidence during screening is explained mainly by lead time of cases detected at screening.

Miller et al[210] have also published updated incidence data (for IBC only) from the Canadian screening trials with 25 years of follow-up from the start of these trials. Results of the two trials were combined in this publication and apply to women aged 40-59 years at recruitment. They report a stable residual excess among the screened group after 15 years of follow-up which they expressed as the proportion of screen-detected cancers (106/424 in total) to give a measure of over-diagnosis of 22%. If calculated as a percentage of cancers diagnosed over the entire follow-up period the estimate would be much lower (~ 4% =117 excess/3133 total cases among controls).

Measures of over-diagnosis from mammography screening trials have also been reported in the series of Cochrane review updates dating from 2006-2013[199, 207, 212, 213]. A summary estimate of 25% excess breast cancer incidence due to screening was provided for trials classified as having adequate randomisation, which included those where screening was not offered to controls at the conclusion of the trial. (Separate estimates were 32% for Malmo, 30% for NBSS I, 26% for NBSS II and 19% for the UK age trial). Estimates from the Malmo trial were based on the same data as used by Moss[70] which included little or no follow-up time once women stopped screening. However, their assessment of over-diagnosis in the Canadian trials was also based on a relatively short follow-up period after screening, although data for longer follow-up periods were published in 2000[214] and 2002[215]. They also included an estimate (19%) based on data from the UK age trial[216], which was designed specifically to examine the effectiveness of annual screening from age 40, and commenced in 1991. The data they present corresponds to less than 6 years of follow-up on average since recruitment into the trial, with most women not having completed the scheduled 8 rounds of screening. This is insufficient to allow for compensatory catch up among controls after screening ceased, so excess incidence during this period is not a valid measure of over-diagnosis. They also provided a summary estimate (33% excess) for three other trials which offered screening to controls based on incidence data for the period before the

trial concluded without long term follow-up. This cannot be considered to indicate the extent of over-diagnosis since there is no allowance for lead time effects. Also because of the short follow-up periods, essentially their denominators in all estimates refer to cancers occurring among women in the screening age range, which inflates the level of over-diagnosis.

The UK independent review panel on breast cancer screening also reviewed available data from screening trails[39], but only considered trials that did not screen the control group with follow-up extending at least 5 years after screening stopped. They excluded the HIP trial because of inconsistent data and the Edinburgh trial due to inadequate randomisation. Their results were very similar to those reported by Moss[70] for the Canadian trial and Zackrisson[71] for the Malmo trial, when comparable denominators were used in the calculation (i.e. ~11% over-diagnosis for all BC including DCIS). They have also presented their summary measure as a proportion of cancers diagnosed women of screening age only (which resulted in an estimate of 19% over-diagnosis).

While follow-up studies of the Canadian screening trials (which commenced in 1980) are held up as being the best available evidence on over-diagnosis, it is possible that results may be inaccurate due to the continuation or uptake of screening by trial participants after the end of the trial period through public screening programs. This is not an unrealistic scenario given that a large proportion of women were still within the target screening age when these trials were completed and that population-based breast screening programs were established only a few years later. Continued screening among those in the screening arms would lead to inflated estimates of over-diagnosis, whereas uptake by the control groups would tend to understate the true extent of over-diagnosis. Hence there is still a considerable degree of uncertainty about the accuracy of estimates based on evidence from randomised controlled trials.

**Table 5.1. Estimates of over-diagnosis based on follow-up of mammography screening trials (IBC+DCIS)**

| Author, date (Trial) | Age at recruitment | Follow-up from start of trial (yrs) | Controls screened | Follow-up >5yrs after trial ended | RR screened / unscreened | % OD | comment |
|---|---|---|---|---|---|---|---|
| **Moss 2005[70]** | | | | | | | |
| HIP | 40-64 | 5 | no | no | 1.04 | | Not generalisable |
| Malmo | 45-70 | 8 | no[a] | no | 1.31 | | Estimated prior to end of trial |
| Canada I | 40-49 | 13 | no | yes | 1.14 | | Follow-up ~6yrs post-screen |
| Canada II | 50-59 | 13 | no | yes | 1.11 | | Follow-up ~6yrs post-screen |
| Edinburgh | 45-64 | 10 | no | no | 1.13 | | Short post-screen follow-up period |
| Two County | 40-74 | 11 | yes | no | 0.99 | | Follow-up after controls screened |
| Stockholm | 40-64 | 5 | yes | no | 0.97 | | Follow-up after controls screened |
| Goteborg | 39-59 | 8 | yes | no | 0.94 | | Follow-up after controls screened |
| **Zackrisson et al. 2006[71]** | | | | | | | |
| Malmo | 55-69[b] | 25 | no | yes | 1.10 | | Follow-up ~15yrs post-screen |
| **Cochrane reviews 2006-2013[199, 207, 212, 213]** | | | | | | | |
| Malmo | 45-70 | 7-9 | no | no | 1.32 | | Estimated prior to end of trial |
| Canada I | 40-49 | 7-9 | no | no | 1.30 | | Short post-screen follow-up period |
| Canada II | 50-59 | 7-9 | no | no | 1.26 | | Short post-screen follow-up period |
| UK age trial | 40-49 | 7-9 | no | no | 1.17 | | Estimated prior to end of trial |
| Weighted Summary | | | | | 1.25 | | |
| Two County | 40-74 | Trial period | yes | no | 1.33 | | Estimated prior to screening of controls |
| Stockholm | 40-64 | Trial period | yes | no | 1.49 | | Estimated prior to screening of controls |
| Goteborg | 39-59 | Trial period | yes | no | 1.13 | | Estimated prior to screening of controls |
| Weighted Summary | | | | | 1.33 | | Estimated prior to screening of controls |
| Total | | | | | 1.29 | | |
| **UK independent panel 2012[39]** | | | | | | | |
| Malmo | 45-70 | 25 | no | yes | | 12% | Excess among screened for entire follow-up |
| Canada I | 40-49 | 13 | no | yes | | 14% | Excess among screened for entire follow-up |
| Canada II | 50-59 | 13 | no | yes | | 10% | Excess among screened for entire follow-up |
| Summary | | | | | | 11% | Excess among screened for entire follow-up |
| **Miller et al. 2014[210]** | | | | | | | |
| Canadian I & II | 40-59 | 25 | no | yes | | 22% | Excess % of screen-detected during trial |

[a] Screening was offered to controls in the youngest cohort only

[b] Youngest cohort excluded because controls received screening at end of trial

**Table 5.2. Micro-simulation, multistate and lead time modelling studies**

| Author / date | Population | Method | Parameter estimates | Measure | Denominator | % over-diagnosis |
|---|---|---|---|---|---|---|
| Yen et al. 2003[88] | Swedish Two-county Service screening UK USA, Australia & NL | Multistate Markov model first 2 screening rounds | Varies across datasets | Derived from model parameters for proportion of non-progressive DCIS | Incidence of DCIS at first and second screening rounds | 1st screen 37% DCIS<br>2nd screen 4% DCIS<br>(average) |
| Paci et al. 2004[79] | Florence service screening | Lead time adjustment assuming exponential distribution | MST=3.42yrs | Observed incidence adjusted for lead time/ expected incidence without screening | Cumulative incidence for age 50-84yrs<br>(IBC+DCIS) | 2% IBC<br>5% IBC+DCIS |
| de Koning et al. 2005[86] | Dutch service screening | Multistate micro-simulation model MISCAN | No details | Modelled excess incidence with screening/Expected incidence without screening | Cumulative incidence over a lifetime 0-100yrs (IBC+DCIS) | 3% |
| Duffy et al. 2005[69] | a. Two-county trial<br>b. Goteborg trial | Multistate Markov model first 3 screening rounds | a. MST=2.7yrs<br>b. MST=1.9yrs<br>Sensitivity=99.9% | Derived from model parameters for proportion of non-progressive BC | Incidence during screening round 1 (and rounds 2&3) including interval cancers<br>(IBC +DCIS) | a. 1st screen 3.1%<br>2nd/3rd screen 0.3%<br>b. 1st screen 4.2%<br>2nd/3rd screen 0.3% |
| Paci et al. 2006[78] | Northern and Central Italy service screening | Lead time adjustment assuming exponential distribution | $MST_{50-59}$=3.7yrs<br>$MST_{60-69}$=4.2yrs | Observed incidence adjusted for lead time/ expected incidence without screening | Cumulative incidence for age 50-74yrs<br>(IBC+DCIS) | 3% IBC<br>5% IBC+DCIS |
| Olsen et al. 2006[205] | Copenhagen service screening | Multistate Markov model first 2 screening rounds | MST=2.7yrs<br>Sensitivity=100% | Modelled excess incidence/Observed incidence among screened | Incidence at first /second screen plus intervals over 4 years (IBC +DCIS) | 1st screen 7.8%<br>2nd screen 0.5%<br>Combined 4.8% |
| Martinez-Alonso et al. 2010[76] | Catalonia, Spain population data | CISNET model to adjust for population screening patterns | Weighted patterns of screening | Excess observed incidence/Expected after adjustment for birth cohort | Cumulative incidence from 0 to maximum age of cohort<br>(IBC only) | By birth cohort<br>1935-1950<br>0.4%, 23%, 31%, 47% |
| Seigneurin et al. 2011[204] | Isere, France population data | Bayesian stochastic (multistate) simulation model | MST=2-4 yrs<br>Sensitivity=90% | Derived from model parameters for proportion of non-progressive IBC and DCIS | Incidence during screening period 50-69yrs (IBC+DCIS) | 1.5% IBC<br>28% DCIS |
| de Gelder et al. 2012[197] | Dutch service screening | Multistate micro-simulation model MISCAN | $MST_{DCIS}$=2.6yrs<br>$MST_{IBC}$= 1.0-3.9yrs*<br>Sensitivity=47-95%*<br>*Depending on stage | Modelled excess incidence with screening /Expected incidence without screening<br>(steady state phase) | Cumulative incidence over a lifetime 0-100yrs (IBC+DCIS) | 2.8% |

% over-diagnosis is for all BC except where specified, MST= mean sojourn time (equates to average lead time), NL=Netherlands

**Micro-simulation, multistate and lead time modelling studies**

A number of modelling studies of varying complexity have been conducted to estimate over-diagnosis from breast screening. These are summarised in Table 5.2. While it is difficult to thoroughly critique many of these studies without detailed knowledge of the modelling mechanics or validity of assumptions on which they are based, a brief description and limited assessment of the approaches used is provided in this section.

Paci et al. have used a relatively simple yet transparent lead time modelling approach in their 2004[79] and 2006[78] publications relating to mammographic screening in Florence and Northern/Central Italy. In the 2004 study, the authors determined the probability that screen-detected cancers would have become clinically symptomatic each year after screening, based on an exponential distribution of lead time with a mean duration of 3.7 years as estimated in the Florence screening pilot study. They applied these probabilities to determine the additional number of cancers expected in each year during and beyond the study period in accordance with the number of screen-detected cancers recorded. Observed incidence was corrected for lead time by subtracting the total number of cancers that would have surfaced clinically after the end of the study period. Lead time adjusted incidence was then compared with the expected incidence without screening, to provide an estimate of over-diagnosis. In this study, age-standardised pre-screening incidence rates were used as the expected incidence without taking into account any increasing time trends. Hence the reported estimates of over-diagnosis may be biased upwards. Despite the lack of adjustment for increasing background incidence they found minimal over-diagnosis for IBC (~2%) and very modest levels when DCIS was included (~5%). In the 2006 study, period trends were taken into account in determining the expected incidence in the absence of screening. The authors also used age-specific lead time estimates from the Swedish Two County Trial (3.7 and 4.2 years for women aged 50-59 and 60-69 years respectively) when adjusting for lead time. Their final estimates were consistent with those reported in the first study (3% over-diagnosis for IBC and 5% for IBC and DCIS combined for women age 50-74 years).

Multistate models have been used to estimate over-diagnosis in studies by Olsen et al. 2006[205], Duffy et al. 2010[69] and Yen et al. 2004 (for DCIS only)[88] using data for the first

few screening rounds from the screening trials and/or from mammography screening programs in various countries. Each of these models included progressive and non-progressive preclinical states. Parameters for the rate of progressive and non-progressive cancers and mean length of duration in the screen-detectable phase for each state were derived from the models that gave the best fit with the observed data for screen-detected and interval cancer rates. While this leaves these studies open to the criticism that lead time lengths may be inflated because over-diagnosed cases are included in the measure of lead time, their models explicitly allow for a proportion of screen-detected cancers to be non-progressive. Each of these studies indicated a higher proportion of non-progressive (over-diagnosed) cases in the first round of screening and very low or negligible proportions in subsequent rounds. Over 2-3 screening rounds the level of over-diagnosis was found to be relatively low (<5% for IBC and DCIS combined). While the proportion of non-progressive DCIS was found to be high in absolute terms, particularly at the first screening round (37%), in-situ cancers only makes a small contribution to the overall level of over-diagnosis because they constitute a relatively small proportion of the cancers detected[88].

De Koning et al. (2006)[86] have developed a more complex micro-simulation model, referred to as MISCAN, which was used to estimate over-diagnosis in the national screening program in the Netherlands. They first modelled the "natural history" of breast cancer in a synthetic cohort of women based on clinical data (incidence by age and stage) before the implementation of screening, then adjusted for screening effects using estimates for various screening parameters (e.g. screening sensitivity and duration in various preclinical stages) to assess which measures give the best fit with observed data. In their 2006 publication the authors only describe their model in very general terms and do not provide details about parameters used in or estimated from the model. They reported estimates of 3% over-diagnosis (which includes in-situ cases) for total life time incidence (ages 0 to 100 years) and 8% for screen-detected cancers. The MISCAN model is described in more detail in a publication by the same research team in 2011[197], which reported 2.8% over-diagnosis over the entire lifetime for the program in its steady state.

Recently Seigneurin et al.[204] estimated over-diagnosis for IBC and DCIS separately, using a relatively complex Bayesian modelling approach which allowed for both progressive and non-progressive cases of DCIS and IBC. In this approach, a large number of datasets were simulated using different values for various unknown parameters (constrained by general knowledge and or other published data). Estimated values for these parameters were derived from the model that produced the best fit with observed data (i.e. incidence data for the provincial region of Isere in France, from 1991-2006). They estimated 3.3% of IBC and 31.9% of DCIS cases detected by screening were over-diagnosed. This corresponds to 1.5% of all IBC and 28% of all DCIS diagnosed in women aged 50-69 years during the study period. An estimate for IBC and DCIS combined was not provided. The authors concluded that while the proportion of over-diagnosed DCIS was high, the contribution to the overall level of over-diagnosis was relatively small given that only 15% of breast cancer cases are in-situ.

Martinez-Alonso et al[76] estimated over-diagnosis in successive cohorts of women eligible for mammography screening in the Catalonia region of Spain using one of the CISNET models developed by Lee and Zelen[217]. Their approach was to model the observed breast cancer incidence for the region using age-period cohort regression with additional variables to capture complete fertility and the proportion of women screening at age 50. The background incidence without screening was then determined by setting the proportion screening at age 50 to zero. The CISNET model was subsequently applied to the background incidence to predict the incidence for a series of screening scenarios (e.g. annual or biennial screening commencing at age x continuing to age y, etc.) with each scenario weighted to reflect population screening patterns determined from survey data. Estimates of over-diagnosis were derived by comparing the predicted cumulative incidence with the observed incidence for separate birth cohorts who would have been eligible for mammography screening. Results indicate a dramatic increase in over-diagnosis by birth cohort, with older cohorts (those born around 1935) having a negligible level of over-diagnosis (0.4%), and younger cohorts (those born around 1950) having extremely high levels (46.6%). The authors offered little explanation for the extreme range of estimates but imply that they are related to increasing screening intensity in more recent times. Their

method for deriving background incidence in the absence of screening is questionable since it involved removing screening effects from observed data, where the screening indicator variable in their model was simply the proportion of women screening at age 50. Also, screening patterns were inferred from survey data and may not reflect the true pattern of screening in the region over the course of the study. There is also some uncertainty about the validity of parameters in the CISNET model which may have been derived from other populations. Estimates for each of the cohorts are also likely to differ because the denominators used would have applied to different age ranges (i.e. younger cohorts would have fewer cancer cases in the denominator leading to higher measures of over-diagnosis).

**Observational studies using aggregate data**

***Comparisons based on temporal trends***

Several studies have attempted to measure over-diagnosis by comparing observed incidence in populations with a mammography screening program with the expected incidence had screening not been implemented (see Table 5.3) by projecting of incidence trends from the period before population-based screening was introduced. This approach was used in studies by Jonsson et al. 2005[218], Jorgensen and Gotzche 2009[82], Morrell et al. 2010[77], and Duffy et al. 2010[69]. In addition to temporal trends, Jonsson et al. also took regional variation in incidence rates into account in predicting background incidence, while Morrell also adjusted for estimated changes in the prevalence in obesity, nulliparity and HRT use before and after implementation of population-based screening in addition to temporal trends. Morrell et al. also used an alternative method to predict background incidence rates which involved interpolation across younger and older age groups. Linear rather than Poisson regression was used in the studies by Morrell et al.[77] and Jorgensen and Gotzsche[82] which may have led to slightly lower estimates of background incidence rates, and hence, inflated estimates of over-diagnosis. In both of these studies the pre-screening period was restricted to avoid any opportunistic screening that may have occurred prior to the introduction of population-based programs. This may not be appropriate as incidence began to increase during this time for reasons other than increases in opportunistic screening, resulting in a larger observed excess compared with expected rates.

By contrast, Zahl et al. (2004 & 2012)[74, 219] compared incidence rates after screening had commenced with rates in the pre-screening era without any adjustment for increases over time or for other risk factor changes. Their justification was that the increases they observed (1% per annum during the 4 years prior to screening implementation in Norway and 0.8% per year over a 15 year period prior to screening implementation in Sweden) were not statistically significant. Furthermore, data presented in the 2004 paper showed increases in age-specific incidence in the Norwegian regions without population-based screening programs, which were not taken into account. Over-diagnosis is therefore likely to be overestimated.

In another paper by Zahl et al.[54], published in 2008, incidence rate comparisons were made for overlapping time periods within the same population (Norwegian counties where screening was first introduced). The 'control' period, 1992-1997, included a 4 year period before screening commenced as well as the first two years of screening (i.e. the first screening round). The 'screening' period, 1996-2001, corresponds to the first three screening rounds offered in these regions and is therefore likely to include a high proportion of prevalence screens in each of the rounds, which will inflate estimates. No adjustment for risk differences were made, since the risk profiles during the two periods were considered to be comparable. Nor was any adjustment made for lead time effects. Over-diagnosis was determined by comparing cumulative incidence rates in women aged 50-64 years over the two time periods. The authors argue that, since the control period included the first round of screening, the cumulative incidence across the two periods should be the same after the 6 year follow-up period. Kalager and Bretthauer have pointed out that two thirds of the women considered in each comparison group are actually the same women[220]. The only groups who would have contributed to differences during the follow-up period were women aged 60-64 years in the control group and 50-54 years in the screened group. This age differential was not taken into account in assessing the level of over-diagnosis. On the basis of their (erroneous) finding of 22% higher cumulative incidence in the screening period, the authors have argued that some screen-detectable breast cancers must spontaneously regress. This study has since been cited as 'evidence' for spontaneous regression of breast cancer[200, 206].

The study by Waller et al.[68] used quite a different approach to measure over-diagnosis in the UK. Rather than predicting expected incidence by extrapolation of pre-screening incidence, they used age-period-cohort regression to model incidence over a 30 year period including the years before and after screening. Their model included screening variables to indicate the expected proportion of women attending screening (for first and subsequent rounds) and who had completed screening by calendar year and year of age. Background incidence trends were predicted by setting these screening variables to zero. Their regression model also incorporated prevalence of HRT use to adjust for the effects of changing patterns of use. Cumulative incidence rates were calculated by summing the age-specific incidence rates based on models with and without screening (by setting screening variables to zero). Results were presented as life-time risk of breast cancer with and without the effect of a population-based mammography screening (8.6% in the presence of screening compared with 7.8% in the absence of screening which is equivalent to 10% excess risk). This is one of the few studies to have attempted to account for changing prevalence of HRT use and therefore may have resulted in more valid estimates than other studies using population level data.

Not all the studies listed in Table 5.3 have adequately adjusted for lead time effects. While studies by Jorgensen and Gotzsche[82], Zahl et al. 2004[74] and Zahl and Maehlen 2012[219] and Duffy et al. 2010[69] report using the compensatory drop method, only the study by Duffy et al.[69] is likely to have adequately adjusted for lead time effects. In their 2004 study, Zahl et al.[74] did not adjust for the decline in incidence among women aged 70-74 years because the risk ratios (0.89 for Norway and 0.98 for Sweden) were not statistically significant. They also ignored the statistically significant decline among women aged 75-79 years (RR=0.88) for Sweden. Furthermore, insufficient time had elapsed after implementation of screening in Norway for the anticipated deficit to be apparent in the post-screening age group. The compensatory drop in older women was also ignored in the 2012 study by Zahl and Maehlen[219], pertaining to data from Norway. Estimates of over-diagnosis in these two papers are therefore likely to be considerably inflated since they are essentially the incidence rate ratios for the screening versus pre-screening period with no adjustment for changes in incidence over time or for lead time effects. The compensatory drop method was also used by

Jorgensen and Gotzche[82], who examined over-diagnosis in 5 different countries using population incidence data extracted from secondary sources. Rate ratios for observed versus expected incidence in the older age groups, who had completed screening, showed relative declines in Manitoba (Canada), Sweden and Norway but not in the UK or in NSW (Australia). Adjustment was made by subtracting the absolute number of cancers corresponding to the drop (where it applied) from the excess number of cancers occurring among those in the targeted screening age group, based on risk ratios in each of these respective age groups in the last year of observation. The summary estimate across all 5 countries indicated 52% over-diagnosis, which includes DCIS as well as IBC. The range for individual countries was 44% to 57%. Except for Manitoba, rather than being founded on actual data for all breast cancer cases, measures were based the assumption that in-situ cancers contributed an additional 10% to incidence rates. This study has been criticised by Puliti et al.[89] for comparing the incidence in the latest year of screening (i.e. when incidence was at its highest). Comparisons which were based on modelled incidence rather than observed incidence during the post screening period, do not capture the temporal nature of the excess and deficit due to screening, and hence do not sufficiently account for lead time effects. In stark contrast to Jorgensen and Gotzsche's estimate of 57% over-diagnosis for UK, Duffy et al. (2010)[69] found 4% over-diagnosis (IBC only) for the same population over a similar period. While the general approach used by both research teams was similar, Duffy et al. had direct access to cancer registry and denominator population data, while Jorgensen and Gotzsche used broad age-specific rates derived from published reports. Furthermore, Duffy et al. used Poisson regression to predict expected incidence, did not restrict the pre-screening period and compared expected incidence with observed incidence for the period 1989-2003. The contrasting results from these two studies serve to highlight the potential impact of specific details of study design on final estimates of over-diagnosis.

Two studies, that of Jonsson et al.[218] and Morrell et al.[77], used the alternative "rate shift" method to adjust for lead time effects. The rationale for this approach is that once a screening program reaches its 'stable' phase, the incidence among women attending screening will also stabilise at a level equivalent the expected incidence for women after the average lead time duration had passed (i.e. the incidence of women

2-5 years older depending on assumption about lead time length). As outlined previously this method however may not adequately adjust for lead time. As a consequence estimates arrived at in the papers by Jonsson et al.[218] and Morrell et al.[77] are likely to overstate the extent of over-diagnosis.

With the exception of the paper by Waller et al.[68], all studies presented in this section have reported the excess (over-diagnosed cases) as the percentage of cancers occurring during the screening age range (generally 50-69 years). Collectively these estimates greatly inflated the measure of over-diagnosis because they have restricted age range to which they refer. By contrast, most studies described in other sections have considered screening and post-screening age groups in the denominator when expressing the level of over diagnosis.

***Geographical comparisons***

Three studies have attempted to estimate over-diagnosis by comparing incidence in regions where mammographic screening programs has been implemented with regions where public programs were not available. This study design is only possible in countries where there was a staggered implementation of screening or pilot programs well before a national screening program was introduced (e.g. Denmark, Sweden and the Netherlands). The fourth study included in the section, by Kaleger et al.[221], also used data from regions with and without screening but in a different way (discussed below). Study design features and results are shown in Table 5.4.

Peeters et al. 1989[72] compared incidence over a 12 year period (1975-1986) in Nijmegen, in the Netherlands, which had a screening program, with incidence in a neighbouring city where there was no program. No adjustment was made for differences in background risk in the two regions, although comparability was tested and deemed acceptable (i.e. non-significant RR of 0.96). Summary rate ratios, adjusted for birth cohort, were calculated for different periods after the introduction of screening. A rate ratio of 1.10 was found for the whole 12 year follow-up period, indicating over-diagnosis accounted for 11% of all breast cancers diagnosed in Nijmegen among women over the age of 35. While incidence was considerably elevated during the initial period spanning the first two rounds of screening (RR=1.30), rate ratios for the two subsequent periods showed very little elevation (1.03 and 1.01

respectively). Puliti et al.[89] considered that this study did not adequately adjust for lead time. This is likely to be the case in relation to estimates from the early period. However, estimates for the later periods do incorporate some lead time adjustment via a compensatory drop in the oldest cohort with longer term follow-up. Findings suggest very little over-diagnosis when lead time is adequately accounted for.

More recently, Jorgensen et al.[75] published a study which compared breast cancer incidence regions with screening (Copenhagen from 1991 and Funen from 1995) and with the rest of Denmark which did not have screening during the study period. They used Poisson regression to estimate incidence rate ratios between regions for women in the screening target age (50-69 years) and women in the post-screening age range (70-79 years) for the whole study period (1998-2003). Their model included adjustment for geographical differences during the pre-screening period (1971-1990) as well as differences in the age distributions, but Puliti et al.[89] judged this risk adjustment to be inadequate because differences in the average incidence in screening and non-screening regions during the pre-screening era were not necessarily applicable to later periods. The compensatory drop method was used to adjust for lead time effects, with the excess and deficit number of cancers in the screening regions determined from the IRR in the screening (50-69 years) and post screening age groups (70-79 years) respectively. The 33% over-diagnosis reported refers to the excess number of cancers due to screening after lead time adjustment as a percentage of expected number occurring during the screening target age. This is equivalent to 18% over-diagnosis if considered as the proportion of cancers in women age 50 to 79 years who were eligible to be screened.

Geographical comparisons were also used by Hellquist et al.[222] in a Swedish study examining the extent of over-diagnosis occurring in women aged 40-49 years. Over-diagnosis was determined by calculating the incidence rate ratio for regions where women in their forties were invited to participate in screening compared with regions where they were not invited. Rate ratios were adjusted for differences in incidence between these regions prior to the introduction of screening in younger age groups. They also applied an incidence rate adjustment to account for lead time effects in age and period trends, based on an average 'population' lead time of 1.2 years (which is

equivalent to the rate shift method). The first round of screening was excluded. Rate ratios were 1.01 for all BC and 0.95 for IBC alone, suggesting no over-diagnosis in this age group (with a possible deficit of invasive cases) in subsequent screening rounds. Whether risk adjustment based on incidence rates prior to the introduction of screening for women 40-49 years in selected areas adequately accounted for risk differences between regions (particularly in relation to risk among those who took up the invitation to screen) is unclear. Uncertainties in relation to the validity of the rate shift method also apply to this study.

A fourth study by Kalager et al.[221], undertaken in Norway, also used data from different geographical regions with and without screening to estimate background trends. Temporal trends in incidence in the regions without screening programs were taken to indicate the changes in incidence due to factors other than screening. Their estimates were based primarily on temporal comparisons in the regions where screening had been implemented (incidence rate ratios for the screening period versus pre-screening period) with adjustment for changes in incidence rates in regions without screening. Final estimates were derived by dividing the IRR for current versus pre-screening incidence in screening regions by the IRR for current versus pre-screening incidence in non-screening regions. Two different approaches were used to adjust for lead time effects: the compensatory drop method, achieved by including a 10 year follow-up period after completion of screening in incidence rate comparisons (i.e. aged 50-79 years); and the rate shift method, whereby rates were shifted upwards (for both age and period) by the lead time period in the comparison population. For the rate shift method, cases detected in the first screening round were excluded. Due to uncertainty around lead time estimates for mammography screening, the authors chose to use two estimates of lead time length: 2 and 5 years. Their estimates of over-diagnosis using the compensatory drop method was 18-25% of expected cancers among women age 50-79 years. The lower estimate applied to the region with a 10 years follow-up period, whereas the upper estimate applied to all screening regions across Norway, with follow-up periods ranging from 2-10 years after the start of screening, which do not allow for lead time effects to be fully compensated. The rate shift method yielded estimates of 15% using a 5 year incidence shift and 25% using a 2 year incidence shift. Given the rate shift method is unlikely to adequately account for lead time effects and

that the follow-up period for screening programs across the whole of Norway is too short to apply the compensatory method, the only estimate where lead time adjustment may be adequate is the compensatory drop method applying to the region with 10 years follow-up (which they report to be 18% for IBC). This estimate is higher than most others studies that which have used similar denominator population.

**Table 5.3 Observational studies using aggregate data in dynamic populations to measure over-diagnosis (OD)**

| Author/date | Country Region | General approach | Risk factor adjustment | Lead time adjustment | Follow-up from start of screening | OD measure | Denominator | OD estimate IBC [All BC] |
|---|---|---|---|---|---|---|---|---|
| Zahl et al. 2004[74] | Norway Sweden | Age-period cohort model with screening term , with "prevalence" and "incidence" screening period in each country | None | Compensatory drop (dismissed as non-significant) | Norway 5yrs Sweden 15yrs | Rate Ratio: observed/expected | Expected rate for 50-69yrs without screening | 1.56 Norway 1.45 Sweden |
| Jonsson et al. 2005[218] | Sweden | Observed vs expected incidence without screening | Temporal Region | Rate shift (1.6 -3yrs) | 11-15yrs | Rate Ratio: observed/expected IBC only | Expected rate within age group without screening | 1.54 50-59yrs 1.21 60-69yrs |
| Waller et al. 2007[68] | UK | Observed vs expected incidence without screening | Temporal Cohort HRT use | *Cumulative incidence* | 14yrs | Lifetime risk with screening vs without screening | Expected CI without screening to age 75yrs | Lifetime risk 8.6 vs 7.8% |
| Zahl et al. 2008[54] | Norway | Temporal comparison 2 overlapping time periods: 6yrs pre-screen to round 1 6yrs screening rounds 1 to 3 | None | No adjustment (Self-referent population) | 6yrs | Cum Inc. Ratio: Screen rounds 1-3/ Pre-screen period to screen round 1 | Cum Incidence over 6yrs from pre-screen to 1st screen 50-64yrs | 1.22 at 6yrs 1.20 at 8yrs |
| Jorgensen & Gotzsche 2009[82] | UK, Sweden, Norway, NSW, Manitoba | Observed vs expected incidence without screening | Temporal | Compensatory drop | Various | Rate Ratio: observed/expected | Expected rate for 50-69yrs | 1.52 all BC (range 44-57%) |
| Morrell et al. 2010[77] | NSW, Australia | Observed vs expected incidence without screening | Temporal BMI, HRT, Nulliparity | Rate shift (2.5 & 5yrs) | 10-12yrs | Rate Ratio :observed/expected | Expected rate in 50-69yrs | Interpolation: 1.42-1.51 Extrapolation: 1.30-1.36 |
| Duffy et al. 2010[69] | England | Observed vs expected incidence without screening | Temporal | Compensatory drop | 15yrs | Cases OD per 1000 women screening for 20 years | Observed cancers in 50-64yr olds screened for 20yrs | 2.3/1000 women/20yrs (4% of IBC 50-64) |
| Zahl & Maehlen 2012[219] | Norway | Observed vs expected incidence without screening | None | Compensatory drop (dismissed as non-significant) | 14yrs | % excess RR screening /pre-screening period among 50-69yrs | Expected rate for 50-69yrs before screening started | 50% increase in incidence rate not explained by lead time or HRT |

**Table 5.4 Observational studies using geographical comparison methods to measure over-diagnosis (OD)**

| Author/date | Country Region | General approach | Risk factor adjustment | Lead time adjustment | Follow-up from start of screening | OD measure | Denominator | OD estimate |
|---|---|---|---|---|---|---|---|---|
| Peeters et al. 1989[72] | Nijmegen, Netherlands | Geographical comparison with and without screening | None | Compensatory drop *(not stated as such)* | 10yrs | Rate ratio: Screening/non-screening region | Rate (cancers/Py) in region without screening 35yrs+ | 1.01 at 10yrs all BC |
| Jorgensen et al. 2009[75] | Denmark | Geographical comparison with and without screening<br>IRR for population screening by Poisson regression | Temporal trends<br>Regional differences in pre-screening period | Compensatory drop | 13yrs | % Excess cancers / all cancers in age group | No. cancers diagnosed ages 50-69, 50-79yrs | 33% 50-69yrs all BC<br>18% 50-79yrs all BC |
| Hellquist et al. 2012[222] | Sweden | Geographical comparison with and without screening<br>Applies to 40-49yrs only<br>(excludes 1st screening round) | Regional diff | Rate adjustment of ~1.2yrs by age and period | Follow-up to age 49yrs | Rate ratio: screening region (adj) / non-screening regions | CI rate for women 40-49yrs in regions with no screening | 0.95 IBC<br>1.01 all BC |
| Kalager et al. 2012[221] | Norway | Compared screening and pre-screening incidence adjusting for temporal trends in non-screening regions<br>(excludes 1st round screen for rate shift method) | Temporal trends in non-screening regions | a) Compensatory drop<br>b) Rate shift (2 & 5yrs) | 2-10yrs | a) % excess CI rate among 50-79yrs<br>b) % excess CI rate among 50-69yrs | Expected CI rate without screening within age group (IBC only) | a) 18% 10yrs f/u<br>25% 2-10yrs f/u<br>b) 15% 5yr shift<br>25% 2yr shift |

**Observational studies using cohort approaches**

For the most part, previous studies measuring over-diagnosis have tended to use aggregate level data relating to populations where screening programs have been implemented. Reliance on aggregated data and the 'dynamic' nature of populations being studied can lead to erroneous estimates of over-diagnosis, as has been described previously. More recently, several studies have been published which use a cohort approach and/or individual level data on both screening exposure and breast cancer outcomes to estimate over-diagnosis. These are summarised in Table 5.5.

Using data from the Florence screening program, Puliti et al. (2009)[81] estimated over-diagnosis by comparing the observed and expected cumulative incidence in a cohort of women who had been invited to participate in the screening when the program was first initiated. Their study only included women aged 60-69 years at the start of screening to allow for at least 5 years of post-screening follow-up after the age of 69 years. The expected incidence in the absence of screening was estimated via projection of pre-screening incidence trends using Poisson regression. Aggregate level data were used in this study. Lead time effects were accounted for by calculating cumulative incidence within cohorts that had at least 5 years follow-up after they reached the maximum screening age. Women younger than 60 years would not have had sufficient follow-up time, given the study was conducted 15 years after the screening program commenced, and hence they were not included in the estimation of over-diagnosis. The authors report a rate ratio for observed to expected cumulative incidence of 1.01 for all BC and 0.99 for IBC, indicating little or no over-diagnosis after 5-15 years of follow-up among women aged 60-69 years when first invited to screen. This longitudinal approach to calculating cumulative incidence offers more reliable estimates than the cross-sectional approaches used in other studies.

In a subsequent study conducted by the same authors[85], over-diagnosis was estimated using a similar study design but with individual level rather than aggregate level data. Again the study was restricted to the invited cohort aged 60-69 years at the commencement of screening 1990. Women were classified as attenders (attended one or more invitations to screen) or non-attenders (never attended screening). Breast cancer incidence over a 15-year follow-up period was determined via linkage with the

regional cancer registry and over-diagnosis estimated by comparing the cumulative incidence in the two groups. Adjustment was made for differences in socioeconomic deprivation and marital status using Poisson regression, which may have partially adjusted for differences in breast cancer risk between attenders and non-attenders. Their estimates of 10% over-diagnosis (RR=1.10) for all BC and 5% for IBC alone (RR=1.05) refer to the extent of over-diagnosis in those who actually participated in screening rather than in the population as a whole, so would be anticipated to be slightly higher than estimates applying to the eligible population, all of whom do not attend screening.

Njor et al.[84] conducted a study in Denmark which included the use of both temporal and geographic comparison cohorts, with individual level data on breast cancer incidence and deaths, where applicable. Three 'control' groups were established: an historical cohort from the regions with screening programs, an historical cohort in non-screening regions and a current cohort in these non-screening regions. The main comparison was between the contemporary cohort and historical cohort from the screening regions. Adjustment for background incidence trends unrelated to screening effects was based on difference in incidence rates in the historic and current cohorts in non-screening regions. Interaction terms were included where appropriate to accommodate differences in temporal trends across regions. The cumulative incidence method, with 8 years of follow-up, was used to incorporate lead time effects. The combined estimate of over-diagnosis across the regions with screening was 2.3% (RR=1.023) for women with at least 8 years post-screening follow-up. The authors also examined the potential for selection bias by comparing the cumulative incidence among never-screened women in regions where screening was available with the expected incidence had screening not been implemented. Adjusting for this difference led to marginally higher estimates of over-diagnosis among participants.

Using individual level data from the British Columbia (Canada) Breast Screening Program linked to cancer registry and death registration data, Coldman et al.[223] measured over-diagnosis by comparing incidence among active screening participants with non-active participants. Time at risk in the active screening group included a 5-year post-screening period (which may not have been long enough to fully account for

lead time). 'Non-active' time at risk was determined by subtracting the time at risk among the active group from the mid-year population estimates. The total cumulative incidence in each group was calculated by summing the 5 year age-specific incidence rates for the period 2005-2009. The screen-detection rate for the first screening round (at age 40 years) was calculated separately and added to the total for the active screening group. This method yielded over-diagnosis estimates of 5.4% for IBC and 17.3% for all BC. Estimates pertained to women aged 40 to 89 years who had participated in screening. No adjustment was made for risk differences between screening participants and non-participants therefore the estimates may be biased. The authors also present over-diagnosis estimates for the whole eligible screening population which would not be affected by self-selection bias. These estimates were based on comparisons between the observed cumulative incidence eligible and the expected cumulative incidence in the absence of screening derived from projection of pre-screening incidence rates, using the compensatory drop method to adjust for lead time effect. Estimates for the eligible population were lower than for screened women, as would be expected (i.e. -1.7% for IBC and 6.7% for all BC). This latter method is less robust due to the uncertainties inherent in predicting of background incidence rates based on projections.

Individual level data were also used in the study by Falks et al.[83], which included all women invited to participate in the Norwegian screening program between 1995 and 2009 with follow-up ranging from 1-10 years post-screening. Incidence rate ratios were calculated comparing invited women who did and did not attend screening (divided into prevalent and subsequent screening invitation periods) and for women who had ever and never attended screening in the post-screening age range (starting from 2 years after the last screening invitation). The estimated IRRs were then applied to reference incidence rates for breast cancer in the absence of screening. The inclusion of the post-screening period allowed for lead time adjustment via the compensatory drop method. This method gives over-diagnosis estimates for women who actually attended screening. However, estimates were also given for all invitees, which were derived by multiplying the original estimates by screening compliance rates. Reference rates were obtained using 3 different approaches: a modelling approach in which observed rates for women 40 years of age between 1993 and 1995

were adjusted by applying age-specific relative risks; a cohort approach which used observed rates for the cohort born 1903-1907; and a period approach which used observed rates for the period 1980-1984. The level of over-diagnosis among invitees (which includes those who did not attend) ranged from 10-11% for IBC and 14-17% for all BC, with reference rates using the cohort approach yielding lower estimates than the other approaches. The authors found little evidence of self-selection bias in younger women but did indicate higher incidence in older non-attenders compared with women who were not invited. No adjustment was made for these differences. The over-diagnosis estimates reported in this study are slightly higher than were those reported in other studies which used individual data. This may relate to the relative 'recentness' of screening in Norway compared with other countries.

**Table 5.5. Observational studies using cohort design and/or individual level data to measure over-diagnosis (OD)**

| Author/date | Country Region | General approach | Estimation of incidence without screening | Risk factor adjustment | Lead time adjustment | Post-screening follow-up | OD measure | Denominator | OD estimate |
|---|---|---|---|---|---|---|---|---|---|
| Puliti et al. 2009[81] | Florence, Italy | Observed vs expected incidence without screening (Aggregate data) | Projection of pre-screening trends using Poisson regression | Temporal | Cumulative incidence | 5-15yrs | Ratio observed CI in invited screening cohort / expected without screening | Expected CI in cohort for invited cohort 60-69yrs, 5-15yrs F/U | 0.99 IBC<br>1.01 all BC |
| Puliti et al. 2012[85] | Florence, Italy | Compare incidence in attenders vs non-attenders with 5-15yrs post-screen follow-up | Cumulative incidence over whole period in non-attenders | Temporal trends<br>SES, Marital status | Cumulative incidence | 5-15yrs | % excess CI among attenders / CI among non-attenders | Expected CI in non-attenders to 74yrs (start age 60-69yrs) | 5% IBC<br>10% all BC |
| Njor et al. 2013[84] | Denmark | Compared observed vs expected CI based on comparison with historical cohort adjusted for trends in non-screening regions | Incidence in historical control group (same age, same region for pre-screen period) adjusted for trends in non-screen regions | Temporal trends<br>Regional diff. (explored self-selection bias) | Cumulative incidence | 8+yrs | % excess CI invited cohort with screening / expected without screening | Expected CI in invited cohort with 8+yrs post-screen follow-up (start age 56/59yrs) | 2.3% all BC |
| Coldman, Phillips 2013[223] | British Columbia, Canada | Compared incidence among active screeners including 5yrs post-screen follow-up with incidence in non-screeners | 1.Incidence in non-screeners estimated by subtracting active screeners data from population-wide data | Not adjusted for BC risk difference | Cumulative incidence | 5+yrs | % excess CI among<br>1. Screened / non-screened women | Observed CI in non-screened women/eligible population | Screened:<br>5.4% IBC<br>17% All BC<br>Eligible: |
| | | | 2.Expected incidence via projection of pre-screening trend | Temporal trends | Compensatory drop | Up to age 89yrs | 2. Observed /expected CI in eligible population | 40-89yrs | -1.7% IBC<br>6.7% All BC |
| Falk et al. 2013[83] | Norway | Estimated IRR for screened, post-screened and never-screened applied to reference IRs without screening based on 3 different approaches | 3 different reference rates:<br>- Age-specific RR x incidence at age 40yrs in 1993-1995<br>-Period 1980-1984<br>-Birth cohort 1903-7 | Temporal<br>Regional diff (explored self-selection bias) | Compensatory drop (cumulative incidence) | Follow-up to 79yrs | % Excess CI among women who screen as per recommendations /CI without screening [For invited/attended] | Expected CI without screening from age 50 -79yrs | Invited:<br>10-11% IBC<br>14-17% all BC<br>Attended:<br>11-13% IBC<br>17-20% all BC |

## SUMMARY

Estimation of over-diagnosis is complex. A variety of methods has been used to measure the extent of over-diagnosis due to mammography screening, resulting in an extremely wide range of estimates (0-54%). A large proportion of the studies have design flaws which could have led to biased estimates of over-diagnosis, generally resulting in overestimation. Accounting for lead time effects and ensuring equivalent background risk of breast cancer in the comparison groups or comparison populations are the two most crucial aspects of study design. The timing of studies in relation to the start of screening and length of follow-up after screening has been completed are also important if there is to be allowance for adequate lead time adjustment and to ensure that estimates are not inflated due to elevated proportions screening for the first time. The potential for bias due to inadequate risk or lead time adjustment is shown in Table 5.6 for various experimental and observational studies that have reported estimates of breast cancer over-diagnosis.

Studies which have used individual level data on exposure and outcomes are likely to offer the most reliable estimates. Fitting this description are studies based on long term follow-up of the original screening trials and observational cohort studies where screening histories and breast cancer incidence are known at an individual level. Of the original screening trials, only the Malmo and Canadian trials were conducted in a way that allowed long term follow-up of participants in both the intervention and control arms. An important advantage of trial data is that risk differences in the comparison arms should be negligible, so long as the lack of 'double blinding' does not lead to important bias. However, questions have been raised about the generalisability of trial results to modern screening technologies and screening protocols. While offering relatively precise measures of exposure and outcomes, observational cohort studies that use individual level data may produce biased estimates if those attending screening have higher or lower levels of risk for breast cancer than those who do not attend (i.e. self-selection bias). Though attempts have been made to control for potential confounders, none of the recently published cohort studies has been able to comprehensively adjust for breast cancer risk factors at the individual level, so there is

potential for estimates of over-diagnosis to be inflated if risk of breast cancer is higher among screening participants.

Observational studies using aggregate level data are likely to provide less reliable estimates of over-diagnosis than those based on individual level data, since breast cancer incidence in the absence of screening must be inferred from other populations or time periods, and cannot be observed directly. Projection of pre-screening incidence rates assumes no change in background trends since screening programs were implemented, which may not be the case if prevalence of risk factors for breast cancers has changed (as may apply for example for fertility rates, HRT use, body mass index, diabetes prevalence, etc.).

In dynamic populations where long-term follow-up of individuals is not possible (i.e. where the cumulative incidence method cannot be applied), lead time effects are generally taken into account using either the compensatory drop method or the rate shift method. Studies which have used the rate shift method to adjust for lead time may overstate the extent of over-diagnosis since shifting rates by the average lead time does not account for the ongoing additive effects of lead time in a dynamic population. The extent to which the compensatory drop method fully captures lead time effects is dependent on the timing of the study, in that the vast majority of women in the post screening age group (e.g. 70-79 years) must have had the opportunity to participate in the screening program. Some studies have not used this method appropriately.

In general, estimates from modelling studies have been lower than those from studies using trial data or from observational studies. Modelling approaches generally rely heavily on assumptions about screening parameters, such as the duration and distribution of lead time, which cannot be directly validated. One possibility is that modelling studies have systematically over estimated lead time length and therefore underestimated over-diagnosis. This however, is difficult to demonstrate empirically. On the other hand, some modelling studies have used simulated populations which ensure comparability of breast cancer risk. Where this is the case estimates will not be distorted by risk differences in comparison populations, which is a major problem that most likely leads to overestimation in observational studies.

The most crucial factor explaining the variation in estimates across studies is the way in which measures of over-diagnosis are calculated and reported. Estimates of over-diagnosis are highly dependent on whether the denominator population includes the post-screening age groups or not. This is clearly evident from the summary of study findings shown in Table 5.6. Restricting the denominator to cancers diagnosed among women in the screening age range or to screen-detected cancers would result in substantially higher measures of over-diagnosis. On the other hand, including all cancers occurring over an extended period after women stop screening will result in much lower estimates. Including post-screening ages better reflects the definition of over-diagnosis, which refers to cancers that would not have become clinically evident during a woman's remaining lifetime, and is therefore the more appropriate denominator.

Estimates of over-diagnosis due to mammography, based on results from the least biased studies which have used individual level data and have included post-screening age groups in the denominator, range from 2-17%. Generally these estimates are referring to the excess number of cancers (including DCIS) among women invited to screen as a proportion of cancers expected without screening from the age when screening starts to at least 10 years after screening ends. With the exception of the finding of 2% over-diagnosis in the study by Njor et al.[84], estimates from trials and observational studies based on individual level data are reasonably consistent (10-14%). Estimates from the least biased observational studies using aggregate level data, with post-screening ages included in the denominator, range from 1-10%, but are less reliable due to imprecision in measures of exposure and outcome. Studies that have expressed over-diagnosis as the excess proportion of cancers among women of screening age only have invariably reported much higher levels of over-diagnosis (22-57%). However, expressing these estimates in terms of risk over the remaining lifetime would result in considerably lower levels of over-diagnosis (i.e. at least 2 fold lower). Denominator issues aside, many of the latter studies have not adjusted adequately for lead time and or for differences in breast cancer risk in comparison populations so estimates of over-diagnosis are likely to be inflated in any case. Estimates from modelling studies, which include post-screening age groups in the denominator, range

from negligible to around 5% for invasive cancers. It is however, difficult to assess the extent of bias in these studies.

Based on the evidence from the least flawed studies, it is likely that the level of over-diagnosis is towards the more modest end of the range. However, it is difficult to give a precise estimate due to uncertainties inherent in all methodological approaches.

**Table 5.6. Summary of previous studies measuring over-diagnosis (grouped according to denominators used to report measures of over-diagnosis)**

| Studies | Data | Likelihood of bias due to: | | Population | Denominator | Estimate All BC | Estimate IBC |
|---|---|---|---|---|---|---|---|
| | | Inadequate risk adjustment | Inadequate lead time adjustment | | | | |
| Moss 2005 (Canadian I &II)[70] | individual (RCT) | low | low | invited | screen + post-screen | 11-14% | - |
| Zackrisson et al. 2006 (Malmo)[71] | individual (RCT) | low | low | invited | screen + post-screen | 10% | - |
| UK Independent Panel 2012[39] | individual (RCT) | low | low | invited | screen + post-screen | 10-14% | - |
| Peeters et al. 1989[72] (later periods) | aggregate/population | high | low | eligible | screen + post-screen | 1% | - |
| Paci et al. 2004[79] | aggregate/population | high | low | eligible | screen + post-screen | 5% | 2% |
| Paci et al. 2006[78] | aggregate/population | low | low | eligible | screen + post-screen | 5% | 3% |
| Waller et al. 2007[68] | aggregate/population | low | low | eligible | screen + post-screen | 10%* | - |
| Puliti et al. 2009[81] | aggregate/cohort | low | low | eligible | screen + post-screen | 1% | -1% |
| Kalager et al. 2012[221] | aggregate/population | low | high | eligible | screen + post-screen | - | 15-25% |
| Puliti et al. 2012[85] | individual/cohort | low | low | screened | screen + post-screen | 10% (screened) | 5% (screened) |
| Njor et al. 2013[84] | individual/cohort | low | low | eligible | screen + post-screen | 2.3% | |
| Falk et al.2013[83] | individual/cohort | low | low | eligible | screen + post-screen | 14-17% | 10-11% |
| Coldman, Phillips 2013[223] | individual/cohort | low | high | screened | screen + post-screen | 17% (screened) | 5.4% (screened) |
| Coldman, Phillips 2013[223] | aggregate/population | low | low | eligible | screen + post-screen | 6.7% | -1.7% |
| Cochrane review (controls not screened)[212] | individual (RCT) | low | high | invited | screen + limited post-screen | 25% | - |
| Cochrane review (controls screened)[212] | individual (RCT) | low | high | invited | screening age only | 29% | - |
| Miller et al. 2014 (Canadian I & II)[210] | individual (RCT) | low | low | invited | screen-detected only | 22% | - |
| Zahl et al. 2004[74] | aggregate/population | high | high | eligible | screening age only | 45-56% | - |
| Jonsson et al. 2005[218] | aggregate/population | low | high | eligible | screening age only | 54% (50-59yrs) | - |
| Jonsson et al. 2005[218] | aggregate/population | low | high | eligible | screening age only | 21% (60-69yrs) | |
| Zahl et al. 2008[54] | aggregate/population | high | high | eligible | screening age only | 22% | - |
| Jorgensen & Gotzsche 2009[82] | aggregate/population | high | high | eligible | screening age only | 44-57% | - |
| Jorgensen et al. 2009[75] | aggregate/population | high | low | eligible | screening age only | 33% | - |
| Duffy et al. 2010[69] | aggregate/population | low | low | eligible | screening age only | - | 4% |
| Morrell et al. 2010[77] | aggregate/population | low | high | eligible | screening age only | 30-42% | - |
| Zahl & Maehlen 2012[219] | aggregate/population | high | high | eligible | screening age only | 50% | - |

Excludes complex modelling studies and studies focussing on screening in women aged under 50 years

# Chapter 6: A novel case-control design to estimate the extent of over-diagnosis of breast cancer due to organised population-based mammography screening.

Paper published in the International Journal of Cancer 2014 (online ahead of print).

Authors: Kerri R Beckmann, John W Lynch, Janet E Hiller, Gelareh Farshid, Nehmat Houssami, Stephen W Duffy, David M Roder

## PREFACE

The study presented in this chapter is the first of two approaches used to estimate the level of over-diagnosis of breast cancer due to organised population-based mammography screening in an Australian setting (South Australia). This study uses a novel case-control design with individual person-level data on exposure to screening and long-term follow-up from the last screening mammogram to avoid reliance on assumptions about lead time distributions. The case-control design allows relative increases and decreases in breast cancer incidence to be estimated for various time intervals after screening. This involves using conditional logistic regression to compare the screening histories of women diagnosed with breast cancer and age-matched controls who were free of breast cancer. Since breast cancer is a relatively rare event, odds ratios will approximate the relative risk (risk ratio) of breast cancer among women who attended screening compared with women who did not. The resulting pattern of odds ratios over the follow-up period represents the lead time effects due of screening (i.e. the initial increase in incidence at the time of screening due to the advancement of diagnoses, and the compensatory drop after screening is complete due to the depletion of cancers that were detected at an earlier time through screening). The extent of over-diagnosis can be estimated by applying the resulting odds ratios to a set of reference incidence rates to calculate the total incidence over an extended follow-up period for the population who have participated in screening and comparing this with the expected total incidence in the absence of screening (based on the reference rates).

## AUTHORSHIP STATEMENT

**A novel case-control design to estimate the extent of over-diagnosis of breast cancer due to organised population-based mammography screening**.
*International Journal of Cancer* 2014. [E-pub ahead of print]

**Kerri Beckmann (Candidate)**
I conceived and designed the study, managed data collation, analysed and interpreted the data, drafted and critically revised the manuscript, coordinated contributions from co-authors and acted as corresponding author for peer review and preparation for publication.

25/09/14
Signed …… Date……

**John Lynch**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

23/09/14
Signed …… Date……

**Janet Hiller**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**Gelareh Farshid**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**Nehmat Houssami**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**Stephen Duffy**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

23/09/14
Signed …… Date……

**David Roder**
My contribution to this publication developing the research methodology, assisting with interpretation of findings and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

22/09/14
Signed …… Date……

## ABSTRACT

Debate about the extent of breast cancer over-diagnosis due to mammography screening has continued for over a decade, without consensus. Estimates range from 0-54%, but many studies have been criticised for having flawed methodology. In this study we used a novel study design to estimate over-diagnosis due to organised mammography screening in South Australia (SA). To estimate breast cancer incidence at and following screening, we used a population-based, age-matched case-control design involving 4,931 breast cancer cases and 22,914 controls to obtain OR for yearly time intervals since women's last screening mammogram. The level of over-diagnosis was estimated by comparing the cumulative breast cancer incidence with and without screening. The former was derived by applying ORs for each time window to incidence rates in the absence of screening, and the latter, by projecting pre-screening incidence rates. Sensitivity analyses were undertaken to assess potential biases. Over-diagnosis was estimated to be 8% (95% CI 2 to 14%) and 14% (95% CI 8 to 19%) among SA women aged 45-85 years from 2006-2010, for invasive breast cancer and all breast cancer respectively. These estimates were robust when applying various sensitivity analyses, except for adjustment for potential confounding assuming higher risk among screened than non-screened women, which reduced levels of over-diagnosis to 1% (95% CI -5 to 7%) and 8% (95% CI 2 to 14%) respectively when incidence rates for screening participants were adjusted by 10%. Our results indicate that the level of over-diagnosis due to mammography screening is modest and considerably lower than many previous estimates, including others for Australia.

## INTRODUCTION

Mammography screening has been shown to be effective in reducing breast cancer mortality, both through randomised controlled trials[24, 36, 38, 39, 224] and observational studies of population-based screening programs[33, 34, 50]. Concerns have been raised about the potential for mammographic screening to lead to the over-diagnosis of breast cancers, i.e. the detection of cancers that would not have presented in a woman's lifetime had she not participated in screening[56, 198]. While some over-diagnosis is inevitable due to early deaths from causes other than breast cancer, it is also possible that screening may detect 'indolent' or very slow growing tumours that would never present symptomatically during a woman's remaining life-time. Treatment for such screen-detected cancers would constitute a harmful unintended consequence of breast screening programs.

Effective screening requires that there be a preclinical phase of the disease, when the cancers are detectable using screening technology, preceding the symptomatic phase. Lead time (the amount of time by which the diagnosis is advanced by screening) must therefore be sufficient to enable benefits from earlier treatment. Screening, will initially lead to higher age-specific incidence rates during screening, due to the shift to earlier diagnosis. Following the cessation of screening, with no over-diagnosis, incidence should decline to compensate for the elevated increase during screening so as to eventually produce similar cumulative incidence as in a non-screened population after lead time has passed[53, 198]. Any excess in the number of cancers detected at screening compared with the deficit following screening provides a measure of over-diagnosis, given sufficient follow-up time and assuming no other influences on background incidence rates.

The extent of over-diagnosis due to mammography screening is contested. Estimates of the level of over-diagnosis are extremely varied, with reports ranging from 0-54%[3, 89, 198]. While some variation may be attributed to differences in policies and practices across screening programs, inadequate adjustment for lead time and risk differences, and differences in the denominators used, are likely to account for a substantial part of this variation[89]. With the exception of the original screening trials, most early studies used aggregate rather than individual level data to quantify over-diagnosis. Such

studies are unable to measure actual exposures of individual women to screening and hence accurately determine the expected decline in incidence after women have completed screening. Alternative approaches using statistical adjustment for lead time rely on various assumptions including, most critically, the assumed distribution of lead time effects[3]. While designs based on long-term follow-up of the original mammographic screening trials offer the most accurate (least biased) estimates, their generalisability for modern service screening programs could be questionable, given that trial findings applied to the use of older screening technology, breast cancer incidence pertaining to the 1960's to 1990's, and often different screening protocols[89]. More recently, several studies have been undertaken using individual level data that capture women's actual exposure to screening[81, 83-85, 223]. These studies provide more reliable estimates than those based on aggregate data, although there is still some variation, with estimates ranging from 2% to 17% for invasive and in-situ cases combined.

The objective of this study was to estimate the level of over-diagnosis of breast cancer due to organised population-based mammography screening in an Australian setting (South Australia). We have used a novel case-control study design with individual person-level data on exposure to screening. Through long term follow-up from the last screening mammogram we have avoided reliance on assumptions about lead time distributions.

## MATERIALS AND METHODS

### The screening program

BreastScreen SA (BSSA) is the South Australian component of Australia's national mammography screening program. It operates as a State-wide service, providing population-based screening through both fixed and mobile units to eligible women at no cost. The program began as a pilot study in 1989 with state-wide rollout from 1991. Women aged 40 years or older with no symptoms of breast cancer are eligible to participate, however the program actively targets women aged 50-69 years through written invitations to those listed on the SA Electoral Roll (extended to 74 years following a policy change in 2013). Screening is generally undertaken every two years, however women with a strong family history of breast cancer are eligible for annual

screening. During the study period, screening involved analogue screen-film technology with two views independently read by two radiologists, with a third reading if results were discordant. The program has been continually evaluated since its inception, with ongoing data collection on participation and screening outcomes.

**Overall study design**

Estimation of over-diagnosis involved two distinct steps. The first step was a case-control study to estimate the increased or decreased breast cancer incidence in women exposed to screening for successive time windows after the most recent screening episode, relative to not having participated in screening. Women were categorised as exposed if they had undergone screening at BreastScreen SA at least once (~70% of the study population), and not exposed if they had never participated in the screening program. Odds ratios (ORs) for breast cancer were determined for each of the time windows since the last screening mammogram. The second step involved calculating the cumulative incidence within the screening population-based on the ORs from the case-control study, the proportion of controls in each of the time windows and the background incidence rates had screening not been introduced. Background rates were derived through projection of pre-screening incidence rates. The level of over-diagnosis was interpreted to be the cumulative excess number of cancers within the population exposed to screening compared with the number expected without screening.

Ethical approval was obtained from the Human Research Ethics Committees of SA Health and the University of Adelaide. Participant consent was not required since all analyses were undertaken using de-identified data sets and the study did not involve contact with participants.

**Study population**

Cases were selected from the South Australian Cancer Registry (SACR) and consisted of all women diagnosed with breast cancer, between January 1, 2006 and December 31, 2010, who were aged between 45 and 85 years and resident in SA at the time of diagnosis. We included both invasive breast cancers (IBC) and ductal carcinoma in-situ (DCIS) as cases. If an individual was diagnosed with both DCIS and IBC during the study period (n=12), each event was included as a separate ‘case’ requiring a different set of

matched controls. Due to the way the data were provided to us it was not possible to exclude these 12 cases. The SACR does not record second primary invasive cancers so only the first IBC case was included.

Five age-matched controls per case with the same month and year of birth were randomly sampled from the South Australian electoral rolls (ER) using an incidence density sampling approach. Incidence density sampling allows the same control to be selected more than once in relation to a different case, and a case to be selected as a control, provided their diagnosis date was after the date of diagnosis of the case they were selected to match. The electoral roll was chosen as the sampling source for controls, since invitations to participate in the screening programs are issued to all women on the ER who are within the target age (approximately 93% of residents in the target age range). To ensure an equivalent chance of being invited to participate in screening we also restricted the eligibility of cases to those on the ER.

Probabilistic record linkage was undertaken to establish the eligibility of cases and controls. For cases, this involved linkage between the case file (SACR data) and ER data. Cases with any diagnosis of breast cancer before the study period were excluded. For controls, record linkage was undertaken between the control file (ER data) and death records from Births, Death and Marriages data for SA to determine date of death, and also with the SACR to determine the date of diagnosis of any previous breast cancer. Controls who had died or were diagnosed with breast cancer before the date of diagnosis of their respective case were excluded.

To determine screening histories for cases and controls, both files were linked with BSSA records, using probabilistic record linkage software (Linkage Wiz[225] and Link Plus[226]). BSSA records included all women invited to participate since the inception of the program, as well as women who have contacted the service of their own accord, irrespective whether they actually undertook mammography screening through the program.

All record linkage tasks were undertaken by the Epidemiology Branch within the SA Department of Health. Each of these linkage processes involved extensive manual clerical review to ensure a high level of accuracy with regard to linked and non-linked

records at all steps. Data provided to the research team were stripped of potentially identifying information and all analyses were performed with anonymised data sets.

**Measures**

The main outcome measure in the case-control study was histologically confirmed diagnosis of breast cancer from the SACR. The key exposure variable was the time that had elapsed since the date of the most recent screening episode with BSSA, prior to diagnosis and their diagnosis date (month and year) or, for controls, the diagnosis date of their matched case. Time was further categorised into discrete successive time windows (i.e. single years since the last screening mammogram), with never having attended screening being used as the reference category.

Socioeconomic position (SEP) and area of residence were included as covariates in our analyses. SEP was assigned on the basis of women's residential postcode at the time of the case's diagnosis ( ascertained from the source data flies, i.e. the SACR for cases or the ER for controls), using the Australian Bureau of Statistics (ABS) Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD) for 2006, which is a measure of the average level of access to resources within society at the neighbourhood level[141]. Women were grouped into quintiles of SEP according to the distribution of IRSAD scores across SA. Similarly, area of residence was assigned according to postcode, using the Accessibility and Remoteness Index for Australia (ARIA+2006 release), also developed by the ABS[227]. ARIA classifies localities according to accessibility to service centres based on road distances. Categories include highly accessible (inner urban), accessible (outer urban), moderately accessible (rural), remote and very remote.

**Analysis**

We used conditional logistic regression to determine ORs for each time window since the last screening mammogram for cases compared with controls. Separate analyses were undertaken for all breast cancer cases (IBC and DCIS) and IBC alone. All logistic regression models included SEP and area of residence as potential confounders. In each case, separate age-stratified models were run to obtain ORs for time windows for each of the 10-year age groups.

To estimate background incidence rates as if breast screening had not been implemented, we projected the breast cancer incidence for SA from the pre-screening era (sourced from the SACR, 1977-1988) using Age-Period Poisson regression with single year of age and calendar year as continuous variables in the model. The estimated residential population (ERP) for SA females[105] for each year of age in each calendar year was used as the denominator. More complex prediction modelling (e.g. age-period-cohort modelling) was found not to be possible due to the relatively short period of observation prior to the establishment of Breast Screen SA (i.e. 12 years).

Our estimates of over-diagnosis (OD) were based on the difference between the cumulative incidence of breast cancer among women exposed to the screening program (derived from the case-control study) and the expected cumulative incidence among that population had they not been exposed to screening, as described in the following formula:

$$OD\% = \frac{CUM\ INC\ (exposed\ to\ screening) - CUM\ INC\ (not\ exposed\ to\ screening)}{CUM\ INC\ (\ exposed\ to\ screening)} * 100$$

This formula corresponds to the UK independent panel's preferred method for reporting over-diagnosis from a population perspective[39].

The cumulative incidence of breast cancer in the absence of the screening program was derived from the projected incidence rates for SA and the ERP for 2006-2010, calculated for 10 year age groups and then summed.

The cumulative incidence in the presence of screening was determined by taking into account the effects of screening at and after the most recent screening episode, based on odds ratios from the logistic regression. Subtotals were calculated for each age stratum separately by summing the number of cancers expected across all time windows. To determine the number of cancers expected within each time window the background age-specific incidence rate was multiplied by the odds ratio and by the number at risk within each time window. The number at risk in each time window was determined by multiplying the proportion within each time window, based on the distribution across the time windows within the control group, by the total person years in the age stratum (from ERP data). Subtotals for each age stratum also included the number of cancers among the women who had never screened (derived from the

background incidence rate multiplied by the proportion of controls who had never screened and by the total person years in that age group). Further detail and rationale are presented in the Appendix of this chapter.

Bootstrapping[228] was undertaken to derive confidence intervals for our final estimates of over-diagnosis. This method involved repeated sampling with replacement, clustered by matched ID, with 200 cycles, to generate a non-parametric distribution of estimates of over-diagnosis based on repeated measures of OR and the screened proportions across time windows.

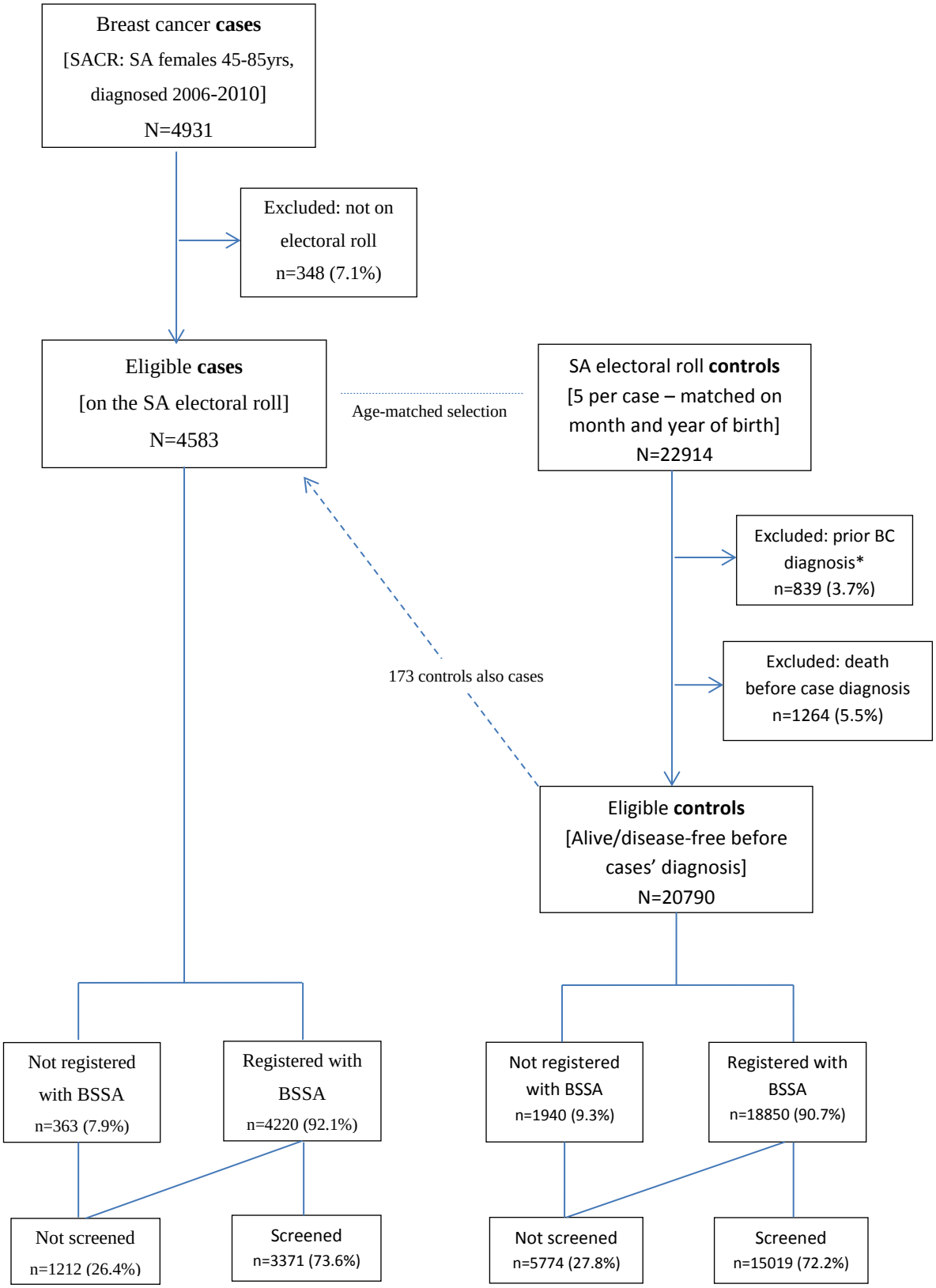
**Sensitivity analyses**

A series of sensitivity analyses were undertaken to examine potential sources of bias that may have distorted our estimates of over-diagnosis. To assess whether results were biased by any misclassification of screening exposure through linkage errors, we restricted analysis to only those participants registered with BSSA. To examine the impact of different 'background' incidence rates we conducted sensitivity analyses using pre-screening incidence rates for SA as the reference 'non-screening' incidence rates. In addition, we tested the effect of altering the age-specific incidence rates, by increasing and decreasing the slope for the age trend by 20%. Finally, to account for any unmeasured confounding due to screening participants potentially having a higher risk of breast cancer compared with non-participants (based on recent SA survey findings[229]), we recalculated our estimates of over-diagnosis by decreasing the incidence rate for screening participants by 2 different orders of magnitude (5 and 10%) and applying these revised rates across each of the screening time windows.

We also investigated the impact of considering shorter follow-up periods after the last screening episode, since there has been considerable debate about what constitutes an adequate follow-up period after screening to account fully for lead time effects. To do this we repeated our analysis with time windows truncated at 5 years (and 8 years a second analysis) and substituted an odds ratio of 1 for periods beyond 5 years (or 8 years) follow-up.

All analyses were undertaken using Stata v12.0[143].

**Figure 6.1. Design of the case-control study to estimate over-diagnosis through population-based mammography screening in South Australia**



## RESULTS

Figure 6.1 shows participant selection and exclusions. Of the 4931 cases selected from the cancer registry, 93% (4583) were identified on the South Australian electoral roll and were therefore eligible. This proportion is consistent with the estimated proportion of SA residents who were registered to vote with the Australian Electoral Commission[230]. Of the 22,914 age-matched controls selected from the ER, 839 (3.7%) were excluded because they had a prior breast cancer diagnosis and 1,264 (5.5%) because they had died before the diagnosis date of their respective case, leaving a total of 20,790 eligible controls. This included 173 selected controls who were also eligible cases (due to a subsequent diagnosis of breast cancer within the study period). Ninety two percent of cases and 91% of controls were registered with BSSA (i.e. had either been invited for screening or contacted the service themselves). Seventy four percent of all eligible cases and 69% of all eligible controls had undergone mammography screening at least once with BSSA prior to the start of this study. Eighty nine percent of eligible cases had invasive cancers while 11% had a diagnosis of DCIS. Characteristics of eligible cases and controls were similar, as shown in Table 6.1.

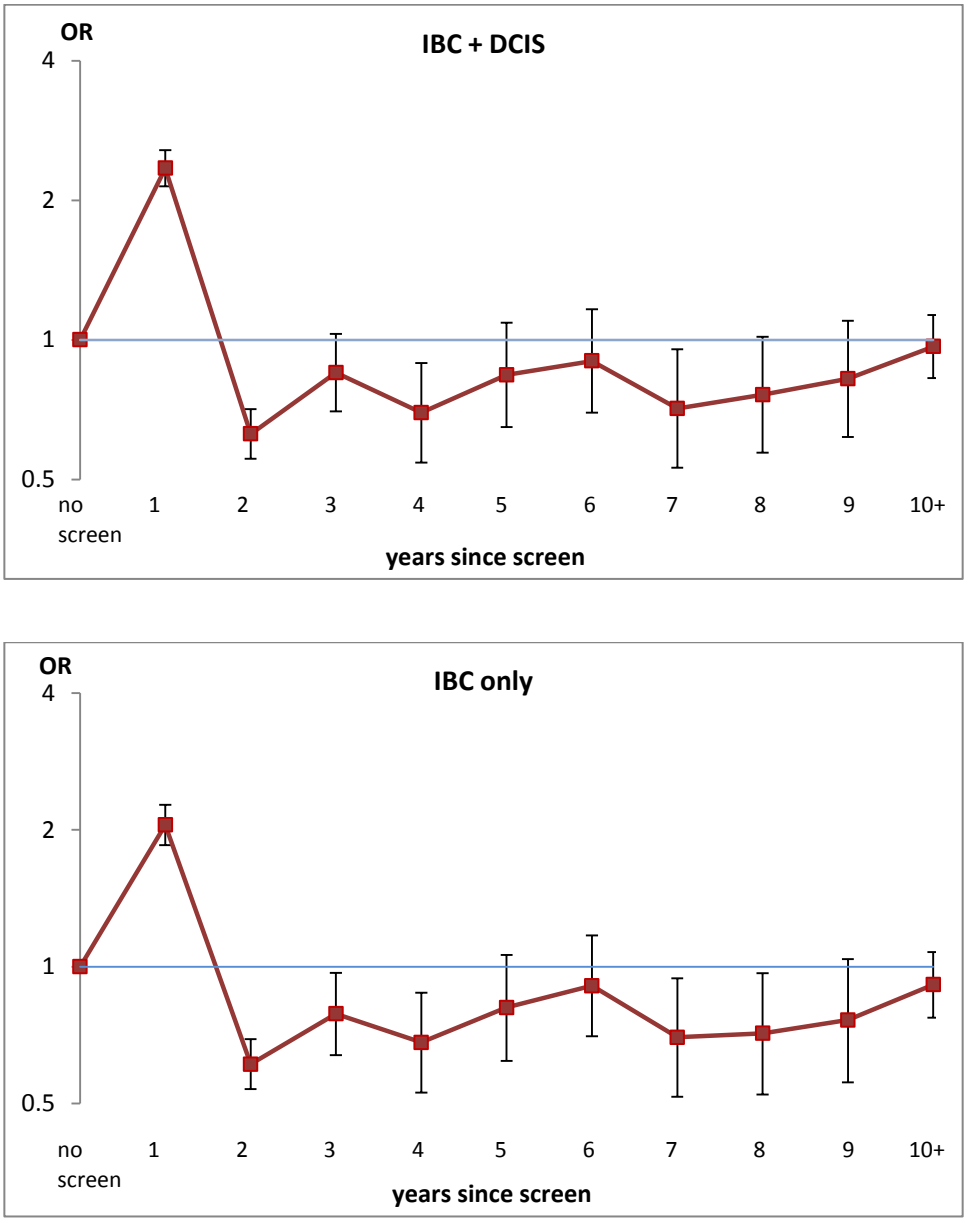
The distribution of time since the last screening mammogram and the date of diagnosis (of the case), however, differed between cases and controls (Table 6.1). Compared with controls, a larger proportion of cases had been screened within a year of diagnosis (46% vs 26%), while a smaller proportion had never had a mammogram at BSSA (27% vs 31%) or had been screened 1-2 years before diagnosis (10% vs 20%). This pattern is consistent with lead time effects resulting in higher incidence at screening and lower incidence immediately after screening.

**Table 6.1 Characteristics of eligible cases and controls**

| Factors | cases | | controls | |
|---|---|---|---|---|
| | Number | Percent | Number | Percent |
| **Total eligible** | 4583 | 100.0 | 20790 | 100.0 |
| **Age group** | | | | |
| 45-54yrs | 1204 | 26.3 | 5880 | 28.3 |
| 55-64yrs | 1520 | 33.2 | 7154 | 34.4 |
| 65-74yrs | 1109 | 24.2 | 4890 | 23.5 |
| 75-85yrs | 750 | 16.4 | 2866 | 13.8 |
| **SEP (area)** | | | | |
| Lowest quintile | 753 | 16.5 | 3851 | 18.6 |
| Low-mid | 949 | 20.7 | 4263 | 20.6 |
| Middle | 889 | 19.4 | 4280 | 20.6 |
| Mid-high | 954 | 20.8 | 4026 | 19.4 |
| Highest quintile | 1033 | 22.6 | 4318 | 20.8 |
| missing | 5 | - | 55 | - |
| **Area of residence** | | | | |
| Inner metropolitan | 3301 | 72.3 | 14052 | 69.9 |
| Outer metropolitan | 669 | 14.6 | 2940 | 14.2 |
| Rural | 446 | 9.7 | 2539 | 12.2 |
| Remote | 137 | 3.0 | 642 | 3.1 |
| Very remote | 25 | 0.6 | 115 | 0.6 |
| missing | 5 | - | 55 | - |
| **Cancer type** | | | | |
| Invasive | 4088 | 89.2 | | |
| Ductal carcinoma in-situ | 495 | 10.8 | | |
| **Record of contact with BSSA** (invitation or self-contact) | | | | |
| Yes | 4220 | 92.1 | 18850 | 90.7 |
| No | 363 | 7.9 | 1940 | 9.3 |
| **Participated in screening at BSSA** | | | | |
| Yes | 3371 | 73.5 | 14294 | 68.7 |
| No | 1213 | 26.5 | 6496 | 31.3 |
| **Years since last screen** | | | | |
| No screening record | 1213 | 26.5 | 6496 | 31.3 |
| 0 - <1yr | 2091 | 45.6 | 5314 | 25.6 |
| 1 - <2yrs | 445 | 9.7 | 4165 | 20.0 |
| 2 - <3yrs | 142 | 3.1 | 922 | 4.4 |
| 3 - <4yrs | 80 | 1.8 | 589 | 2.8 |
| 4 - <5yrs | 76 | 1.7 | 453 | 2.2 |
| 5 - <6yrs | 79 | 1.7 | 390 | 2.1 |
| 6 - <7yrs | 57 | 1.2 | 375 | 1.9 |
| 7 - <8yrs | 60 | 1.3 | 351 | 1.8 |
| 8 - <9yrs | 61 | 1.3 | 301 | 1.7 |
| 9 - <10yrs | 44 | 1.0 | 237 | 1.5 |
| 10+yrs | 210 | 4.6 | 758 | 3.6 |
| **Average no. screening rounds**[a] (excluding never screened) | | | | |
| mean (range) | 5.1 (1-18) | | 4.9 (1-16) | |

[a] based on controls' most recent screen prior to respective cases diagnosis date

**Figure 6.2 Odds ratios[a] for the most recent screening mammogram at BreastScreen SA within discrete time periods since the last screening episode among cases (with breast cancer) compared with controls (without breast cancer), all ages 45-85yrs**




[a]Derived from conditional logistic regression models including all eligible participants adjusted for socioeconomic position and area of residence.

**Table6.2 Estimates of over-diagnosis due to organised mammography screening, for all breast cancer and invasive breast cancer only, South Australia 2006-2010.**

| | | All Breast Cancers | | | Invasive Breast Cancers | | |
|---|---|---|---|---|---|---|---|
| Age group | Person years | Projected incidence rate/100,000 | No. cancers without screening | No. Cancers with screening | Projected incidence rate/100,000 | No. cancers without screening | No. Cancers with screening |
| 45-54yrs | 573868 | 176 | 1010 | 1394 | 161 | 924 | 1192 |
| 55-64yrs | 490675 | 223 | 1094 | 1354 | 206 | 1011 | 1139 |
| 65-74yrs | 323303 | 284 | 918 | 1010 | 263 | 850 | 873 |
| 75-85yrs | 267409 | 363 | 971 | 873 | 338 | 904 | 790 |
| Cumulative incidence | | | 3993 | 4631 | | 3689 | 3993 |
| % over-diagnosis (CI) | | | | 13.8 (8.1, 18.7) | | | 7.6 (1.6-13.6) |

% Over-diagnosis = [(Cum Inc. with screening – Cum Inc. without screening)/Cum Inc. with screening]*100

Estimates apply to the total population who would have been eligible for screening at BSSA (i.e. SA females 45-85yr)

CI - Confidence intervals (derived by bootstrapping, 200 replications)

Note: Over diagnosis estimates apply to the total population who would have been eligible for screening at BSSA for ages 45-85yrs. Excess cancers within age strata do not indicate the level of over-diagnosis for individual age groups since the proportion with adequate post screening follow-up will vary, and will be insufficient to adjust for lead time in the younger age groups since most will still be actively screening.

Results of conditional logistic regression for the whole study population, showing the odds ratio by yearly time windows with respect to the most recent screening episode (Figure 6.2), reflect this pattern of difference between cases and controls. For all ages combined, the odds of having been screened within 12 months of the diagnosis date among cases was double that of controls ($OR_{all\ BC}$ =2.35, 95% CI 2.14-2.57; $OR_{IBC}$=2.05, 1.67-2.25), while the odds of having been screened in the 1-2 years prior to diagnosis was less than two thirds that of controls ($OR_{all\ BC}$= 0.63, 0.55-0.71; $OR_{IBC}$=0.61, 0.53-0.69). The odds of screening within a specific time window tended toward no difference as time since screening increased.

From Poisson regression models, IBC incidence was predicted to have increased by 2.5% (95% CI 2.3-2.7%) per year for each additional year of age and by 0.8% (0.02-1.6%) per year for each calendar year. For all BC, incidence increased by 2.4% (2.3-2.7%) and 1.1% (0.03-1.9%) per year, for age and calendar year respectively. The expected mean annual age-specific incidence rates in the absence of screening, which represent the background non-screening incidence in SA for the period 2006-2010, are shown in Table 6.2 along with the expected incidence for each age stratum with and without screening, the cumulative incidence, and subsequent estimates of the proportion of over-diagnosed cases. Over-diagnosis of invasive breast cancers was estimated to be 7.6% (95%CI 1.6-13.6%) of all cancers diagnosed in the population, based on analyses using all eligible study participants, and 13.8% (8.1-18.7) for IBC and DCIS combined.

When analysis was restricted to those participants registered with BreastScreen SA estimates of over-diagnosis were 7.1% (-1.3 – 15.6%) of IBC and 14.0% (9.2-18.7) for all BC (Table 6.3). Sensitivity analyses, using the mean annual incidence rates for the pre-screening era in SA as the background reference incidence rates without screening, produced very similar estimates of over-diagnosis in relation to all breast cancers and IBC alone. Changing the age trends by 20% however did have a small impact on estimates. Decreasing the age gradient resulted in slightly higher estimates (14.5% for all BC and 8.3% for IBC) while increasing the slope resulted in lower estimates of over-diagnosis (13.1% and 6.9% respectively). In further sensitivity analysis in which we adjusted for higher breast cancer risk among screening participants compared with

non-participants based on risk differences noted in a recent survey conducted in SA[229], estimates of over-diagnosis were considerably lower than in the original unadjusted analysis. For example, estimates of over-diagnosis were 1% for IBC and 8% for IBC and DCIS combined after adjustment for 10% higher breast cancer risk among screening participants. Truncation of the follow-up period after screening resulted in higher estimates of over-diagnosis. The estimate for all BC was 16.4% and 15.0% with only 5 years and 8 years of post-screening follow-up. The corresponding figures for IBC were 10.9% and 9.3%.

**Table 6.3 Sensitivity analyses for estimates of over-diagnosis due to organised mammography screening**

| Sensitivity analyses | % over-diagnosis (95% CI) | |
|---|---|---|
| | IBC+DCIS | IBC only |
| Participants registered with BSSA[a] | 14.0 (9.2, 18.7) | 7.1 (-1.3, 15.6) |
| 5% adjustment for increased BC risk[a,b] | 10.7 (4.9, 16.6) | 4.3 (-1.8, 10.5) |
| 10% adjustment for increased BC risk[a,b] | 7.7 (1.8, 13.7) | 1.2 (-5.0, 7.4) |
| Using pre-screening incidence rates (1977-1986)[b] | 13.9 (6.8, 20.8) | 7.6 (2.1, 13.0) |
| Slope for age trend reduced by 20%[a,b] | 14.5 (8.6, 20.5) | 8.3 (2.5, 14.0) |
| Slope for age trend increased by 20%[a,b] | 13.1 (5.7,18.3) | 6.9 (0.9, 12.9) |
| Truncation of post-screening follow-up at 5 years[a,b] | 16.4 (11.6, 21.2) | 10.9 (5.6-16.2) |
| Truncation of post-screening follow-up at 8 years[a,b] | 15.0 (10.1, 20.0) | 9.3 (3.6, 15.0) |

[a] using projected incidence rates for SA 1977-1988

[b] including all eligible participants

CI - Confidence intervals (derived by bootstrapping, 200 replications)

IBC - invasive breast cancer; DCIS - ductal carcinoma in-situ; BSSA – BreastScreen SA; BC – Breast cancer

## DISCUSSION

We estimated the proportion of over-diagnosed breast cancers resulting from organised mammography in the SA population from 2006-2010 to be 8% for invasive breast cancers and 14% for all breast cancer cases including DCIS, without adjustment for differences in risk profiles among screen attendees and non-attendees. Our estimate of over-diagnosis applies to the proportion of breast cancers among all women aged 45-85 years who would have been eligible for screening through BSSA, of which around 70% have participated at least one.

In calculating the extent of over-diagnosis we have assumed that background incidence rates in SA continued to increase at the same rate as had occurred in the pre-screening era (1978-1988). This cannot easily be verified since there is no comparable population which has not been exposed to mammography screening in some form. Furthermore, projections of rates for all breast cancer may be inaccurate due to the practice of SACR overwriting in-situ cancer cases if women subsequently developed invasive cancer, which occurred up until 1996. However, our estimates of over-diagnosis did not vary substantially when pre-screening incidence rates in SA were used as the reference (background) rates, indicating that our over-diagnosis estimates using this novel methodology are largely unaffected by variation in background incidence rates. Our estimates are, however, dependent on the relative difference in rates across age strata. If rates increased more quickly in some age groups compared with others, the extent of over-diagnosis may be over or under-estimated, depending on which age groups experienced faster rate increases. It is plausible that incidence has increased more rapidly among younger than older cohorts due to the more recent trend toward fewer births and older age at first birth, which may tend to under-state the level of over-diagnosis.

On the other hand, our empiric estimates are likely to overstate the level of over-diagnosis, due to unmeasured confounding in relation to breast cancer risk. Although we adjusted our estimates for SEP (at the neighbourhood level) and area of residence in our logistic regression models, other potential confounders could not be included due to lack of data at the individual level. Recent surveys examining differences in breast cancer risk profiles among SA women indicated higher prevalence of several

breast cancer risk factors among breast screen participants compared with non-participants[48, 229]. BSSA participants were more likely to have had a family history of breast cancer, higher prevalence of past or current hormone replacement therapy (HRT) use and increased frequency of breast biopsy or breast surgery. Based on comparisons of summary measures of risk derived using the Pfeiffer breast cancer risk prediction model[231], women who had ever participated in BSSA had 13% higher risk of developing breast cancer based on the 2012 survey. However, given the uncertainties about causality in relation to previous breast biopsy or surgery and use of hormone replacement therapy, the extent to which rates actually differ between participants and non-participants is difficult to determine. Despite the uncertainty, it is unlikely that confounding due to risk differences would result in levels of over-diagnosis higher than 8% (for IBC), since there is no evidence of lower risk among screening participants in our population. Indeed, with adjustment of incidence rates among screening participants to account for a potential 10% higher risk, as informed by the related survey[229], the level of over-diagnosis of IBC was negligible (i.e. 1%).

Our findings demonstrate the importance of an extended period of follow-up after screening for estimating over-diagnosis. Over-diagnosis estimates were considerably higher when we simulated shorter follow-up periods, suggesting that the deficit in incidence after screening continues beyond 8 years of follow-up. We therefore recommend a follow-up period of at least 10 years to fully capture lead time effects when estimating over-diagnosis.

A potential weakness of this study is that screening histories were determined through probabilistic record linkage, with linkage errors leading to misclassification of exposure to breast screening. While considerable effort was made to ensure high quality matching including the use of two different matching programs, the inclusion of maiden names and aliases, where known, and extensive manual review, some misclassification may still have occurred. It is also possible that some individuals who were not resident in SA while in the target age range (hence were true non-matches) had participated in screening in other jurisdictions. In either case, it is unlikely that misclassification would be substantial or differential with respect to case-control status. Given that results were very similar when analysis was restricted to only those

participants registered with BSSA, we are confident that the impact of any misclassification bias is minimal.

Our measure of over-diagnosis does not include the impact of private mammography screening. Survey data indicate that approximately one quarter of South Australian women aged 40-84 years have participated in mammography screening outside of the BreastScreen program[229]. The proportion screening privately did not vary according to whether or not women had participated in the publicly funded screening program. The relative timing of women's use of private and public screening was not ascertainable from these cross-sectional data.

Unlike many other study designs which rely on estimates of lead time to determine the extent of over-diagnosis, this study design makes no assumptions about lead time duration or distribution. Lead time effects are accounted for through the categorisation of screening exposure into discrete time periods, and the availability of cases and controls with ten years or more since screening. OR for each time windows provide an estimate of the relative increases or decreases in incidence following screening, compared with those who never participated in screening. The pattern of ORs we observed is consistent with theoretical expectations, with an increased rate ratio in the index year due to screening bringing cancer diagnoses forward, and decreased rate ratios in subsequent years after leaving the program.

Our estimate of 8% over-diagnosis (without risk adjustment) for invasive breast cancer lies at the more moderate end of the range of estimates published over the past decade and is within the range of estimates from recent studies published by Njor et al.[84] from Denmark (2%), Puliti et al.[85] from Italy (10%) and Falk et al.[83] from Norway (11-13%). Our results are consistent with estimates derived from screening mammography trials of 10% (all BC)[71], 12% (IBC only)[69] and 11-14% (all BC)[70], and with data presented in the recent independent review of breast screening in the UK[39]. It should be noted that our denominator differs from most of these studies. We chose to use the incidence in the population in the presence of screening as our denominator, which corresponds with the UK Independent Panel's preferred option for expressing over-diagnosis from a population perspective[39], rather than the expected incidence in

the absence of screening. Our estimates would be slightly higher if expressed as the proportion of cancers expected in the absence of screening (i.e. 8.2% of IBC).

In contrast, several authors have reported considerably higher levels of over-diagnosis, in the order of 30% or more[75, 77, 80, 82, 232]. It has been argued in a recent review that most studies reporting high proportions of over-diagnosis had not adequately adjusted for lead time effects and, or for breast cancer risk[89]. Only 2 studies have reported over-diagnosis estimates for an Australian population, both of which greatly exceed our estimates. Morrell et al[77] reported 30-42% over-diagnosis (IBC only) among 50-69 year old women in New South Wales, while Jorgensen and Gotzsche[82] reported 53% over-diagnosis (including in-situ cancers). In the latter study, estimates of over-diagnosis were based on relative rate ratios of projected pre-screening incidence compared with observed rates, with no adjustment for lead time effects. Furthermore, the time point for their comparison of observed and expected incidence rates corresponded with the period of peak HRT use in Australia[64], which may have contributed to higher than expected incidence of breast cancer. Potential changes in risk factor prevalence were not considered in their analyses. While the former study[77] did take account of trends in risk factor prevalence (with respect to BMI, parity and HRT use), their adjustment for lead time may not be adequate according to recent work by Duffy and Parmar, which indicates that shifting the expected incidence forward by a uniform 5 years may not be sufficient[117]. These authors have expressed their estimate as a proportion of expected cancers among 50-69 year olds. (Their estimate of 30% using the extrapolation method and 5 years lead time shift is equivalent to 23% of observed cancers in this age group).

While our results may be generalizable to other Australian jurisdictions, the extent to which they apply in other countries is less clear, due to the fact that screening programs vary considerably in relation to protocols used, target age range, frequency of screening, as do the risk factor profiles in various countries. However there are many programs that feature biennial screening of women aged 50-69 years with 2 views and double-reading, which began in the early 1990's, particularly across Europe, to which these findings may be applicable[233]. Also, our estimates may not apply to digital mammography, which is becoming more widespread throughout programs across the world, including BreastScreen Australia. Recent reports suggest higher

detection rates for specific subgroups of screening participants[234], however whether over-diagnosis would be larger in relation to digital technologies remains unclear.

**Conclusion**

Our case-control design is a novel approach to measuring over-diagnosis and may have advantages over other methods since it does not rely upon assumptions about the duration or distribution of lead time effects and uses individual measures of screening exposure. Furthermore it provides an estimate of over-diagnosis within a more contemporary service screening setting than the original trials and is not highly dependent on accurate estimates of background (non-screening) incidence rates.

We found only a modest level of over-diagnosis of invasive breast cancer due to population-based mammography screening of 8% among those eligible for screening, which decreased further with adjustment for potential differences in risk factors between attendees and non-attendees. This is considerably lower than previous estimates for Australia, and is useful to inform screening practice and also to inform women about this issue.

**Acknowledgements**

We wish to acknowledge the enormous contributions of Mr Graeme Tucker and Ms Minh Nguyen from the Epidemiology Branch of SA Health and Ms Gillian Rogers from BreastScreen SA in relation to the record linkage and data management tasks that were essential for undertaking this study.

Ms Beckmann is the recipient of a NHMRC PhD scholarship and Professor Lynch is supported through a NHMRC Australian Fellowship. A/Prof Houssami is supported through a National Breast Cancer Foundation Practitioner Fellowship. This work was funded by the National Breast Cancer Foundation's Novel Concept Award (#NC-11-53).

## APPENDIX (Chapter 6):

**Calculation of cumulative incidence in the presence and absence of mammography screening.**

The cumulative incidence of breast cancer in the absence of the screening program was derived from the projected incidence rates for SA and the ERP for 2006-2010, calculated for 10 year age groups and then summed. This can be described mathematically as:

$$CUM\ INC_{ns} = \sum_{age\ groups} [\ (IR_{ns} * N)]$$

where $IR_{ns}$ is the projected age-specific incidence rate in the population not exposed to screening and *N* is the person years in each age stratum for the period 2006-2010.

To calculate the cumulative incidence of breast cancer for the current population exposed to screening we used the following formula:

$$CUM\ INC_{scr} = \sum_{age\ groups} \left[ (IR_{ns} * Pr_{ns} * N) + \sum_{tw} (OR * IR_{ns} * Pr_{sc} *\ N) \right]$$

where $Pr_{sc}$ is the proportion of controls in each time window (tw), $Pr_{ns}$ is the proportion of controls who had never been screened and OR is the odds ratio for each time window. This calculation also incorporates the background incidence among non-participants by including the proportion never screened among controls.

This formula essentially describes the process of summing (within each age stratum) the number of cancers across all time windows. Subtotals were calculated for each age stratum separately, then summed to give the total cumulative incidence for the whole eligible population. The cumulative incidence subtotals for each age stratum, in the presence of the screening program, were calculated by summing the number of cancers expected across each of the time windows (yearly intervals since the most recent screening mammogram). To determine the number of cancers expected within each time window the background incidence rates were multiplied by the odds ratio and by the number of women at risk within each time window. The number at risk in each time window was based on the distribution of controls across the time windows

as a proportion of the total person years in the respective age group for 2006-2010, derived from ERP data. (The distribution among controls should be representative of the distribution in the wider population since controls were selected by random sampling.) Subtotals for each age stratum also included the number of cancers among the women who had never screened (derived from the background incidence rate multiplied by the proportion of controls who had never screened and by total person years in that age group).

The logic behind this approach was to use a case-control design to determine the extent to which incidence was altered at the time of screening and during the follow-up period by measuring the relative differences in incidence at specific time intervals since screening was last undertaken. It is anticipated that incidence will be considerably higher at the time of screening (year 1) relative to the background incidence rate (in the non-screened population). Immediately after screening, incidence should be decreased relative to background rates since many of the cancers would have presented clinically during this period. With increasing time since the last screening episode incidence rates should gradually return the background level. The extent of over-diagnosis can be determined from the net excess of breast cancers across the entire follow-up period, provided it is long enough to allow for lead time effects of screening to be fully compensated for by the deficit. Since breast cancer is a relatively rare event, odds ratios derived from the logistic regression will closely approximate the relative risk of breast cancer relative to women who have not participated in screening and can be used in relation to calculating the expected number of cancers across the time windows at and after screening. We calculated the cumulative incidence separately for age strata, to give better precision in the estimation of over-diagnosis. This allows for the anticipated differences in the distribution of women across the time windows within the different age groups and includes follow-up of women who elected not to rescreen before reaching the upper "target" age.

# Chapter 7: Estimates of over-diagnosis of breast cancer due to population-based mammography screening in South Australia after adjustment for lead time effects.

## PREFACE

The following chapter details the second of the studies undertaken to measure breast cancer over-diagnosis in South Australia. The lead time modelling approach used in this study was inspired by the work of Duffy and Parmer[117] showing the effects of lead time on population incidence rates over a 30 year period. Their study considered the ideal scenario of 100% participation in biennial mammography screening from ages 50-69 years in a static population (i.e. no change in incidence or population size). They used a fixed estimate of the average lead time (40 months) to determine the number of cancers with advanced diagnosis due to screening and applied this to a stable population profile. In the present study I have extended this approach by including 'real' data on screening participation and population denominators for the South Australian population. I also used age-specific estimates of lead time derived from using South Australian incidence and interval cancer rates. Using the lead time model I developed, background incidence rates (i.e. expected incidence without screening) were adjusted to reflect the effects of screening, based on screening participation and lead time estimates for South Australia. Over-diagnosis was then estimated by comparing the lead time adjusted cumulative incidence with observed incidence. This approach is similar to the approach used by Paci et al. in 2004[79] and 2006[78], although in their studies they adjusted the observed incidence to remove lead time effects and compared this with expected incidence in the absence of screening.

In the study outlined in this chapter, age-specific lead time estimates for South Australia were derived mathematically using background incidence rates (without screening) and interval cancer rates from BSSA data. Since the calculations were based on symptomatic cancers only, lead time estimates do not include 'overdiagnosed' cases, which has been one of the criticisms of lead time modelling approaches. As was

established in Chapter 3, HRT use substantially increases the risk of interval cancers. To accurately estimate screening lead times it was necessary to adjust for the effects of HRT on interval cancer rates, due to the high prevalence of HRT use among South Australian women over much of the lifetime of the screening program. These adjustments were based on empirical data from BSSA pertaining to the prevalence of HRT use and the incidence rate ratios for interval cancer rates (based on Poisson regression models as outlined in Chapter 3).

I gratefully acknowledge the assistance of Professor Duffy who provided the mathematical formulae for deriving lead time estimates using background incidence and interval cancer rates. With regard to all other aspects, including the development of a model incorporating screening attendance data along with dynamic population denominators and incidence rates, this chapter constitutes my own original work.

## AUTHORSHIP STATEMENT

**Estimates of over-diagnosis of breast cancer due to population-based mammography screening in South Australia after adjustment for lead time effects.**
[Submitted to *Journal of Medical Screening*]

**Kerri Beckmann (Candidate)**
I conceived and designed the study, managed data collation, analysed and interpreted the data, drafted and critically revised the manuscript, coordinated contributions from co-authors and handled revisions prior to print.

25/09/14
Signed …… Date……

**Stephen Duffy**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

24/09/14
Signed …… Date……

**John Lynch**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

23/09/14
Signed …… Date……

**Janet Hiller**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**Gelareh Farshid**
My contribution to this publication involved advising on methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

19/09/14
Signed …… Date……

**David Roder**
My contribution to this publication involved advising on research design and methodology, assisting with data interpretation and reviewing the manuscript. I give consent for Kerri Beckmann to present this paper for examination towards the Doctor of Philosophy.

22/09/14
Signed …… Date……

## ABSTRACT

**Objective:** This study aims to estimate over-diagnosis due to population-based mammography screening in South Australia (SA) using a lead time adjustment approach where lead time measures are based on symptomatic cancers only.

**Setting:** South Australian women aged 40-84 years in 2005-2009, who were eligible to participate in mammography screening.

**Methods:** Over-diagnosis was estimated by comparing the total number of observed breast cancer cases with the expected number, after adjustment for lead time. Lead time effects were modelled using age-specific estimates of lead time and screening sensitivity, projected background breast cancer incidence rates (in the absence of screening) and proportions undergoing screening, by age and calendar year. Age-specific lead time estimates were derived from interval cancer rates and predicted background incidence, using maximum likelihood methods.

**Results:** Age-specific estimates of lead time were 12, 26, 43 and 53 months, for women aged 40-49, 50-59, 60-69 and 70-79 years respectively. Background incidence rates were estimated to have increased by 0.9% and 1.2% per year for invasive and all breast cancer. The percentage of over-diagnosis among all women (aged 40-84 years) was estimated to be 7.9% (0.1-12.0%) for invasive cases and 12.0% (5.7-15.4%) when ductal carcinoma in-situ was included.

**Conclusions:** We estimated 8% over-diagnosis for invasive breast cancer and 12% inclusive of ductal carcinoma in-situ cancers due to mammography screening among women aged 40-84 years. These estimates may overstate the extent of over-diagnosis if the increasing prevalence of breast cancer risk factors has led to higher background incidence than predicted.

## INTRODUCTION

Concern has been raised about the potential for screening mammography to lead to over-diagnosis, i.e. the detection of cancers that would not have presented with clinical symptoms in a person's life time had they not undergone screening[56]. At present there is no consensus about the level of over-diagnosis due to mammography screening, with reported estimates ranging from 0 to 54% of expected numbers of breast cancers in the absence of screening.[89, 197, 198] Variation is likely to be due to inadequate adjustment for lead time and/or differences in risk between comparison populations, and use of different denominators in reporting estimates.

The aim of this study was to determine the level of over-diagnosis due to population-wide mammography screening in South Australia (SA), based on the excess cumulative incidence of breast cancer after adjusting for lead time effects. Lead time is the amount of time that the diagnosis is advanced by screening from the point when the cancer would have become clinically evident had screening not occurred. Since the aim of screening is to diagnose disease at an earlier stage when treatments are more effective in reducing mortality or morbidity, it is essential that there be adequate lead time to ensure such an advantage. Very long lead times however (as in the case of indolent tumours) can result in over-diagnosis.

To adjust for lead time we based our approach on the method described by Duffy and Parmar[117]. They used this method to demonstrate the effect of lead time in the 'ideal' scenario of 100% participation in biennial mammography screening starting at age 50 and stopping at age 69, assuming constant incidence rates and population profiles over a 30 year period and an average lead time of 40 months for all women irrespective of age. Our study expands on this approach by including the observed proportion of women who underwent screening each year in South Australia, in addition to changes in background incidence rates and population profiles over time. Furthermore, we applied age-specific lead time estimates which were estimated from interval cancer rates and projected background incidence for South Australia and hence reflect lead times for clinically relevant breast cancers.

## METHODS

The population-based mammography screening program (BreastScreen SA) was piloted in South Australia in 1989 and rolled out across the state from 1991. It offers free biennial mammographic screening to asymptomatic women aged 40 years and over, with written invitations targeting women aged 50-69 years. Prior to 2010, the program used two-view, film-based mammography, read by two independent radiologists, with a third reading in discordant cases.

Incidence data for all breast cancers from 1977-2009, among SA women aged 40-84 years, were obtained from the South Australian Cancer Registry. Incidence rates were calculated based on the Estimated Residential Population (ERP) for SA at June 30 of each year by individual year of age, sourced from the Australian Bureau of Statistics (ABS)[105]. These rates represent the 'observed' incidence prior to and following the introduction of mammography screening in SA. The numbers who participated in population-based screening, by year of age and calendar year from 1989 to 2009, were obtained from BreastScreen SA (BSSA).

Expected incidence, by year and age, had population-based screening not been implemented in SA, was determined through projection of breast cancer incidence rates during the period 1977-1988, using age-period Poisson regression with individual year of age and calendar year as continuous variables. There was no evidence of overdispersion in these models. Projections were extended to 2019. For the period 1989-2011, we used the ERP as the denominator for predicting the number of cases, while, for the period 2012-2019, we used the projected population estimates provided by the ABS (series C)[235].

**Estimation of lead time and sensitivity**

We estimated age-specific lead times and screening sensitivity for 10 year age groups (40-49, 50-59, 60-69 and 70-79 years) by calculating the expected interval cancer rates in the first and second year following screening as described by Day[236], according to the following formula:

$$I(t) = I(1-S)t + IS\left\{t + \frac{e^{-\lambda t} - 1}{\lambda}\right\}$$

where *I* is the incidence of cancer in the absence of screening, *S* is the screening sensitivity, t is the time after a screen and 1/λ is the average duration of the preclinical screen-detectable period ( i.e. lead time), usually referred to as the mean sojourn time (MST). The expected rate of interval cancers in the first year after screening is $p_1=I(1)$ and the expected rate in the second year is given as $p_2=I(2)-I(1)$. . The log-likelihood is:

$$lnL = i_1 \ln(p_1 N) - p_1 N + i_2 \ln(p_2 N) - p_2 N$$

assuming the incidence of interval cancers has a Poisson distribution.($i_1$ and $i_2$ are the observed interval cancers in the first and second years respectively.) We maximised the expected number of interval cancers in the first and second year after screening, using observed interval cancers rates for 1990-2009 adjusted for the effect of hormone replacement therapy (HRT) use at screening, over a grid of values of λ and S, as in Day and Walter[237], to obtain maximum likelihood estimates of these parameters. Values for λ ranged from 0.1 to 1.2, which is equivalent to lead times of 10 months to 10 years, while values for sensitivity (*S*) ranged from 0.51 to 0.99. We obtained 95% confidence intervals on each of these parameters using the profile likelihood method[238].

Adjustment for HRT effects was necessary due to the high prevalence of HRT use among SA women over much of the period and the strong association between HRT use and higher interval cancer rates[239], which would have inflated the number interval cancer and distorted lead time estimates. Adjustment was based on empirical data from BreastScreen SA on the prevalence of HRT use at screening, and incidence rate ratios for interval cancers in the first and second years after screening for HRT users compared with non-users. Further details are provided in the Appendix (online supplement).

**Lead time adjustment**

For any given year in which screening takes place, calculating the expected incidence adjusted for lead time requires

1. projecting the pre-screening incidence trends to estimate the expected underlying incidence if screening had not taken place (as described above);

2. calculating the expected cancers brought forward from future years as a result of screening lead time, and adding these to the expected incidence; and

3. calculating the expected cancers detected in previous years due to lead time from screening in those previous years, and subtracting these from the expected incidence.

Based on the exponential distribution of lead time, (2) and (3) can be calculated as shown by Duffy and Parmar[9]. Using this method, we calculated the number of cancers expected after accounting for lead time effects of screening, in each calendar year and year of age for the period since screening was implemented in 1989 to 2009 for women aged 40-84 years. 'Lead time adjusted rates' were determined by dividing the number of expected cases by the estimated population for each year of age and calendar year.

To estimate the extent of over-diagnosis we calculated the excess difference between the observed cumulative incidence (from cancer registry data) and the expected cumulative incidence after lead time adjustment for all women 40-84 years during the period 2005-2009. We only counted cancers whose diagnosis was advanced by ten years or less, that is, we assumed no lead times in excess of ten years. Confidence limits were determined by applying the values for the upper and lower 95% confidence intervals for lead time estimates in modelling lead time effects.

Finally, we conducted several sensitivity analyses to examine the impact of varying the background incidence rates. These analyses included modelling lead time effects assuming that the rate of increase in background incidence trends was 10% and 20% higher or lower than predicted. Direct modelling of HRT effects on background incidence rates was not possible due to the lack of fine grained data on HRT use in the entire population spanning the pre- and post-implementation period. Also, to exclude any 'artificial' elevation in incidence rates due to increased opportunistic screening in the later period before the public breast screening program commenced, we modelled incidence rates for the period 1977-1985, rather than 1977-1988. We also present estimates of excess breast cancer incidence over a period of two decades of screening (1990-2009).

All analyses were undertaken for invasive breast cancer (IBC) and ductal carcinoma in-situ (DCIS) combined, and for IBC alone.

## RESULTS

Person years of screening, numbers of interval cancers and background incidence rates used to derive estimates of age-specific lead time and screening sensitivity for SA are shown in Table 7.1, along with the final estimates of lead time and sensitivity. Based on data for 1990-2009 from BSSA, the mean age-specific lead times for IBC from mammography screening were 12 months, 26 months, 43 months and 52 months, respectively, for women aged 40-49, 50-59, 60-69 and 70-79 years. Lead time estimates for all BC were slightly shorter. Estimates of screening sensitivity ranged from 56% for women in their 40's, to 97% for women in their 70's.

The proportions of cancers which would have their diagnosis advanced by 1-10 years due to lead time effects, after accounting for screening sensitivity, are shown in Table 7.2. These were derived from the age-specific lead time and sensitivity estimates from BSSA screening data. As evident from the table, longer lead times resulted in higher proportions of cancers with advanced diagnosis, over a longer period.

Age-period Poisson modelling of incidence during the pre-screening era (1977-1989) for all women aged 40-84 years indicated an annual increase in incidence of 0.9% (0.1-1.6%) for invasive breast cancer and 1.2% (0.0-2.4%) for invasive and DCIS combined. Incidence rates increased by 2.4% (2.2-2.6%) for IBC and 2.3% (2.1-2.5%) for all BC, for each increasing year of age. Projections from these models using ABS population data for the denominators provided the background incidence rates to be adjusted for lead time effects.

**Table 7.1 Estimates of lead time and screening sensitivity derived from person years of screening, number of interval cancers and background (projected) incidence rates in the absence of screening and South Australia 1990-2009**

| Age group | No. screening episodes[a] | Background incidence rate/100,000 | No. Interval cancers | | Lead time | | Screening sensitivity | |
|---|---|---|---|---|---|---|---|---|
| | | | 0-12mth | 13-24mth | months | 95% CI | % | 95% CI |
| **IBC** | | | | | | | | |
| 40-49 | 167,541 | 132 | 143 | 192 | 12.0 | 10.1-23.1 | 56 | 51 - 77 |
| 50-59 | 479,626 | 169 | 253 | 458 | 26.1 | 20.0-35.3 | 86 | 78 - 94 |
| 60-69 | 360,407 | 216 | 168 | 317 | 42.9 | 32.4-60.0 | 90 | 84 - 96 |
| 70-79 | 74,190 | 277 | 27 | 63 | 52.2 | 40.0-109 | 97 | 87 - 98 |
| **ALL BC** | | | | | | | | |
| 40-49 | 167,541 | 142 | 149 | 206 | 11.7 | 10.1-23.5 | 60 | 51 - 80 |
| 50-59 | 479,626 | 181 | 263 | 495 | 24.5 | 19.4-32.4 | 88 | 80 - 96 |
| 60-69 | 360,407 | 230 | 182 | 353 | 38.7 | 30.0-52.2 | 91 | 85 - 97 |
| 70-79 | 74,190 | 293 | 31 | 73 | 48.0 | 36.4-85.7 | 97 | 87 - 98 |

[a] Screening data for South Australia derived from BSSA screening records

[b] Lead time (mean sojourn time) = $1/\lambda_1 *12$ where $\lambda_1$ was determined using the method described by Day[236]

CI: confidence interval; IBC: invasive breast cancer; BC: breast cancer

**Table 7.2 Proportion of cancers expected to have diagnosis advanced by screening for each subsequent year after mammography screening, based on lead time estimates and sensitivity of screening (for IBC) derived from BreastScreen SA data**

| Age group (yrs) | 40-49 | 50-59 | 60-69 | 70-79 |
|---|---|---|---|---|
| Lead time (months) | 12.0 | 26.1 | 42.9 | 52.2 |
| Screening sensitivity | 0.56 | 0.86 | 0.9 | 0.97 |
| IBC rate (per100000) | 132 | 169 | 216 | 277 |
| **Proportion of cancers advanced by screening#** | | | | |
| Years after screening | | | | |
| 1 | 0.35 | 0.69 | 0.79 | 0.87 |
| 2 | 0.13 | 0.44 | 0.59 | 0.69 |
| 3 | 0.05 | 0.28 | 0.45 | 0.55 |
| 4 | 0.02 | 0.18 | 0.34 | 0.44 |
| 5 | 0.01 | 0.11 | 0.26 | 0.35 |
| 6 | 0.01 | 0.07 | 0.20 | 0.28 |
| 7 | 0.01 | 0.05 | 0.15 | 0.23 |
| 8 | 0.01 | 0.04 | 0.12 | 0.18 |
| 9 | 0.00 | 0.03 | 0.09 | 0.15 |
| 10 | 0.00 | 0.02 | 0.08 | 0.12 |

IBC = Invasive breast cancer

# after accounting for screening sensitivity

Proportions of cancers with advanced diagnosis through screening were calculated as described by Duffy and Parmar (2013) [117] based on average annual incidence of IBC/100,000 in absence of screening (1990-2009) and age-specific estimates of lead time and screening sensitivity. See Appendix for detailed methodology.

Figure 7.1 presents time trends in lead time adjusted rates compared with actual observed incidence rates and projected (background) incidence rates for invasive breast cancer in SA from 1977 to 2009. These trends indicate that the incidence rate pattern after adjusting for lead time effects resembles the 'true' observed incidence pattern, although the adjusted rate is consistently lower than the observed rate.

Differences in the total number of cancers expected after lead time adjustment and the actual observed cumulative incidence, as a percentage of observed breast cancers, for the period 2005-2009 are shown in Table 7.3. Excess cumulative incidence would be equivalent to the level of over-diagnosis due to organised mammography screening, provided trends in incidence rates have not changed substantially since the pre-screening era, and our estimates of lead time apply to progressive cancers only. Based on these assumptions, over-diagnosis constitutes 7.9 % (confidence limits: 0.1 - 12.1%) of all invasive breast cancer among 40-84 year old women and 12.0% (5.7 - 15.4%) of all breast cancers when DCIS is included.

Results of various sensitivity analyses relating to estimates of excess incidence for all BC and for IBC alone are shown in Table 7.4. If background incidence rates were based on projected rates for the period 1977-1985, rather than 1977-1988, our estimates of over-diagnosis were lower (6.4% and 9.7%, for IBC and all BC respectively). A 10% reduction in the rate of increase in background incidence over time resulted in increased estimates of over-diagnosis (10.8 % and 15.2% respectively), while a 10% increase resulted in lower estimates of over-diagnosis (5.3% and 8.8%, respectively). Estimates calculated for over the 2 decades of screening are slightly higher (8.7% and 13%) than estimates for the period 2005-2009, which may be due to the high prevalence of HRT over the first 14 years of the program.

**Figure 7.1 Trends in observed, lead time adjusted and projected (background) invasive breast cancer incidence rates for South Australian women aged 40-84yrs, 1977-2009**

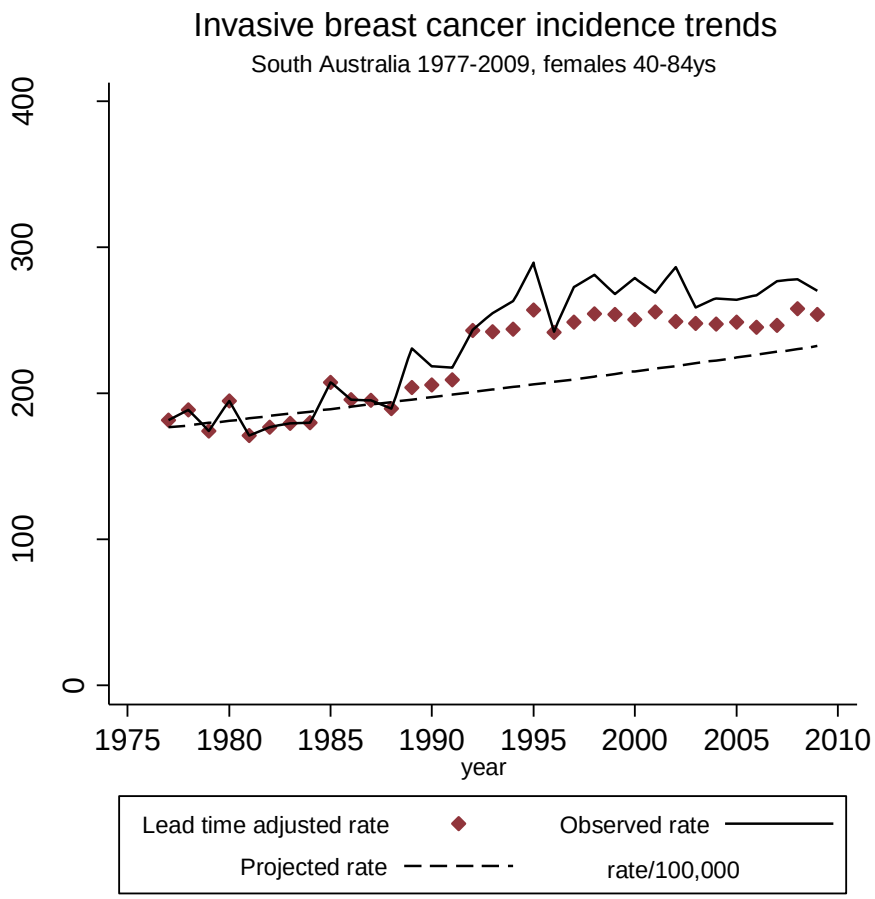


**Table 7.3. Observed, lead time adjusted and percent excess breast cancer incidence for South Australia (2005-2009)**

| Age group | IBC | | | All BC# | | |
|---|---|---|---|---|---|---|
| | Observed | Expected (lead time adjusted[a]) | Excess | Observed | Expected (lead time adjusted[a]) | Excess |
| | No. | No. | % (CL) | No. | No. | % (CL) |
| 40-49yrs | 887 | 856 | 3.5 (1.4, 3.8) | 984 | 935 | 5.0 (2.4, 5.3) |
| 50-59yrs | 1293 | 1174 | 9.2 (-0.2, 14.7) | 1509 | 1264 | 16.2 (8.5, 20.6) |
| 60-69yrs | 1413 | 1269 | 10.2 (-5.2, 19.7) | 1597 | 1327 | 16.9 (4.9, 24.6) |
| 70-84yrs | 1203 | 1117 | 7.1 (5.2, 6.6) | 1288 | 1203 | 6.6 (5.7, 5.8) |
| All ages (40-84yrs) | 4796 | 4418 | **7.9 (0.1, 12.1)** | 5378 | 4731 | **12.0 (5.7, 15.7)** |

BC: breast cancer; IBC: invasive breast cancer

CL: Confidence Limits derived using lead time estimates corresponding to the upper and lower 95% confidence intervals.

[a] lead time adjustment modelling used age-specific lead time estimates based on interval rates and background incidence for all BC (which are similar to but not identical to those for IBC).

Excess incidence represents the % over-diagnosis in the population due to mammography, under the assumption of no change in background incidence rates based on projected pre-screening incidence trends.

% Excess cumulative incidence among all observed breast cancers within the respective age range

= (observed cumulative incidence–lead time adjusted cumulative incidence)/observed cumulative incidence*100

**Table 7.4. Sensitivity analyses for estimates of excess cumulative incidence (over-diagnosis), in SA women aged 40-84yrs .**

| Sensitivity analyses | | | Excess Cumulative Incidence | | | |
|---|---|---|---|---|---|---|
| **Pre-screen period** | **Background rate adjustment** | **Estimate period** | **IBC** | | **All BC** | |
| | | | % | CL | % | CL |
| Base model | - | 2005-2009 | 7.9 | (0.1, 12.1) | 12.0 | (5.7, 15.7) |
| 1977-1985 | - | 2005-2009 | 6.4 | (-1.5, 10.7) | 9.7 | (3.1, 13.2) |
| 1977-1988 | Decreased 10% | 2005-2009 | 10.8 | (2.8, 14.5) | 15.2 | (9.1, 18.4) |
| 1977-1988 | Decreased 20% | 2005-2009 | 12.9 | (5.7, 16.8) | 18.2 | (12.4-21.3) |
| 1977-1988 | Increased 10% | 2005-2009 | 5.3 | (-2.8, 9.6) | 8.8 | (2.1, 12.3) |
| 1977-1988 | Increased 20% | 2005-2009 | 2.6 | (-5.6, 7.1) | 5.4 | (-1.6, 9.1) |
| 1977-1988 | - | 1990-2009 | 8.7 | (0.9, 12.9) | 13.8 | (7.4, 17.3) |

BC: breast cancer; IBC: invasive breast cancer

CL: Confidence Limits derived using lead time estimates corresponding to the upper and lower 95% confidence intervals.

% excess cumulative incidence among all observed breast cancers with in the respective age range

= (observed cumulative incidence–lead time adjusted cumulative incidence/ observed cumulative incidence*100)

## DISCUSSION

Our findings indicate a modest level of over-diagnosis due to population-based mammography screening in South Australia. Trends in lead time adjusted incidence rates closely resemble the observed pattern for SA, although adjusted rates were consistently lower than observed rates. Based on the excess number of breast cancers after lead time adjustment, over-diagnosis due to population-based mammography screening in SA was estimated to be 8% for IBC, and 12% for all breast cancers (IBC + DCIS) for women aged 40-84 years during 2005-2009. These estimates are consistent with findings from a previous study undertaken in SA by our research group (which found 8% and 14% over-diagnosis for IBC and all BC respectively)[240].

Our estimates of over-diagnosis are consistent with those from randomised control trials (RCT) of mammography screening. Moss[70] reported 11-14% over-diagnosis for all BC based on follow-up for trials which did not offer screening to the control arm at the end of the trial period. Similarly, Zackrisson et al.[71] reported 7% over-diagnosis for IBC alone and 10% for all BC based on data from the Malmo trial. Our findings also concur with data presented in the 2012 independent review of breast screening in the UK[39]. There is also good agreement between our findings and several other observational studies including those by Falk et al.[83] (11-13% for IBC in Norway); Waller et al.[68] (~11% for IBC in the UK); and Puliti et al.[85] (10% for all BC in Northern Italy). However, several studies have reported lower levels of over-diagnosis ranging from 0-5% for IBC[69, 78, 79, 81, 84, 223], while others have reported much higher rates of over-diagnosis ranging from 22% to 54%[54, 77, 82, 219, 221]. Choice of denominator explains the higher levels of over-diagnosis reported in some of the latter studies[3].

Our estimates are slightly higher than other studies that have used a similar lead time modelling approach, eg. Paci et al.[78, 79] (5% of all BC in Northern/Central Italy). Differences may reflect variation in the screening programs themselves, or be due to other factors that influence background incidence such as differences in HRT use. Also, our assumption of no lead time beyond 10 years may overestimate over-diagnosis, particularly for older ages where the duration of the preclinical screen-detectable period is relatively long. By contrast, our results are considerably lower than those reported for studies which adjusted for lead time by shifting incidence rates by the

estimated lead time. This was the approach used in an Australian study[77] which reported levels of over-diagnosis of between 30-54% of the expected rate for New South Wales, Australia (equating to 23-35% of observed rates). The results of Duffy and Parmar[9] suggest that the incidence shift method may lead to overestimation of over-diagnosis. This is because the lead time does not simply exchange future incidence for current incidence: it adds a proportion of future incidence to current incidence.

We have made a number of assumptions in estimating over-diagnosis through modelling lead time effects. The first is that there were no major changes in underlying incidence trends. Due to a lack of baseline data during the pre-screening period we were unable to include breast cancer risk factors in our prediction modelling. With the exception of HRT use, prevalence of several known breast cancer risk factors (e.g. body weight, alcohol consumption, parity, age at first birth, diabetes rates) among SA women has steadily increased over the two decades since screening was implemented[241, 242], which may have served to further increase incidence rates. The impact of changing patterns of HRT use on background incidence is more difficult to assess, but the high prevalence of use for much of the time since screening was introduced is likely to have led to higher than predicted background incidence[138]. Use of private mammography screening[243] and increasing diagnostic vigilance[118] may also have inflated background incidence rates over the past 2 decades. As indicated by sensitivity analyses, adjusting the background trend rate has a moderate impact on our estimates. Even with a twenty percent lower rate of increase over time, excess incidence (i.e. over-diagnosis) was just under 13% for IBC. However, the most likely scenario is higher rather than lower background incidence rates, given the cumulative effect of increasing prevalence across multiple breast cancer risk factors[241], in which case over-diagnosis would be lower than our estimates of 8%.

Our model also assumes that background incidence trends before the implementation of organised mammography screening were not greatly affected by any uptake of opportunistic screening during that period. Data from New South Wales, Australia, indicate a marked increase in Medicare-funded diagnostic mammography between 1985 and 1992[243]. The authors believe that some diagnostic mammography during this

period constituted de facto or opportunistic screening. Any parallel increase in opportunistic screening in SA may have inflated incidence trends. However, sensitivity analyses, which restricted the pre-screening era to the period 1977-1985 to remove any effect of opportunistic screening, actually yielded slightly lower estimates of over-diagnosis in our study.

Finally, we have assumed an exponential distribution for lead time, which applies only to cancers that are truly progressive. The assumption that lead time is exponentially distributed is based both on the biological models of cell growth for tumours in the screen-detectable size range[202] and empirical evidence which indicates that an exponential distribution gives the best fit when modelling lead time effects[237]. Zahl et al.[200, 206] have argued lead time adjustment methods underestimate over-diagnosis because lead time estimates include slowly growing or dormant tumours that would never have arisen clinically during a person's lifetime (i.e. over-diagnosed cases). However, the method used in this study to derive age-specific lead times was based on projected background incidence rates in the absence of screening and on interval cancer rates, which for the most part, are truly progressive tumours.

The strengths of this study include the application of age-specific lead times derived directly from the study population. Importantly our lead time estimates were calculated from parameters for symptomatic cancers only so do not include any contribution from non-progressive cancers detected at screening. Our model also includes adjustment for screening sensitivity for different age groups. Furthermore, we have used fine grained data on screening participation to model lead time effects in this study.

## CONCLUSION

Our findings suggest a relatively low level of over-diagnosis due to organised population-based mammography screening in South Australia. Assuming a constant increase in background incidence rates based on incidence trends prior to the implementation of mammography screening, much of observed increase in breast cancer can be explained by lead time effects. Our estimate of 8% for IBC is consistent with findings from RCTs and the recent Independent Review of Breast Screening in the UK. and in agreement with results from our previous case-control study[240]. Our

estimate may overstate the extent of over-diagnosis if projections have not accounted adequately for effects of increases in breast cancer risk factors in the decades following commencement of the screening program.

## ACKNOWLEDGEMENTS

Special thanks go to Mr Graeme Tucker and Ms Minh Nguyen from the Epidemiology Branch of SA Health for their painstaking work in relation to the record linkage and assembly of relevant datasets, and to Ms Gillian Rogers from BreastScreen SA data for her assistance with data extraction and compilation, without which this study would not have been possible. Ms Beckmann is the recipient of a NHMRC PhD scholarship and Professor Lynch is supported through a NHMRC Australian Fellowship. This work was funded by the National Breast Cancer Foundation's Novel Concept Award (#NC-11-53).

The authors declare they have no competing interests.

## APPENDIX

**Estimating lead times**

Day[236] showed that in an interval of time $t$ following a screen, the expected cumulative incidence of interval cancers, which were not in the preclinical screen-detectable period at the time of the screen and so could not have been detected then, is

$$I_1(t) = I \int_0^t F(t)dt$$

where $I$ is the incidence rate of progressive, that is non-overdiagnosed, cancers and F($t$) is the distribution function of the preclinical screen-detectable period. The corresponding cumulative incidence of interval cancers which were in the preclinical screen-detectable period at the time of the screen but which were missed at the screen is

$$I_1(t) = I(1-S) \int_0^t (1 - F(t))dt$$

where $S$ is the screening sensitivity.

The overall incidence of interval cancers ($I$) in time $t$ after a screen will therefore be

$$I(t) = I(1-S) + IS \int_0^t F(t)dt$$

The exponential distribution has been found to be a good fit for the preclinical screen-detectable period[236], so we can express F($t$) as

$$F(t) = 1 - e^{-\lambda t}$$

Where 1/λ is the average duration of the preclinical screen-detectable period, usually referred to as the mean sojourn time (MST). This gives

$$I(t) = I(1-S)t + IS\left\{t + \frac{e^{-\lambda t} - 1}{\lambda}\right\}$$

Since these pertain to symptomatic interval cancers, the parameters, $I$, λ and $S$ apply to progressive, non-overdiagnosed cancers.

For purposes of estimation, we had interval cancer incidence for the first and second year after screening ($i_1$ and $i_2$ respectively), stratified by age and period of diagnosis. The expected rate of interval cancers in the first year after screening is p1=$I$(1) and the expected rate in the second year is p2=$I$(2)-$I$(1). Assuming the incidence of interval cancers has a Poisson distribution, the log-likelihood is

$$lnL = i_1 \ln(I(1)N) - I(1)N + i_2 \ln(I(2)N) - I(2)N$$

We maximised this with over a grid of values of λ and $S$ as in Day and Walter[237], to obtain maximum likelihood estimates of these parameters. We obtained 95% confidence intervals on the parameters using the profile likelihood method[238].

**Calculating interval cancer incidence, taking account of HRT effects**

An earlier study undertaken among BreastScreen SA participants[239] showed a two-fold increase in interval cancer rates among women using HRT at the time of screening, for all age groups except women aged 40-49 years. The high prevalence of HRT use in SA would distort estimates of mammography screening lead time due to the increase in the number of interval cancers and therefore needed to be accounted for when estimated lead time length. We used BreastScreen SA data on the prevalence of HRT use and IRR for interval cancer rates among HRT user in the screening population to determine the number of interval cancers that were likely to be due to HRT use and subsequently adjusted the observed number of interval caners to remove the effect of HRT use. The HRT adjusted number of intervals cancers (for each year and each age group) were calculated as follows:

Adjusted no. interval ca. = Observed no. interval ca. / [(1-Pr) + (Pr * IRR)]

where Pr is the proportion of women using HRT at the time of screening and IRR is the incidence rate ratio for interval cancers among HRT users. The IRR were derived using Poisson regression modelling in the BreastScreen SA cohort for the period 1992-2009 in each age strata. (Modelling is described in detail in Chapter 3). The prevalence of HRT use was also determined directly from BreastScreen records for each of the age strata over the same period. Adjustments were made for interval rates in the first and

second year after screening in each age strata. However, no adjustment was made in relation to women aged 40-49 years since IRRs were non-significant. The relevant incidence rate ratios and prevalence rates are shown in Table 7.5.

The background incidence rates for calculating lead times were based on projections of incidence trends prior to the start of screening in SA (ie 1977-1988). Even though HRT use was also associated with a moderate increase in overall BC incidence no adjustment for HRT was made. This assumes that the impact of HRT use was minimal during the pre-screening period. This assumption is reasonable given that combined oestrogen-progestin formulas which are associated with the highest risk, only became popular in SA during the mid- to late-1980s (anecdotal evidence) and effects tend to be associated with longer term use (5+ years). All calculations are based on data for the whole period of observation 1992-2009, so will represent the average effect over the period 1992-2009.

The data used and final adjusted number of interval cancers is shown in Table 7.6.

**Table 7.5 Impact of current HRT use on screen-detection and interval cancer rates (1992-2009)**

| Age group (yrs) | Prevalence@ Screening (%) | IRR (current HRT use), 95% CI | | | |
|---|---|---|---|---|---|
| | | Screen-detected | Interval ca. 0-12m | Interval ca. 13-24m | All interval ca. 0-24mths |
| **INVASIVE** | | | | | |
| 40-49 | 17.83 | 0.98 (0.75-1.27) | 0.88 (0.55-1.40) | 0.98 (0.67-1.43) | 0.96 (0.79-1.17) |
| 50-59 | 37.05 | 1.14 (1.03-1.25) | 1.90 (1.52-2.37) | 1.71 (1.44-2.02) | 1.32 (1.22-1.42) |
| 60-69 | 28.84 | 1.32 (1.21-1.45 | 2.00 (1.53-2.62) | 2.15 (1.77-2.60) | 1.48 (1.37-1.60) |
| 70-79 | 19.75 | 1.20 (0.99-1.44) | 2.98 (1.57-5.65) | 2.00 (1.25-3.20) | 1.34 (1.13-1.58) |
| **ALL BC** | | | | | |
| 40-49 | 17.83 | 0.93 (0.74-1.17) | 0.84 (0.52-1.32) | 1.00 (0.70-1.44) | 0.93 (0.78-1.11) |
| 50-59 | 37.05 | 1.11 (1.02-1.21) | 1.92 (1.54-2.38) | 1.67 (1.42-1.97) | 1.27 (1.19-1.37) |
| 60-69 | 28.84 | 1.26 (1.16-1.37) | 2.04 (1.58-2.63) | 2.01 (1.67-2.42) | 1.41 (1.31-1.51) |
| 70-79 | 19.75 | 1.15 (0.97-1.37) | 2.72 (1.47-5.02) | 1.78 (1.13-2.81) | 1.27 (1.09-1.48) |

HRT Hormone replacement therapy, CI confidence interval

IRR incidence rate ratio: derived from Poisson regression modelling with BSSA data with age (year) and diagnosis year as covariates

**Table 7.6 Number of interval cancers (SA 1992-2009), adjusting for HRT effects**

| Age Group (yrs) | Projected incidence /100,000 | Prevalence of HRT (%) | Interval cancers 0-12mths | | | Interval cancers 13-24mths | | |
|---|---|---|---|---|---|---|---|---|
| | | | IRR | Observed No. | Adjusted No. | IRR | Observed No. | Adjusted No. |
| IBC | | | | | | | | |
| 40-49 | 132 | 17.83 | 1.00 | 143 | 143 | 1.00 | 192 | 192 |
| 50-59 | 169 | 37.05 | 1.90 | 337 | 253 | 1.71 | 579 | 458 |
| 60-69 | 216 | 28.84 | 2.00 | 217 | 168 | 2.15 | 422 | 317 |
| 70-79 | 277 | 19.75 | 2.98 | 38 | 27 | 2.00 | 75 | 63 |
| ALL BC | | | | | | | | |
| 40-49 | 142 | 17.83 | 1.00 | 149 | 149 | 1.00 | 206 | 206 |
| 50-59 | 181 | 37.05 | 1.92 | 353 | 263 | 1.67 | 618 | 495 |
| 60-69 | 230 | 28.84 | 2.04 | 237 | 182 | 2.01 | 456 | 353 |
| 70-79 | 293 | 19.75 | 2.72 | 42 | 31 | 1.78 | 84 | 73 |

IBC invasive breast cancer; HRT hormone replacement therapy; IRR incidence rate ratio

**Calculating the proportions of cancers with advanced diagnosis due to lead time, by year since screen**

Using the formula from Duffy and Parmar[117], the probability of symptomatic cancer occurring within '$t$' years for screening is

$$\mathrm{Pr}_t = 1 - e^{-t\lambda_0} - \frac{\lambda_0 \left(e^{-t\lambda_0} - e^{-t\lambda_1}\right)}{(\lambda_1 - \lambda_0)}$$

where $\lambda_0$ is the rate of transition from no detectable disease to pre-symptomatic screen-detectable disease and $\lambda_1$ is the rate of transition from screen-detectable to symptomatic disease. The rate of transition from asymptomatic to preclinical disease ($\lambda_0$) is assumed to be equal to the rate for breast cancer in the absence of screening, while the rate of transition to symptomatic disease ($\lambda_1$) corresponds to 1/MST (the mean sojourn time or lead time in years). The probability of symptomatic disease in the second year after screening is the difference between $Pr_2$ and $Pr_1$, and in the third year is $Pr_3$-$Pr_2$, etc.

Example: Given a lead time of 26.1 months (2.2 years) and background breast cancer incidence rate of 169 per 100000 per year among women aged 50-59 years, $\lambda_0$ = 0.46 and $\lambda_1$=0.00169. Hence, the probability of symptomatic disease in the first year after screening is 0.000335 or 19.8% of the expected incidence (0.198=0.000335/0.00169). This indicates that 80.2% of cancers expected to occur during the first year after screening would have had been detected at screening. Likewise the probability of symptomatic disease occurring in the two years after screening=0.001168. The probability of cancer occurring symptomatically in the second year after screening is therefore 0.000833 (0.000833 =0.001168-0.000335) or 49% of the expected incidence. The proportion of cancers expected in the second year after screening that would have had their diagnosis advanced by screening would be 51%. After adjustment for screening sensitivity ($S$=0.86) the proportion of cancers with advanced diagnosis from the first year and second year after screening would be 0.69 and 0.44, respectively, for women aged 50-59 years.

**Lead time adjustment**

The number of cancers brought forward by screening from each future year was derived as follows:

$$\sum_{\mathrm{i}}^{0-10} \mathrm{N\_sc}\,(\mathrm{year0}, \mathrm{age0}) * \mathrm{IR}(\mathrm{year} + \mathrm{i}, \mathrm{age} + \mathrm{i}) * \mathrm{Pr\_lt}(\mathrm{year} + \mathrm{i}, \mathrm{age} + \mathrm{i}) * S$$

where N_sc is the number of women attending screening in the index year and age, IR is the incidence rate in the relevant future years (e.g. year+1/age+1; year+2, age+2; etc), Pr_lt is the proportion expected to be detected due to lead time in each future year, and $S$ is the age-specific screening sensitivity. This total was added to original base-line number of cases in the index year/year of age.

The number of cancers to be subtracted from the detectable pool in future years/ages was derived as follows:

$$\sum_{\mathrm{i}}^{0-10} \mathrm{Pr\_sc}\,(\mathrm{year} - \mathrm{i}, \mathrm{age} - \mathrm{i}) * \mathrm{Nca}(\mathrm{year0}, \mathrm{age0}) * \mathrm{Pr\_lt}(\mathrm{year} - \mathrm{i}, \mathrm{age} - \mathrm{i}) * \mathrm{S}$$

where Pr_sc is the proportion of women screening in the index year, Nca is the number of cancers expected in future years/ages (adjusted all previous screening) and Pr_lt is the proportion expected to be detected by screening in the respective years/age groups. This process was repeated year by year, from the initial year when screening began as a pilot program in 1989 through to 2009.

# Chapter 8: Conclusion

The central question addressed in this thesis is 'to what extent has mammography screening led to over-diagnosis of breast cancer?' (i.e. the detection of BC that would never have emerged clinically in a woman's lifetime had she not participated in screening[57]). This question is examined in the context of the population-based mammography screening program in South Australia, BreastScreen SA.

This body of work first investigated whether screening had led to an increase in breast cancer incidence and whether the excess incidence was greater than expected based on trends prior to the introduction of screening. Next it explored the potential for the excess to be explained (in part) by confounding factors (e.g. temporal changes or differences in the prevalence of risk factors). Particular attention was paid to establishing a causal link between HRT use and breast cancer risk in the SA population, which was identified as a potential confounder. Finally, following a review of study designs to measure the extent of over-diagnosis due to mammography screening, two different approaches were presented: a case-control design using individual level data and a lead time modelling approach using population level data.

## KEY FINDINGS

- **The advent of population-based mammography screening is likely to be the key factor contributing to the observed increase in recorded breast cancer incidence over the past 2 decades.**

Results from the descriptive analysis of incidence patterns, presented in Chapter 2, clearly indicate that breast cancer incidence rates rose sharply after the implementation of population-based mammography screening in South Australia, which was indicative of screening being a major influencing factor. The timing of this increase coincided with the start of the BreastScreen SA program. Stabilisation of incidence rates from the late 1990's is consistent with expected incidence patterns as a screening program matures to the point where the proportion and age profile of women attending for their first screening mammogram remains constant.

- **Breast cancer incidence has remained higher than expected based on projections of trends during the pre-screening era**

Based on comparison of projected and observed rates, breast cancer incidence has remained higher than expected some twenty years after the introduction of screening. Possible explanations include increases in the underlying risk of breast cancer due to increasing obesity, decreasing fertility rates, delayed childbirth. This may have resulted in a steeper rate of increase than during the pre-screening era (on which projections were based); ongoing lead time effects within a dynamic population (through continual recruitment of first time screening participants, ongoing participation after 69 years of age); increased medico-legal vigilance with regard to diagnosis; and over-diagnosis of cancers that would not have progressed to become clinically evident.

With the exception of HRT use, the prevalence of the key breast cancer risk factors examined in Chapter 2 has continued to increase steadily over the past two decades. This is likely to have had an ongoing influence on increasing 'background' breast cancer incidence rates, and may have resulted in increases beyond projected rates over the period.

There was no evidence of the anticipated decrease in incidence in the 'post-screening' age group. This may also be due to other influences that have led to increases in underlying incidence which have masked the expected decline. Though difficult to gauge from the available data on fertility trends, changes in fertility rates may have led to higher than expected rates in older women. Higher incidence in this age group may also be due in part to the increasing proportion of women electing to continue to participate beyond the upper target screening age (~15%).

- **Changing prevalence of HRT use is also likely to have had a major influence on breast cancer trends. The impact of HRT use needs to be considered in relation to quantifying the extent of over-diagnosis.**

Closer examination of age-specific incidence trends suggests that other factors, in addition to the introduction of mammography screening, may have influenced breast cancer incidence trends in South Australia. While increases in incidence appear to have coincided with the introduction of breast screening in South Australia, the pronounced

decrease in women aged 50-59 years in the late 1990's is not consistent with the expected effects of screening. The decline in incidence in this age group is likely to be due to the changing patterns of hormone replacement therapy use. Incidence in women aged 60-69 years however, has continued to increase despite declining use of HRT. As I proposed in Chapter 2, this apparent inconsistency could be explained by a shift to later diagnosis of breast cancers following the decline in use of hormone therapy, if HRT acts primarily as a promoter of tumour growth (rather than an initiator).

Due to the descriptive nature of the study presented in Chapter 2 it was not possible to claim a causal association between HRT use (as used by women in South Australia) and breast cancer incidence. However, findings presented in Chapter 3 provide convincing evidence of a direct causal link. A 30-40% increase in screen-detected cancers and a 90% increase in interval cancer rates were observed among screening participants who were using HRT at or immediately prior to having a mammogram. Interestingly, HRT use did not increase the incidence of DCIS. This may indicate a different etiological pathway for ductal carcinoma in-situ. Alternatively HRT may increase both the development of DCIS and the rate of transition to invasive disease, so that no overall change in DCIS incidence presented. Also, no impact was observed among women under the age of 55 years, suggesting either that risk is only increased among post-menopausal women, or that risk is associated with longer term use and hence not evident among younger women. Results in Chapter 3 also show that the increased risk of breast cancer was associated with current use but not past use. This provides more definitive evidence that the effects of HRT are transient, and dissipate shortly after stopping use, which is consistent with incidence patterns in women age 50-59 years. What is not clear and warrants further investigation is whether, and to what extent, the risk of breast cancer over a woman's entire lifespan is elevated by use of HRT if the primary biological mechanism is primarily through the promotion of occult tumours, rather than initiation of new tumours.

The direct link between HRT use and elevated risk of breast cancer is likely to have confounded and complicated attempts to estimate over-diagnosis in SA. South Australian women were relatively high users of hormone replacement therapy[138].

Based on data from BreastScreen SA, around 50% of women aged 50-59 years indicated current use during the mid-1990s. A similarly high prevalence of use was observed among women aged 60-69 years during the late 1990's. Even in 2008/9, levels of HRT use were relatively high with around 1 in 5 women aged in their 50's and 60's indicating 'current use'.

- **Women who participated in screening at Breast Screen have a higher risk of breast cancer than those who do not.**

Results of the comparative study outlined in Chapter 4 show that prevalence of several breast cancer risk factors was higher among BreastScreen participants than non-participants. The most notable differences were in relation to current or past HRT use (OR =3.9, 95%CI 2.3-6.1) and previous breast biopsy or surgery (OR=2.2, 95% CI 1.2-4.0) which may be indicative of increased benign breast disease. Due to the cross sectional nature of the survey it is not possible to determine whether a history of breast biopsy/surgery led to or was the result of increased participation in screening. The elevated odds ratio for use of HRT among screening participants is consistent with anecdotal evidence that General Practitioners in South Australia were strongly encouraging women who were being prescribed HRT to undertake mammography screening as a precautionary measure in light of evidence of increased risk of breast cancer, and that women were heeding this message.

Based on the Pfieffer risk prediction model[231] and on a simple count of the total number of risk factors, the overall risk of breast cancer was found to be higher among BreastScreen participants than non-participants. Elevated risk among screening participants will potentially inflate estimates of over-diagnosis in the SA context, in the absence of adequate adjustment, especially when methodologies involve the use of individual level data on screening participation, as in the case-control study outlined in Chapter 6.

- **Accounting for lead time effects due to screening and for differences in risk between comparison populations are critical to deriving unbiased estimation of over-diagnosis.**

The review of literature on over-diagnosis provided in Chapter 5 highlights the complexities associated with trying to quantify the level of over-diagnosis due to mammography screening. Adequate account needs to be taken of lead time effects from screening and potential differences in breast cancer risk in comparison populations. Previous estimates of over-diagnosis range from 0% to more than 50% of numbers of cancers expected in the absence of screening. However, a number of these studies are likely to have overstated the extent of over-diagnosis because they have not adequately accounted for lead time effects or for differences in breast cancer risk. How measures of over-diagnosis are calculated and reported is also a major source of variation in estimates, with estimates at the higher end of the range often referring to cancers expected to be diagnosed during the screening age range rather than for women's remaining lifetime.

- **The two different approaches used to estimate over-diagnosis due to mammography screening in South Australia indicate a modest level of over-diagnosis.**

Two different approaches were used to quantify over-diagnosis due to mammography screening in South Australia. The first of these was a case-control study which used individual level data on screening attendance and breast cancer outcomes to determine breast cancer risk at different time intervals after the last screening mammogram to account for lead time effects. The second method involved modelling lead time effects at the population level, using data on projected population incidence trends and screening participation rates, and estimates of lead time and screening sensitivity based on interval cancer rates. Estimates from both studies apply to the eligible screening population, with the observed number of breast cancers in the population exposed to screening as the denominator.

Findings from both the case-control study and the lead time modelling study indicate similar levels of excess cumulative incidence in the eligible population of around 8% of diagnosed IBC and 12-14% for all BC, without adjustment for any risk differences. However, they may overstate the level of over-diagnosis in the population due to the strong influence of HRT use on breast cancer incidence. In relation to the case-control study, estimates may be incorrect if the risk of breast cancer is not equivalent in those

who have attended screening and those who have not. Results shown in Chapter 4 indicate a higher risk of breast cancer among women who have attended screening, due predominately to higher prevalence of HRT use, which may have inflated the estimate of over-diagnosis. Sensitivity analyses, which adjusted for the elevated risk in screening participants, indicated substantially lower levels of over-diagnosis, when this risk difference was taken into account. Changing risk factor prevalence is also problematic in relation to the lead time modelling approach, given that modelling is predicated on estimating the background incidence in the absence of screening. If the background incidence was actually higher than predicted based on projections from before the commencement of screening (a likely scenario given evidence in Chapter 2), estimates based on comparing observed and lead time adjusted incidence would be inflated because some of the difference would be due to factors other than lead time effects. It is therefore likely that the estimates reported in this body of work represent a maximum estimate of over-diagnosis in the SA population.

Estimates for South Australia are consistent with findings from the long term follow up of the original mammography screening trials which did not offer screening to the control arm[39, 70, 71] and with the findings from several well-designed cohort studies that used individual level measures of screening and breast cancer incidence[83, 85]. Based on the evidence presented in this thesis and from other well-designed studies using individual data, mammography screening is likely to result in some over-diagnosis of breast cancer but, when considered over the remaining lifetime for women participating in screening, the extent of over-diagnosis is quite modest (~8% of IBC and 12-14% for all BC).

## STRENGTHS AND LIMITATIONS

All studies presented in this thesis have made use of quality population-wide data sources which are likely to provide highly accurate measures of exposure to screening and breast cancer outcomes. South Australia has a statutorily mandated reporting system with whole of population coverage, whereby hospitals, radiotherapy centres and pathology laboratories are required to notify all new cases of cancer (including in-situ cases for breast cancer) to the central cancer registry. The state also has one overarching breast screening service, allowing for a single centralised state-wide data

collection. BreastScreen SA has had a continuous and consistent record system which has collected participant and screening outcomes data for all screening episodes since the inception of the program. Linkage between the BreastScreen SA record system and the South Australia Cancer Registry is regularly undertaken to accurately identify interval cancers. Pathological details for all biopsy procedures are received directly from pathology reports. BreastScreen data are routinely audited for the purpose of national accreditation. Together, these systems have provided that high quality data underpin the research projects outlined in this thesis.

Data on self-reported use of HRT (at/around the time of screening) was also consistently collected in BreastScreen SA records from 1992 onwards and therefore is accessible at an individual level for all screening participants. Reporting of HRT use prior to screening was complete for almost all screening participants and is likely to be of reasonable quality given the practice of GPs to encourage women to participate in screening when prescribing hormone replacement therapies. These data were used to confirm that there was a direct association between HRT use and increased risk of breast in the screening population, which could not be assumed from due to the lack of specific information about the type and duration of use in the SA population.

Though not measurable at an individual level, information on potential confounders was obtained from the same population in which over-diagnosis was measured (i.e. South Australian women who were eligible to participate in the breast screening program). Weighted population-based survey data were used to determine trends in the prevalence of key risk factors over the 2 decades that screening has operated. These data were collected through the South Australian Health Omnibus Survey which is an annual face to face survey of a representative sample of the population. This survey has been conducted under the auspices of South Australia's Health Department in a consistent, statistically rigorous fashion since it was first undertaken in 1991 and is regarded as the gold standard health monitoring tool for the State[244]. Data used to compare risk factor profiles between BreastScreen participants and non-participants were also collected through the SA Health Omnibus Survey (in 2012) so are again representative of population of women eligible for screening in South Australia.

A key limitation in this work is the uncertainty regarding the background incidence rate in the absence of screening. The final measures of over-diagnosis in both study designs required estimation of the background incidence rates without screening. Projections assume no major change in trends over a 20 year period, which is unlikely to have been the case. Assessment of trends in incidence had screening not been implemented is fundamental to the lead time adjustment approach, as the background incidence is the starting point for modelling lead time effects. While results from the case-control are less dependent on the overall rate of increase in background incidence rates, inaccurate measures may still result if the rate of increase in incidence had differed across age groups, given the final estimates of over-diagnosis were derived by summing effects across age strata. The lack of data on HRT use and on the prevalence of other risk factors at a population level over the long term, i.e. 30 years, limits the ability to determine trends in background incidence rates more precisely. Likewise not having data on exposure to the key breast cancer risk factors, including HRT use, at an individual level is a major impediment to being able to accurately adjust for confounding factors when estimating over-diagnosis in the case-control study.

Another problematic area (which is relevant to the modelling study) relates to measuring lead time duration. Estimates were based on background incidence in the absence of screening and on interval cancer rates. The method used in this thesis circumvents the criticisms that lead time estimates include over-diagnosed cases because it was based on interval cancer rates which would, for the most part, present clinically and by definition cannot be overdiagnosed cases. However, our lead time estimates were also dependent on predicting background rates in the absence of screening, which cannot directly be observed. While it was possible to remove the effect of HRT on interval cancer rates using data for the screening population to estimate age-specific lead times, adjustment of the background incidence rates was not possible due to the lack of data on HRT prevalence in the population prior to the introduction of screening. Estimates of lead time have relied on the assumption that the projected pre-screening trend were not distorted by increasing use of HRT, which is difficult to validate.

These factors, along with limited knowledge about how HRT use affects tumour growth patterns, result in a degree of uncertainty about the validity of estimates of over-diagnosis arising from the studies presented in this thesis. These same concerns will also apply more generally to most other observational studies which have attempted to quantify the level of over-diagnosis.

## POLICY IMPLICATIONS

Despite the uncertainty that remains in relation to the exact level of over-diagnosis resulting from mammography screening, the ongoing debate is already having an impact on screening programs. This is exemplified by a recent review of mammography screening in Switzerland which stopped just short of calling a halt to screening but instead recommended no further expansion of screening in regions where programs are not in place[245].

I believe it would be unwise to make any changes to current screening programs or protocols in Australia based on available evidence regarding the extent of over-diagnosis due to screening. This thesis has highlighted the complexities in estimating over-diagnosis. Findings from a number of previously published studies may be inflated due to lead time effects not being adequately taken into account and/or differences in risk between comparison groups[89]. Estimates for South Australia are also uncertain due to limitations in design and the very real potential that other factors have impacted on breast cancer rates at both the population and individual level.

Furthermore, this thesis has focused on measuring one side of the 'balance sheet' of benefits and harms from mammography screening (i.e. reduced risk of breast cancer death, less invasive treatments versus increased investigation of false positives and over-diagnosis). Examination of both sides of the equation is important in determining policy implications for screening programs. Controversies also extend to estimating the benefits of mammography screening and opinions are divided. Generally the same group of researchers who claim that mammography leads to high levels of over-diagnosis also argue that there is little benefit in terms of mortality reduction due to screening[246], suggesting that, to some extent, divisions may be driven by subjectivity and *a priori* predispositions. While a critical review of the evidence for mortality and morbidity reductions due to mammography screening is beyond the scope of this

thesis, the evidence does appear to be favourable. Seven of the eight original screening trials indicated a reduction in breast cancer mortality, though not all findings were statistically significant. No reduction in breast cancer mortality was observed in the Canadian NBSS trials I and II[210]. However, the first systematic review of mammography screening by the Cochrane team, concluded that 'Screening for breast cancer is unjustified ...(as)… there is no reliable evidence that screening decreases breast cancer mortality'[247]. This review has been highly influential in informing the debate (or misinforming according to many). The authors based their conclusion on a meta-analysis of only two trials (the Canadian and Goteborg trials), on the grounds that these were the only trials of sufficient quality to justify inclusion in their analysis. More recently, the Cochrane team have conceded a small reduction in mortality due to screening (reporting 19% reduction based on 7 trials but assuming 15% due to the suboptimal quality of 4 of these trials[207, 212, 213]). Meta-analyses undertaken by other peak bodies have assessed the overall mortality reduction among women invited to screening and aged 50-69 years to be in the order of 20-25%[24, 39, 248, 249]. Observational studies conducted in the context of more recent service screening have confirmed benefits in terms of mortality reductions[33, 34, 50, 250-252]. The reduction observed in the context of screening programs (i.e. 20-30% for invited cohorts and 25-50% for women who were actually screened) is at least as large if not larger than that observed in trials[50, 250, 252]. Others have argued that advances in treatment have had a far greater impact on reducing breast cancer mortality than mammography screening[253, 254]. While improved treatment could explain some of the reduction in population-wide trends in breast cancer mortality, it cannot account for mortality differences noted in case-control and cohort studies. There would need to be substantial and consistent differences in the treatments protocols for breast cancers detected through screening and those that present clinically to explain the sizable reductions in breast cancer mortality observed among screening participants compared with non-participants. Favourable findings across many studies of differing designs, conducted in different settings, suggest the benefits of screening in terms of mortality reduction are real and independent of improvements in treatment.

Criticism has been levelled at cancer screening programs for traditionally emphasising the benefits of screening and ignoring or down-playing the risks, to encourage high

participation rates[255-257]. Most screening advocates and opponents agree that women should be provided with information about the risks and benefits associated with mammography screening so that they can make better informed decisions about whether or not to participate. While little is known about what constitutes the best approach to risk communication and what information actually enhances decision making, some underlying principles have been suggested[258]. These include ensuring information is 'exact', 'concrete' and 'simple'. In relation to breast screening these principles are hard to achieve, due to the methodological complexities, divergent opinions and uncertainties regarding final estimates.

Several attempts have been made to develop a balance sheet of the risks and benefits associated with mammography screening to help inform women's decision making[259-262]. Two recent examples highlight the problems associated with such endeavours, in the absence of any consensus about the extent of risks and benefits. The balance sheet developed by the Euroscreen Working Group quantifies the benefits (lives saved) and risks (over-diagnosed cases and false positive assessments) per 1000 women screening biennially from age 50 to 69 years with 10 years of post-screening follow-up[259, 260]. An equivalent guide to breast cancer screening decisions developed by researchers in the USA quantifies a similar set of benefits and risks for annual screening for a 10 year period (for different age groups)[261]. Both sets of figures were developed on the basis of extensive literature reviews and meta-analyses. Euroscreen's balance sheet reported 7-9 lives saved, 4 over-diagnosed cases and 170 false positive recalls for assessment per 1000 women screening[260]. In contrast the US balance sheet suggested that one life would be saved per 1000 women screening (for women screening in their 50's), one in five screen-detected cancers would be overdiagnosed and one instance of false positive recall for every women attending over the 10 years of annual screening [261]. While some differences are likely to be due to variation in screening protocols and threshold criteria for recall for further assessment, it is obvious that the use of different data sources and interpretation of study findings can lead to very different balance sheets. The potential for women to be confused by the very different messages regarding the level of benefits and risks that may be presented from various "expert" bodies is considerable.

To date, only two studies have examined women's preferences for and reactions to information about over-diagnosis due to mammography screening. These were conducted in NSW, Australia[263] and in the UK[264]. Findings from both of these qualitative studies were strikingly similar. Even though the issue of over-diagnosis was confusing for many, most were able to understand the concept and felt strongly that women should be informed about the risks. For the most part, being given information about over-diagnosis was unlikely to greatly influence women's future intentions to participate in screening. Most were not dissuaded and preferred to "err on the side of caution". In the Australian study, where women were presented with a range of estimates of over-diagnosis (1-10%, 30% and 50%), reactions varied according to the level of over-diagnosis. While 10% or less was dismissed as negligible and 30% considered acceptable, 50% over-diagnosis led to higher levels of concern which could potentially deter women from screening[263]. Based on these results, estimates of over-diagnosis for South Australia (8% for invasive BC and 12-14% all BC) are well within the range considered to be acceptable.

Interestingly, many women also viewed over-diagnosis as a treatment issue, with some women indicating that they may delay or consider alternative treatments in light of information about over-diagnosis. Careful consideration regarding messages about over-diagnosis is required, given that the implications are broader than simply informing the decision to screen.

My personal view, based on the literature and my own research findings, is that it is not possible to provide a definitive measure of the extent of over-diagnosis due to the uncertainties and complexities associated with quantifying something that is essentially unobservable. Information provided to women should therefore be more general in nature. The risk of over-diagnosis could be described as low to moderate, while acknowledging the high degree of uncertainty about the precise level. Information should clearly indicate that it is not possible to tell at an individual level which cancers will and won't develop into life-threatening disease.

## FUTURE RESEARCH

One aspect that warrants further investigation is the potential for the effects of screening to have been modified by hormone replacement therapy use. Findings

presented in this thesis point to some of the increase in incidence being due to the high prevalence of HRT use in SA. Even during the period of decline (from 2002 onwards) the influence of HRT may still have been impacting on incidence through a shift toward later diagnosis of breast cancer due to a slowing of growth rates in hormone responsive tumours when HRT was withdrawn. Incidence trends among the 50-59 and 60-69 year olds support this proposition. It is also highly likely that the impact of mammography screening on incidence was exaggerated over the first 15 years of the life of the program due to the high prevalence of HRT use in South Australia, given that HRT most likely increases tumour growth rates and that screening detects tumour with relatively small diameters compared to clinical presentation. Differences in the uptake of HRT may also be an explanation for the variable findings across studies, which to date has received little attention. Analysis of trends in hormone receptor status of screen-detected breast cancers among HRT users and non-users may be helpful in this regard.

Determining the biological mechanism by which HRT use increases risk has important implications for estimating over-diagnosis. Whether HRT acts solely as a promoter of breast cancer or also acts as an initiator of new tumours is yet to be established. As a promoter, HRT use would serve to shift the age distribution in relation to breast cancer diagnoses, with little or no increase in the overall number of breast cancers between users and non-users, and thus could be accounted for in a similar way to lead time effects. If, on the other hand the use of hormone therapy plays even a small role in the initiation of breast cancers, confounding would be a serious issue, where prevalence of HRT use was high, due to an increase in the number of breast cancers among users.

Some researchers believe that over-diagnosis may be due to detection of breast cancers that would have regressed and therefore never surfaced clinically. While direct evidence for regression in breast cancer is scant, the notion of regression is not implausible. Indirect evidence from animal models, modelling studies, and the success of hormonal therapies in treating breast cancer all support the concept of hormonally induced regression of breast tumours[265, 266]. Further investigation into the relationship between hormone use and tumour regression is warranted, particularly in the light of the dramatic changes in prevalence of HRT use during and since the introduction of

screening. (While the above discussion refers to HRT as a single entity it is clear that effects vary according to the specific type of HRT use, the mode of delivery, duration of use and timing of commencement after menopause and any future research efforts need to take this into account).

A definitive answer regarding the level of over-diagnosis due to mammography screening may never be reached. Conducting a randomised controlled trial to quantify over-diagnosis is not feasible in the current environment where screening is widespread. Furthermore, RCTs would not yield answers for several decades, by which time the screening technologies in question would be outdated. A historic cohort design or a case-control study (as demonstrated in this thesis) provides the best alternative option for measuring over-diagnosis. However, the validity of such approaches relies on the availability of data at the individual level on a range of breast cancer risk factors to adequately control for confounding due to self-selection into screening programs. The case-control study outlined in this thesis would have been greatly enhanced had individual level data on breast cancer risk factors been available or collectable for all cases and controls. Rarely would data covering the whole population over such an extended period of time be accessible. Unless such data are available to provide an accurate profile of risk at the individual level, continuing attempts to measure the extent of over-diagnosis may not be the best use of resources.

The question of how much over-diagnosis is due to screening may be better reframed as one of how best to treat small, low grade, node negative, receptor positive tumours (detected at screening) in more minimally invasive ways. Findings from the studies presented in this thesis suggest at least 30-40% of the over-diagnosis burden is due to the detection of pre-invasive DCIS. Questions relating to appropriate management of ductal carcinoma in-situ are therefore relevant to the issue of over-diagnosis. In light of this, consideration could be given to conducting a trial of active surveillance without immediate treatment for low grade, pre-invasive cancers. Prospective or retrospective observational studies of women who elect not to have treatment may be a more feasible alternative given that recruitment may be difficult for a trial of 'no treatment'. Only one study to date has looked at outcomes for women electing non-surgical

surveillance with endocrine therapy alone in relation to ER-positive DCIS, which included only 12 patients[267]. This issue has recently been suggested as one of the top ten priorities in relation to future research into the management of DCIS[268] and is supported by others[269].

Efforts may also be better directed toward developing tools to accurately identify those cancers with little or no potential to progress. Over the past two decades, a new model of breast cancer tumorgenesis has emerged in which low and high grade cancer subtypes are seen as separate entities that follow distinct paths (i.e. low grade DCIS develops into low grade invasive ductal carcinoma, while high grade DCIS develops into high grade IBC). This contrasts with the traditional view of tumorgenesis as a process of incremental transition from low grade lesions to high grade cancers. Genetic profiling has identified several different molecular subtypes of breast cancer which differ with respect to prognostic outcomes. This has spawned a plethora of gene expression profiling tests aimed at providing a more accurate assessment of risk of progression/recurrence and guiding treatment decision-making. While potentially promising, risk stratification based on molecular profiling has not yet been shown to offer any advantage over conventional, clinically-based classification methods using tumour grade, size, nodal involvement, receptor status or mitotic activity[270].

## CONCLUDING REMARKS

Debate over the extent of over-diagnosis due to mammography screening has been fierce and opinion continues to be divided. Estimation of over-diagnosis is compounded by the complex interplay of multiple factors which influence breast cancer incidence. Claims which state with certainty that the level of over-diagnosis due to mammography screening is high (i.e. in the order of 30% or higher) should be viewed with some scepticism. While it is highly likely that mammography screening does lead to some over-diagnosis, results of the two studies presented in this thesis suggest the level of over-diagnosis is fairly modest in the context of lifetime risk of breast cancer.

# References

1. Welch HG, Black WC. Overdiagnosis in cancer. J Natl Cancer Inst 2010;102:605-13.

2. Biesheuvel C, Barratt A, Howard JM, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. Lancet Oncol 2007;8:1129-38.

3. de Gelder R. Interpreting overdiagnosis estimates in population-based mammography screening. Epidemiol Rev 2011;33:111-21.

4. Australian Institute of Health and Welfare, Cancer in Australia 2010: an overview. Cancer series no. 60. Cat. no. CAN 56. AIHW 2010.

5. International Agency for Research on Cancer. World Cancer Report 2008 ed. Lyon, France IARC Press 2008.

6. Miller BA, Feuer EJ, Hankey BF. Increasing incidence of breast cancer since 1982: relevance of early detection. Cancer Causes Control 1991;2:67-74.

7. Garfinkel L, Boring CC, Health CW. Changing trends. An overview of breast cancer incidence and mortality. Cancer 1994;74:222-7.

8. Bray F, McCarron P, Parkin DM. The changing global patterns of female breast cancer incidence and mortality. Breast Cancer Res 2004;6:229-39.

9. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. GLOBOCAN 2008 v1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 10 [Internet]. Lyon, France: International Agency for Research on Cancer, 2010.

10. South Australian Cancer Registry, Cancer in South Australia 2007 - with projections to 2010. Adelaide: South Australian Department of Health 2010.

11. Australian Institute of Health and Welfare, Australasian Association of Cancer Registries, Cancer in Australia: an overview, 2008. Cancer Series no. 46. Cat. no. CAN 42. Canberra: AIHW 2008.

12. Australian Institute of Health and Welfare, Breast cancer in Australia: an overview. Cancer Series No. 50. Cat. no. CAN 46. Canberra: AIHW 2009.

13. Jones JL. Overdiagnosis and overtreatment of breast cancer: progression of ductal carcinoma in-situ: the pathological perspective. Breast Cancer Res 2006;8:204.

14. Bombonati A, Sgroi DC. The molecular pathology of breast cancer progression. J Pathol 2011;223:307-17.

15. Tsikitis VL, Chung MA. Biology of ductal carcinoma in-situ classification based on bological potential. Am J Clin Oncol 2006;29:305-10.

16. Sgroi DC. Preinvasive Breast Cancer. Annu. Rev. Pathol. Mech. Dis. 2010;5:193-221.

17. Schnitt SJ. Molecular biology of breast tumor progression: a view from the other side. Int J Surg Pathol 2010;18:170-3S.

18. Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in-situ of the breast: a review. Breast Cancer Res Treat 2006;97:135-44.

19. Nguyen DX, Massague J. Genetic determinants of cancer metastasis. Nat Rev Genet 2007;8:341-52.

20. Goss PE, Chambers AF. Does tumour dormancy offer a therapeutic target? Nat Rev Cancer 2010;10:871-7.

21. Kerlikowske K. Epidemiology of ductal carcinoma in-situ. J Natl Cancer Inst Monogr 2010;2010:139-41.

22. Luke C, Priest K, Roder D. Changes in Incidence of in situ and invasive breast cancer by histology type following mammography screening. Asian Pacific J Cancer Prev 2006;7:69-74.

23. Estevez LG, Alvarez I, Segui MA, Munoz M, Margeli M, Miro C, Rubio C, Lluch A, Tusquets I. Current perspectives of treatment of ductal carcinoma in-situ. Cancer Treat Rev 2010;36:507-17.

24. International Agency for Research on Cancer, IARC Handbooks of Cancer Prevention Volume 7: Breast Cancer Screening. Lyon, France: IARC Press 2002.

25. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 2. Breast cancer detection and death rates among women aged 50 to 59 years. CMAJ 1992;147:1477-88.

26. Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study: 1. Breast cancer detection and death rates among women aged 40 to 49 years. CMAJ 1992;147:1459-76.

27. Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study: update on breast cancer mortality. J Natl Cancer Inst Monogr 1997:37-41.

28. Shapiro S, Venet W, Strax P, Venet L, Roeser R. Selection, follow-up, and analysis in the Health Insurance Plan Study: a randomized trial with breast cancer screening. Natl Cancer Inst Monogr 1985;67:65-74.

29. Tabar L, Fagerberg CJ, Gad A, Baldetorp L, Holmberg LH, Grontoft O, Ljungquist U, Lundstrom B, Manson JC, Eklund G, et al. Reduction in mortality from breast cancer after mass screening with mammography. Randomised trial from the Breast Cancer Screening Working Group of the Swedish National Board of Health and Welfare. Lancet 1985;1:829-32.

30. Alexander FE, Anderson TJ, Brown HK, Forrest AP, Hepburn W, Kirkpatrick AE, Muir BB, Prescott RJ, Smith A. 14 years of follow-up from the Edinburgh randomised trial of breast cancer screening. Lancet 1999;353:1903-8.

31. Bjurstam N, Bjorneld L, Warwick J, Sala E, Duffy SW, Nystrom L, Walker N, Cahlin E, Eriksson O, Hafstrom LO, Lingaas H, Mattsson J, et al. The Gothenburg Breast Screening Trial. Cancer 2003;97:2387-96.

32. Frisell J, Lidbrink E, Hellstrom L, Rutqvist LE. Followup after 11 years -update of mortality results in the Stockholm mammographic screening trial. Breast Cancer Res Treat 1997;45:263-70.

33. Walter SD. Mammographic screening: case-control studies. Ann Oncol 2003;14:1190-2.

34. Gabe R, Duffy SW. Evaluation of service screening mammography in practice: the impact on breast cancer mortality. Ann Oncol 2005;16:153-62.

35. Green BB, Taplin SH. Breast cancer screening controversies. J Am Board Fam Pract 2003;16:233-41.

36. Kerlikowske K, Grady D, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. JAMA 1995;273:149-54.

37. Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. Lancet 2001;358:1340-2.

38. U.S. Preventive Services Task Force. Screening of breast cancer: recommendations and rationale. Ann Intern Med 2002;137:344-6.

39. Independent UK Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. Lancet 2012;380:1778-86.

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

41. Moss S, Waller M, Anderson TJ, Cuckle H. Randomised controlled trial of mammographic screening in women from age 40: predicted mortality based on surrogate outcome measures. Br J Cancer 2005;92:955-60.

42. Djulbegovic B, Lyman GH. Screening mammography at 40-49 years: regret or no regret? Lancet 2006;368:2035-7.

43. Cox B. The effect of service screening on breast cancer mortality rates. Eur J Cancer Prev 2008;17:306-11.

44. Duffy SW, Tabar L, Chen HH, Holmqvist M, Yen MF, Abdsalah S, Epstein B, Frodis E, Ljungberg E, Hedborg-Melander C, Sundbom A, Tholin M, et al. The impact of

organized mammography service screening on breast carcinoma mortality in seven Swedish counties. Cancer 2002;95:458-69.

45. 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, et al. Effectiveness of service screening: a case-control study to assess breast cancer mortality reduction. Br J Cancer 2008;99:423-7.

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

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

48. Roder D, Houssami N, Farshid G, Gill G, Luke C, Downey P, Beckmann K, Iosifidis P, Grieve L, Williamson L. 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 2008;108:409-16.

49. Thompson RS, Barlow WE, Taplin SH, Grothaus L, Immanuel V, Salazar A, Wagner EH. A population-based case-cohort evaluation of the efficacy of mammographic screening for breast cancer. Am J Epidemiol 1994;140:889-901.

50. Schopper D, de Wolf C. How effective are breast cancer screening programmes by mammography? Review of the current evidence. Eur J Cancer 2009;45:1916-23.

51. BreastScreen SA, The impact of BreastScreen SA, 20 year report 1989-2008: with trend data for 1989-2004. Adelaide: Government of South Australia 2010.

52. Duffy SW, Agbaje O, Tabar L, Vitak B, Bjurstam N, Bjorneld L, Myles JP, Warwick J. Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. Breast Cancer Res 2005;7:258-65.

53. Morrison AS, Screening in Chronic Disease. Monographs in Epidemiology and Biostatistics Oxford: Oxford University Press, 1985.

54. Zahl PH, Maehlen J, Welch HG. The natural history of invasive breast cancers detected by screening mammography. Arch Intern Med 2008;168:2311-6.

55. Barrett A. Cancer screening benefits, harms and making informed choice. Australian Family Physician 2006;35:39-42.

56. Paci E, Duffy S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis and overtreatment in service screening. Breast Cancer Res 2005;7:266-70.

57. Puliti D, Miccinesi G, Paci E. Overdiagnosis in breast cancer: Design and methods of estimation in observational studies. Preventive Medicine 2011;53:131-3.

58. National Breast and Ovarian Cancer Centre. Over-diagnosis from mammography screening - NBOCC Position Statement. Sydney: NBOCC, 2008.

59. Duffy SW, Lynge E, Jonsson H, Ayyaz S, Olsen AH. Complexities in the estimation of overdiagnosis in breast cancer screening. Br J Cancer 2008;99:1176-8.

60. Neutel CI, Morrison H. Could recent decreases in breast cancer incidence really be due to lower HRT use? Trends in attributable risk for modifiable breast cancer risk factors in Canadian women. Can J Public Health 2010;101:405-9.

61. Farquhar C, Marjoribanks J, Lethaby A, Suckling JA, Lamberts Q. Long term hormone therapy for perimenopausal and postmenopausal women. Cochrane Database Syst Rev 2009:CD004143.

62. MacLennan AH, Gill TK, Broadbent JL, Taylor AW. Continuing decline in hormone therapy use: population trends over 17 years. Climacteric 2009;12:122-30.

63. Main P, Robinson M. Changes in utilisation of hormone replacement therapy in Australia following publication of the findings of the Women's Health Initiative. Pharmacoepidemiol Drug Saf 2008;17:861-8.

64. Canfell K, Banks E, Clements M, Kang YJ, Moa A, Armstrong B, Beral V. Sustained lower rates of HRT prescribing and breast cancer incidence in Australia since 2003. Breast Cancer Res Treat 2009;117:671-3.

65. Crocetti E, Buzzoni C, Falcini F, Cortesi L, De Lisi V, Ferretti S, Tumino R, Russo A, Paci E. Disentangling the roles of mammographic screening and HRT in recent breast cancer incidence trends in Italy by analyses based on calendar time and time since screening activation. Breast J 2010;16:350-5.

66. Marshall SF, Clarke CA, Deapen D, Henderson K, Largent J, Neuhausen SL, Reynolds P, Ursin G, Horn-Ross PL, Stram DO, Templeman C, Bernstein L. Recent breast cancer incidence trends according to hormone therapy use: the California Teachers Study cohort. Breast Cancer Res 2010;12:R4.

67. Martin RM, Wheeler BW, Metcalfe C, Gunnell D. What was the immediate impact on population health of the recent fall in hormone replacement therapy prescribing in England? Ecological study. J Public Health (Oxf) 2010;32:555-64.

68. Waller M, Moss S, Watson J, Moller H. The effect of mammographic screening and hormone replacement therapy use on breast cancer incidence in England and Wales. Cancer Epidemiol Biomarkers Prev 2007;16:2257-61.

69. Duffy SW, Tabar L, Olsen AH, Vitak B, Allgood PC, Chen TH, Yen AM, Smith RA. Absolute numbers of lives saved and overdiagnosis in breast cancer screening, from a randomized trial and from the Breast Screening Programme in England. J Med Screen 2010;17:25-30.

70. Moss S. Overdiagnosis and overtreatment of breast cancer: overdiagnosis in randomised controlled trials of breast cancer screening. Breast Cancer Res 2005;7:230-4.

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

72. Peeters PH, Verbeek AL, Straatman H, Holland R, Hendriks JH, Mravunac M, Rothengatter C, Van Dijk-Milatz A, Werre JM. Evaluation of overdiagnosis of breast

cancer in screening with mammography: results of the Nijmegen programme. Int J Epidemiol 1989;18:295-9.

73. Svendsen AL, Olsen AH, von Euler-Chelpin M, Lynge E. Breast cancer incidence after the introduction of mammography screening: what should be expected? Cancer 2006;106:1883-90.

74. Zahl PH, Strand BH, Maehlen J. Incidence of breast cancer in Norway and Sweden during introduction of nationwide screening: prospective cohort study. Br. Med. J. 2004;328:921-4.

75. Jorgensen KJ, Zahl PH, Gotzsche PC. Overdiagnosis in organised mammography screening in Denmark. A comparative study. BMC Womens Health 2009;9:36.

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

77. Morrell S, Barratt A, Irwig L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. Cancer Causes Control 2010;21:275-82.

78. 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, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. Breast Cancer Res 2006;8:R68.

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

80. Zahl PH. Overdiagnosis of breast cancer in Denmark. Br J Cancer 2004;90:1686; author reply 7.

81. Puliti D, Zappa M, Miccinesi G, Falini P, Crocetti E, Paci E. An estimate of overdiagnosis 15 years after the start of mammographic screening in Florence. Eur J Cancer 2009;45:3166-71.

82. Jorgensen KJ, Gotzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. Br. Med. J. 2009;339:b2587.

83. Falk RS, Hofvind S, Skaane P, Haldorsen T. Overdiagnosis among women attending a population-based mammography screening program. Int J Cancer 2013;133:705-12.

84. Njor SH, Olsen AH, Blichert-Toft M, Schwartz W, Vejborg I, Lynge E. Overdiagnosis in screening mammography in Denmark: population-based cohort study. Br. Med. J. 2013;346:f1064.

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

86. de Koning HJ, Draisma G, Fracheboud J, de Bruijn A. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based on observed screen and clinical data. Breast Cancer Res 2006;8:202.

87. Mandelblatt JS, Cronin KA, Bailey S, Berry DA, de Koning HJ, Draisma G, Huang H, Lee SJ, Munsell M, Plevritis SK, Ravdin P, Schechter CB, et al. Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. Ann Intern Med 2009;151:738-47.

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

89. Puliti D, Duffy SW, Miccinesi G, de Koning H, Lynge E, Zappa M, Paci E. Overdiagnosis in mammography screening for breast cancer in Europe: a literature review. J Med Screen 2012;19 Suppl 1:42-56.

90. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan BK, Humphrey L. Screening for breast cancer: an update for the U.S. Preventive Services Task Force. Ann Intern Med 2009;151:727-37, W237-42.

91. Kopans DB. The most recent breast cancer screening controversy about whether mammographic screening benefits women at any age: nonsense and nonscience. AJR Am J Roentgenol 2003;180:21-6.

92. Krieger N, Chen JT, Waterman PD. Decline in US breast cancer rates after the Women's Health Initiative: socioeconomic and racial/ethnic differentials. Am J Public Health 2010;100 Suppl 1:S132-9.

93. Canfell K, Banks E, Moa AM, Beral V. Decrease in breast cancer incidence following a rapid fall in use of hormone replacement therapy in Australia. Med J Aust 2008;188:641-4.

94. De Prithwish C, Neutel CI, Olivotto I, Morrison H. Breast cancer incidence and hormone replacement therapy in Canada. J Natl Cancer Inst 2010;102:1489-95.

95. Daubisse-Marliac L, Delafosse P, Boitard JB, Poncet F, Grosclaude P, Colonna M. Breast cancer incidence and time trend in France from 1990 to 2007: a population-based study from two French cancer registries. Ann Oncol 2011;22:329-34.

96. Hofvind S, Sakshaug S, Ursin G, Graff-Iversen S. Breast cancer incidence trends in Norway explained by hormone therapy or mammographic screening? Int J Cancer 2012;130:2930-8.

97. Holford TR, Cronin KA, Mariotto AB, Feuer EJ. Changing patterns in breast cancer incidence trends. J Natl Cancer Inst Monogr 2006:19-25.

98. Gompel A, Plu-Bureau G. Is the decrease in breast cancer incidence related to a decrease in postmenopausal hormone therapy? Ann N Y Acad Sci 2010;1205:268-76.

99. Jemal A, Ward E, Thun MJ. Recent trends in breast cancer incidence rates by age and tumor characteristics among U.S. women. Breast Cancer Res 2007;9:R28.

100. Glass AG, Lacey JV, Jr., Carreon JD, Hoover RN. Breast cancer incidence, 1980-2006: combined roles of menopausal hormone therapy, screening mammography, and estrogen receptor status. J Natl Cancer Inst 2007;99:1152-61.

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

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

103. Roder D, de Silva P, Zorbas H, Kollias J, Malycha P, Pyke C, Campbell I, Webster F. Survival from synchronous bilateral breast cancer: the experience of surgeons participating in the breast audit of the Society of Breast Surgeons of Australia and New Zealand. Asian Pac J Cancer Prev 2012;13:1413-8.

104. Roder D, Zorbas H, Kollias J, Pyke C, Walters D, Campbell I, Taylor C, Webster F. Factors predictive of treatment by Australian breast surgeons of invasive female breast cancer by mastectomy rather than breast conserving surgery. Asian Pac J Cancer Prev 2013;14:539-45.

105. Australian Bureau of Statistics. 3101.0 - Australian Demographic Statisitics, June 2012 Canberra: ABS, 2012.

106. National Cancer Insitute. Joinpoint Regression Program, Version 3.5.4 - Aug 2012. In: Statisitical Methodology and Applications Branch, ed. Bethesda, MD.: NCI, 2012.

107. Wilson DH, Wakefield M, Taylor AW. The South Australian Health Omibus Survey. Health Promotion J Aust 1992;2:47-9.

108. Australian Bureau of Statistics. 3105.0.65.001 - Australian Historical Population Statistics, 2008 Canberra: ABS, 2008.

109. Chlebowski RT, Kuller LH, Prentice RL, Stefanick ML, Manson JE, Gass M, Aragaki AK, Ockene JK, Lane DS, Sarto GE, Rajkovic A, Schenken R, et al. Breast cancer after use of estrogen plus progestin in postmenopausal women. N Engl J Med 2009;360:573-87.

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

111. Dietel M. Hormone replacement therapy (HRT), breast cancer and tumor pathology. Maturitas 2010;65:183-9.

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

113. Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. Epidemiol Rev 1993;15:36-47.

114. Brooks PJ, Zakhari S. Moderate alcohol consumption and breast cancer in women: from epidemiology to mechanisms and interventions. Alcoholism, Clinical and Experimental Research 2013;37:23-30.

115. Eliassen AH, Colditz GA, Rosner B, Willett WC, Hankinson SE. Adult weight change and risk of postmenopausal breast cancer. JAMA 2006;296:193-201.

116. Cleary MP, Grossmann ME. Minireview: Obesity and breast cancer: the estrogen connection. Endocrinology 2009;150:2537-42.

117. Duffy SW, Parmar D. Overdiagnosis in breast cancer screening: the importance of length of observation period and lead time. Breast Cancer Res 2013;15:R41.

118. Nassar A. Core needle biopsy versus fine needle aspiration biopsy in breast cancer - a historical perspective and opportunities in the modern era. Diagnostic Cytopathology 2011;39:380-8.

119. Australian Institute of Health and Welfare, Dixon T, Walters A, A growing problem: trends and patterns in overweight and obesity among adult Australians, 1980 to 2001 Canberra: AIHW, 2003.

120. Australian Bureau of Statistics, Smoking, risky drinking and obesity - 4102.0 - Australain Social Trends. Canberra: ABS, 2009.

121. Reeves GK, Beral V, Green J, Gathani T, Bull D. Hormonal therapy for menopause and breast cancer risk by histological type: a cohort study and meta-analysis. Lancet Oncol 2006;7:910-8.

122. Ross RK, Paganini-Hill A, Wan PC, Pike MC. Effect of hormone replacement therapy on breast cancer risk: estrogen versus estrogen plus progestin. J Natl Cancer Inst 2000;92:328-32.

123. Brinton LA, Richesson D, Leitzmann MF, Gierach GL, Schatzkin A, Mouw T, Hollenbeck AR, Lacey JV, Jr. Menopausal hormone therapy and breast cancer risk in the NIH-AARP Diet and Health Study Cohort. Cancer Epidemiol Biomarkers Prev 2008;17:3150-60.

124. Li CI, Malone KE, Porter PL, Weiss NS, Tang MT, Cushing-Haugen KL, Daling JR. Relationship between long durations and different regimens of hormone therapy and risk of breast cancer. JAMA 2003;289:3254-63.

125. Bakken K, Fournier A, Lund E, Waaseth M, Dumeaux V, Clavel-Chapelon F, Fabre A, Hemon B, Rinaldi S, Chajes V, Slimani N, Allen NE, et al. Menopausal hormone therapy and breast cancer risk: impact of different treatments. The European Prospective Investigation into Cancer and Nutrition. Int J Cancer 2011;128:144-56.

126. Boyd NF, Martin LJ, Li Q, Sun L, Chiarelli AM, Hislop G, Yaffe MJ, Minkin S. Mammographic density as a surrogate marker for the effects of hormone therapy on risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2006;15:961-6.

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

128. Roman R, Sala M, Salas D, Ascunce N, Zubizarreta R, Castells X. Effect of protocol-related variables and women's characteristics on the cumulative false-positive risk in breast cancer screening. Ann Oncol 2012;23:104-11.

129. Kavanagh AM, Mitchell H, Giles GG. Hormone replacement therapy and accuracy of mammographic screening. Lancet 2000;355:270-4.

130. Banks E. Hormone replacement therapy and the sensitivity and specificity of breast cancer screening: a review. J Med Screen 2001;8:29-34.

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

132. Banks E, Reeves G, Beral V, Bull D, Crossley B, Simmonds M, Hilton E, Bailey S, Barrett N, Briers P, English R, Jackson A, et al. Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: cohort study. Br. Med. J. 2004;329:477.

133. Banks E, Reeves G, Beral V, Bull D, Crossley B, Simmonds M, Hilton E, Bailey S, Barrett N, Briers P, English R, Jackson A, et al. Impact of use of hormone replacement therapy on false positive recall in the NHS breast screening programme: results from the Million Women Study. BMJ 2004;328:1291-2.

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

135. Litherland JC, Stallard S, Hole D, Cordiner C. The effect of hormone replacement therapy on the sensitivity of screening mammograms. Clin Radiol 1999;54:285-8.

136. Virnig BA, Wang SY, Shamilyan T, Kane RL, Tuttle TM. Ductal carcinoma in-situ: risk factors and impact of screening. J Natl Cancer Inst Monogr 2010:113-6.

137. Calvocoressi L, Stowe MH, Carter D, Claus EB. Postmenopausal hormone therapy and ductal carcinoma in-situ: a population-based case-control study. Cancer Epidemiol 2012;36:161-8.

138. Taylor AW, MacLennan AH, Avery JC. Postmenopausal hormone therapy: who now takes it and do they differ from non-users? Aust N Z J Obstet Gynaecol 2006;46:128-35.

139. Roman R, Sala M, De La Vega M, Natal C, Galceran J, Gonzalez-Roman I, Baroja A, Zubizarreta R, Ascunce N, Salas D, Castells X. Effect of false-positives and women's characteristics on long-term adherence to breast cancer screening. Breast Cancer Res Treat 2011;130:543-52.

140. National Breast Cancer Centre, Breast imaging: a guide for practice. Camperdown NSW: National Breast Cancer Centre 2002.

141. Australian Bureau of Statistics, Information Paper: An Introduction to Socio-Economic Indexes for Areas (SEIFA) 2006. Canberra: ABS, 2008.

142. Zou G. A modified poisson regression approach to prospective studies with binary data. Am J Epidemiol 2004;159:702-6.

143. StataCorp. Stata Statisitical Software: Release 12 College Station, TX: StatCorp LP, 2010.

144. Henrich JB, Kornguth PJ, Viscoli CM, Horwitz RI. Postmenopausal estrogen use and invasive versus in-situ breast cancer risk. J Clin Epidemiol 1998;51:1277-83.

145. Persson I, Thurfjell E, Bergstrom R, Holmberg L. Hormone replacement therapy and the risk of breast cancer. Nested case-control study in a cohort of Swedish women attending mammography screening. Int J Cancer 1997;72:758-61.

146. Kavanagh AM, Davidson N, Jolley D, Heuzenroeder L, Chapman A, Evans J, Gertig DM, Amos A. Determinants of false positive recall in an Australian mammographic screening program. Breast 2006;15:510-8.

147. Kavanagh AM, Giles GG, Mitchell H, Cawson JN. The sensitivity, specificity, and positive predictive value of screening mammography and symptomatic status. J Med Screen 2000;7:105-10.

148. Sweeney KJ, Boran S, Fitzpatrick P, Merrigan AB, Moran R, Flanagan F, Kerin MJ. Postmenopausal hormone therapy and prevalent round screening mammography. Int J Cancer 2009;124:1490-1.

149. Seradour B, Esteve J, Heid P, Jacquemier J. Hormone replacement therapy and screening mammography: analysis of the results in the Bouches du Rhone programme. J Med Screen 1999;6:99-102.

150. Crouchley K, Wylie E, Khong E. Hormone replacement therapy and mammographic screening outcomes in Western Australia. J Med Screen 2006;13:93-7.

151. Warren R. Hormones and mammographic breast density. Maturitas 2004;49:67-78.

152. Weedon-Fekjaer H, Tretli S, Aalen OO. Estimating screening test sensitivity and tumour progression using tumour size and time since previous screening. Stat Methods Med Res 2010;19:507-27.

153. Santen RJ, Yue W, Heitjan DF. Modeling of the growth kinetics of occult breast tumors: role in interpretation of studies of prevention and menopausal hormone therapy. Cancer Epidemiol Biomarkers Prev 2012;21:1038-48.

154. Gapstur SM, Morrow M, Sellers TA. Hormone replacement therapy and risk of breast cancer with a favorable histology: results of the Iowa Women's Health Study. JAMA 1999;281:2091-7.

155. Stanford JL, Weiss NS, Voigt LF, Daling JR, Habel LA, Rossing MA. Combined estrogen and progestin hormone replacement therapy in relation to risk of breast cancer in middle-aged women. JAMA 1995;274:137-42.

156. Stefanick ML, Anderson GL, Margolis KL, Hendrix SL, Rodabough RJ, Paskett ED, Lane DS, Hubbell FA, Assaf AR, Sarto GE, Schenken RS, Yasmeen S, et al. Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. JAMA 2006;295:1647-57.

157. Kerlikowske K, Miglioretti DL, Ballard-Barbash R, Weaver DL, Buist DS, Barlow WE, Cutter G, Geller BM, Yankaskas B, Taplin SH, Carney PA. Prognostic

characteristics of breast cancer among postmenopausal hormone users in a screened population. J Clin Oncol 2003;21:4314-21.

158. Reeves GK, Pirie K, Green J, Bull D, Beral V. Comparison of the effects of genetic and environmental risk factors on in-situ and invasive ductal breast cancer. Int J Cancer 2012;131:930-7.

159. Burrell HC, Pinder SE, Wilson AR, Evans AJ, Yeoman LJ, Elston CW, Ellis IO. The positive predictive value of mammographic signs: a review of 425 non-palpable breast lesions. Clin Radiol 1996;51:277-81.

160. Cherel P, Becette V, Hagay C. Stellate images: anatomic and radiologic correlations. Eur J Radiol 2005;54:37-54.

161. Cui Y, Page DL, Lane DS, Rohan TE. Menstrual and reproductive history, postmenopausal hormone use, and risk of benign proliferative epithelial disorders of the breast: a cohort study. Breast Cancer Res Treat 2009;114:113-20.

162. Rohan TE, Negassa A, Chlebowski RT, Habel L, McTiernan A, Ginsberg M, Wassertheil-Smoller S, Page DL. Conjugated equine estrogen and risk of benign proliferative breast disease: a randomized controlled trial. J Natl Cancer Inst 2008;100:563-71.

163. Maxwell CJ, Bancej CM, Snider J. Predictors of mammography use among Canadian women aged 50-69: findings from the 1996/97 National Population Health Survey. CMAJ 2001;164:329-34.

164. Schueler KM, Chu PW, Smith-Bindman R. Factors associated with mammography utilization: a systematic quantitative review of the literature. J Womens Health 2008;17:1477-98.

165. von Euler-Chelpin M, Olsen AH, Njor S, Vejborg I, Schwartz W, Lynge E. Socio-demographic determinants of participation in mammography screening. Int J Cancer 2008;122:418-23.

166. Smith ML, Hochhalter AK, Ahn S, Wernicke MM, Ory MG. Utilization of screening mammography among middle-aged and older women. J Womens Health 2011;20:1619-26.

167. Vernon SW, Laville EA, Jackson GL. Participation in breast screening programs: a review. Soc Sci Med 1990;30:1107-18.

168. Meissner HI, Breen N, Taubman ML, Vernon SW, Graubard BI. Which women aren't getting mammograms and why? (United States). Cancer Causes Control 2007;18:61-70.

169. Duffy S, Cuzick J, Tabar L, Vitak B, Chen TH, Yen MF. Correcting for non-compliance in case-control studies to evaluate cancer screening programmes. Journal of the Royal Statistical Society . Series C (Applied Statistics) 2002;51:235-43.

170. Hay JL, McCaul KD, Magnan RE. Does worry about breast cancer predict screening behaviors? A meta-analysis of the prospective evidence. Prev Med 2006;42:401-8.

171. McCaul KD, Branstetter AD, Schroeder DM, Glasgow RE. What is the relationship between breast cancer risk and mammography screening? A meta-analytic review. Health Psychology;15:423-9.

172. Kim J, Jang SN. Socioeconomic disparities in breast cancer screening among US women: trends from 2000 to 2005. Journal of Preventive Medicine and Public Health, 2008;41:186-94.

173. Kjellen M, von Euler-Chelpin M. Socioeconomic status as determinant for participation in mammography screening: assessing the difference between using women's own versus their partner's. Int J Public Health 2010;55:209-15.

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

175. dos Santos Silva I, Beral V. Socioeconomic differences in reproductive behaviour. In: Kogevinas M, Pearce N, Susser M, Boffett P. Social Inequalities and

Cancered., vol. IARC Scientific Publications No.138. Lyon International Agency for Research on Cancer, 1997.

176. Lagerlund M, Sontrop JM, Zackrisson S. Do reproductive and hormonal risk factors for breast cancer associate with attendance at mammography screening? Cancer Causes Control 2013;24:1687-94.

177. Lopez-de-Andres A, Martin-Lopez R, Hernandez-Barrera V, Carrasco-Garrido P, Gil-de-Miguel A, Esteban y Pena MM, Jimenez-Garcia R. Predictors of breast and cervical cancer screening in a Spanish metropolitan area. J Womens Health 2010;19:1675-81.

178. Coughlin SS, Uhler RJ, Hall HI, Briss PA. Nonadherence to breast and cervical cancer screening: what are the linkages to chronic disease risk? Preventing Chronic Disease 2004;1:A04.

179. Miles A, Cockburn J, Smith RA, Wardle J. A perspective from countries using organized screening programs. Cancer 2004;101:1201-13.

180. Walsh B, Silles M, O'Neill C. The importance of socio-economic variables in cancer screening participation: a comparison between population-based and opportunistic screening in the EU-15. Health Policy 2011;101:269-76.

181. Palencia L, Espelt A, Rodriguez-Sanz M, Puigpinos R, Pons-Vigues M, Pasarin MI, Spadea T, Kunst AE, Borrell C. Socio-economic inequalities in breast and cervical cancer screening practices in Europe: influence of the type of screening program. Int J Epidemiol 2010;39:757-65.

182. The IPAQ Group. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ) - Short and long forms www.ipaq.ki.se/scoring.pdf, 2005.

183. World Cancer Research Fund, American Institute for Cancer Research, Second Expert Report, Food, Nutrition, Physical Activity, and the Prevention of Cancer: A Global Perspective Washington, DC: AICR, 2007.

184. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011;30:377-99.

185. von Hippel PT. Regression with missing Ys: An improved strategy for analyzing multiply imputed data. Sociological Methodology 2007;37:83-117.

186. Gail MH, Costantino JP, Bryant J, Croyle R, Freedman L, Helzlsouer K, Vogel V. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. J Natl Cancer Inst 1999;91:1829-46.

187. Wee CC, McCarthy EP, Davis RB, Phillips RS. Obesity and breast cancer screening. J Gen Intern Med 2004;19:324-31.

188. Fontaine KR, Heo M, Allison DB. Body weight and cancer screening among women. Journal of Women's Health & Gender-Based Medicine 2001;10:463-70.

189. Berz D, Sikov W, Colvin G, Weitzen S. 'Weighing in' on screening mammography. Breast Cancer Res Treat 2009;114:569-74.

190. van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, Fraser G, Goldbohm RA, Graham S, Kushi L, Marshall JR, Miller AB, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol 2000;152:514-27.

191. Hartmann LC, Sellers TA, Frost MH, Lingle WL, Degnim AC, Ghosh K, Vierkant RA, Maloney SD, Pankratz VS, Hillman DW, Suman VJ, Johnson J, et al. Benign breast disease and the risk of breast cancer. N Engl J Med 2005;353:229-37.

192. Arnold M, Razum O, Coebergh JW. Cancer risk diversity in non-western migrants to Europe: An overview of the literature. Eur J Cancer 2010;46:2647-59.

193. Wu Y, Zhang D, Kang S. Physical activity and risk of breast cancer: a meta-analysis of prospective studies. Breast Cancer Res Treat 2013;137:869-82.

194. Narod SA. Hormone replacement therapy and the risk of breast cancer. Nat Rev Clin Oncol 2011;8:669-76.

195. Meads C, Ahmed I, Riley RD. A systematic review of breast cancer incidence risk prediction models with meta-analysis of their performance. Breast Cancer Res Treat 2012;132:365-77.

196. Kopans DB, Smith RA, Duffy S. Mammographic Screening and "Overdiagnosis". Radiology 2011;260:616-20.

197. de Gelder R, Heijnsdijk EAM, van Ravesteyn NT, Fracheboud J, Draisma G, de Koning HJ. Interpreting Overdiagnosis Estimates in Population-based Mammography Screening. Epidemiol. Rev. 2011;33:111-21.

198. Biesheuvel C, Barratt A, Howard K, Houssami N, Irwig L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. Lancet Oncol. 2007;8:1129-38.

199. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2006:CD001877.

200. Zahl PH, Jorgensen KJ, Gotzsche PC. Lead time Models Should Not Be Used to Estimate Overdiagnosis in Cancer Screening. J Gen Intern Medicine 2014;29:1283-6.

201. Walter SD, Day NE. Estimation of the duration of a pre-clinical disease state using screening data. Am J Epidemiol 1983;118:865-86.

202. Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. Breast Cancer Res 2008;10:R41.

203. Spratt JA, von Fournier D, Spratt JS, Weber EE. Decelerating growth and human breast cancer. Cancer 1993;71:2013-9.

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

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

206. Zahl PH, Jorgensen KJ, Gotzsche PC. Overestimated lead times in cancer screening has led to substantial underestimation of overdiagnosis. Br J Cancer 2013;109:2014-9.

207. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2011:CD001877.

208. Duffy SW, Michalopoulos D, Sebuodegard S, Hofvind S. Trends in aggregate cancer incidence rates in relation to screening and possible overdiagnosis: a word of caution. J Med Screen 2014;21:24-9.

209. Lambe M, Wigertz A, Holmqvist M, Adolfsson J, Bardage C, Fornander T, Karlsson P, Odlind V, Persson I, Ahlgren J, Bergkvist L. Reductions in use of hormone replacement therapy: effects on Swedish breast cancer incidence trends only seen after several years. Breast Cancer Res Treat 2010;121:679-83.

210. Miller AB, Wall C, Baines CJ, Sun P, To T, Narod SA. Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial. BMJ 2014;348:g366.

211. McCann J, Treasure P, Duffy S. Modelling the impact of detecting and treating ductal carcinoma in-situ in a breast screening programme. J Med Screen 2004;11:117-25.

212. Gotzsche PC, Jorgensen KJ. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2013;6:CD001877.

213. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. Cochrane Database Syst Rev 2009:CD001877.

214. Miller AB, To T, Baines CJ, Wall C. Canadian National Breast Screening Study-2: 13-year results of a randomized trial in women aged 50-59 years. J Natl Cancer Inst 2000;92:1490-9.

215. Miller AB, To T, Baines CJ, Wall C. 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 2002;137:305-12.

216. Johns LE, Moss SM, Trial Management G. Randomized controlled trial of mammographic screening from age 40 ('Age' trial): patterns of screening attendance. J Med Screen 2010;17:37-43.

217. Lee S, Zelen M. A stochastic model for predicting the mortality of breast cancer. J Natl Cancer Inst Monogr 2006:79-86.

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

219. Zahl PH, Maehlen J. Overdiagnosis of breast cancer after 14 years of mammography screening. Tidsskr Nor Laegeforen 2012;132:414-7.

220. Kalager M, Bretthauer M. Spontaneous regression of invasive breast cancer: does this study answer the question? Arch Intern Med 2009;169:997.

221. Kalager M, Adami HO, Bretthauer M, Tamimi RM. Overdiagnosis of invasive breast cancer due to mammography screening: results from the Norwegian screening program. Ann Intern Med 2012;156:491-9.

222. Hellquist BN, Duffy SW, Nystrom L, Jonsson H. Overdiagnosis in the population-based service screening programme with mammography for women aged 40 to 49 years in Sweden. J Med Screen 2012;19:14-9.

223. Coldman A, Phillips N. Incidence of breast cancer and estimates of overdiagnosis after the initiation of a population-based mammography screening program. CMAJ 2013;185:E492-8.

224. Glasziou P, Houssami N. The evidence base for breast cancer screening. Preventive Medicine 2011;53:100-2.

225. LinkageWiz Software. LinkageWiz Data Matching Software Payneham, Australia: LinkageWiz Software, 2002.

226. Centre for Disease Control and Prevention. Registry Plus Link Plus CDC, Atlanta, GA, 2010.

227. Department of Health and Aged Care, Measuring Remoteness: Accessibility/Remoteness Index of Australia (ARIA), Revised Edition. Dept Health and Aged Care, 2001.

228. Efron B, Tibshirani RJ. An Introduction to Bootstraped.: Boca Raton: Chapman & Hall, 1993.

229. Beckmann KR, Roder DM, Hiller JE, Farshid G, Lynch JW. Do breast cancer risk factors differ among those who do and do not undertake mammography screening? J Med Screen 2013;20:208-19.

230. Australian Electoral Commission. Size of the electoral roll and estimated participation rate 2010. http://www.aec.gov.au/Enrolling_to_vote/Enrolment_stats/national/2010.htm, 2013.

231. Pfeiffer RM, Park Y, Kreimer AR, Lacey JV, Jr., Pee D, Greenlee RT, Buys SS, Hollenbeck A, Rosner B, Gail MH, Hartge P. Risk Prediction for Breast, Endometrial, and Ovarian Cancer in White Women Aged 50yrs or Older: Derivation and Validation from Population-Based Cohort Studies. PLoS Medicine 2013;10:e1001492.

232. Zahl PH. Overdiagnosis of invasive breast cancer due to mammography screening. Ann Intern Med 2012;157:220-1; author reply 1-2.

233. Giordano L, von Karsa L, Tomatis M, Majek O, de Wolf C, Lancucki L, Hofvind S, Nystrom L, Segnan N, Ponti A, Eunice Working G, Van Hal G, et al. Mammographic screening programmes in Europe: organization, coverage and participation. J Med Screen 2012;19 Suppl 1:72-82.

234. Iared W, Shigueoka DC, Torloni MR, Velloni FG, Ajzen SA, Atallah AN, Valente O. Comparative evaluation of digital mammography and film mammography: systematic review and meta-analysis. Sao Paulo Medical Journal 2011;129:250-60.

235. Australian Bureau of Statistics. Population Projections, Australia, 2012 (base) to 2101. Cat. no. 3222.0, Table C4, ed. 26/11/2013 Canberra: ABS, 2013.

236. Day NE. Estimating the sensitivity of a screening test. J Epidemiol Community Health 1985;39:364-6.

237. Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. Biometrics 1984;40:1-14.

238. Silvey SD. Statistical Inferenceed. Harmondsworth, England: Penguin Books Ltd., 1970.

239. Beckmann KR, Farshid G, Roder DM, Hiller JE, Lynch JW. Impact of hormone replacement therapy use on mammographic screening outcomes. Cancer Causes Control 2013;24:1417-26.

240. Beckmann KR, Lynch JW, Hiller JE, Farshid G, Houssami N, Duffy SW, Roder DM. A novel case-control design to estimate the extent of over-diagnosis of breast cancer due to organised population-based mammography screening. Int J Cancer 2014.

241. Beckmann K, Roder D, Hiller JE, Farshid G, Lynch J. The influence of mammographic screening on breast cancer incidence trends in South Australia. Asia Pac J Cancer Prev 2014;15:3105-12.

242. Chittleborough CR, Grant JF, Phillips PJ, Taylor AW. The increasing prevalence of diabetes in South Australia: the relationship with population ageing and obesity. Public Health 2007;121:92-9.

243. Taylor R, Boyages J. Estimating risk of breast cancer from population incidence affected by widespread mammographic screening. J Med Screen 2001;8:73-6.

244. Taylor AW, Dal Grande E, Wilson DH. The South Australian Health Omnibus Survey 15 years on: Has public health benefited? Public Health Bulletin 2006;Edition 3/2006:30-2.

245. Biller-Andorno N, Juni P. Abolishing mammography screening programs? A view from the Swiss Medical Board. N Engl J Med 2014;370:1965-7.

246. Marmot MG. Sorting through the arguments on breast screening. JAMA 2013;309:2553-4.

247. Gotzsche PC, Olsen O. Is screening for breast cancer with mammography justifiable? Lancet 2000;355:129-34.

248. Canadian Task Force on Preventive Health C, Tonelli M, Connor Gorber S, Joffres M, Dickinson J, Singh H, Lewin G, Birtwhistle R, Fitzpatrick-Lewis D, Hodgson N, Ciliska D, Gauld M, et al. Recommendations on screening for breast cancer in average-risk women aged 40-74 years. CMAJ 2011;183:1991-2001.

249. U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009;151:716-26, W-236.

250. Broeders M, Moss S, Nystrom L, Njor S, Jonsson H, Paap E, Massat N, Duffy S, Lynge E, Paci E, Group EW. The impact of mammographic screening on breast cancer mortality in Europe: a review of observational studies. J Med Screen 2012;19 Suppl 1:14-25.

251. Njor S, Nystrom L, Moss S, Paci E, Broeders M, Segnan N, Lynge E, Euroscreen Working G. Breast cancer mortality in mammographic screening in Europe: a review of incidence-based mortality studies. J Med Screen 2012;19 Suppl 1:33-41.

252. Nickson C, Mason KE, English DR, Kavanagh AM. Mammographic screening and breast cancer mortality: a case-control study and meta-analysis. Cancer Epidemiol Biomarkers Prev 2012;21:1479-88.

253. Jatoi I, Miller AB. Why is breast cancer mortality declining? Lancet Oncol 2003;4:251-4.

254. Burton RC, Bell RJ, Thiagarajah G, Stevenson C. 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 2012;131:949-55.

255. Gotzsche PC, Hartling OJ, Nielsen M, Brodersen J, Jorgensen KJ. Breast screening: the facts or maybe not. BMJ 2009;338:b86.

256. Zapka JG, Geller BM, Bulliard JL, Fracheboud J, Sancho-Garnier H, Ballard-Barbash R, Group ICW. Print information to inform decisions about mammography screening participation in 16 countries with population-based programs. Patient Educ Couns 2006;63:126-37.

257. Gotzsche PC, Jorgensen KJ. The Breast Screening Programme and misinforming the public. J. R. Soc. Med. 2011;104:361-9.

258. Giordano L, Cogo C, Patnick J, Paci E, Euroscreen Working G. Communicating the balance sheet in breast cancer screening. J Med Screen 2012;19 Suppl 1:67-71.

259. Paci E, Broeders M, Hofvind S, Puliti D, Duffy SW, the EWG. European Breast Cancer Service Screening Outcomes: A First Balance Sheet of the Benefits and Harms. Cancer Epidemiol Biomarkers Prev 2014;23:1159-63.

260. Paci E, Group EW. 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 2012;19 Suppl 1:5-13.

261. Pace LE, Keating NL. A systematic assessment of benefits and risks to guide breast cancer screening decisions. JAMA 2014;311:1327-35.

262. Welch HG, Passow HJ. Quantifying the benefits and harms of screening mammography. JAMA Internal Medicine 2014;174:448-54.

263. Hersch J, Jansen J, Barratt A, Irwig L, Houssami N, Howard K, Dhillon H, McCaffery K. Women's views on overdiagnosis in breast cancer screening: a qualitative study. BMJ 2013;346:f158.

264. Waller J, Douglas E, Whitaker KL, Wardle J. Women's responses to information about overdiagnosis in the UK breast cancer screening programme: a qualitative study. BMJ open 2013;3.

265. Jordan VC, Ford LG. Paradoxical clinical effect of estrogen on breast cancer risk: a "new" biology of estrogen-induced apoptosis. Cancer prevention research 2011;4:633-7.

266. Santen RJ. Menopausal hormone therapy and breast cancer. The Journal of Steroid Biochemistry and Molecular Biology 2014;142:52-61.

267. Meyerson AF, Lessing JN, Itakura K, Hylton NM, Wolverton DE, Joe BN, Esserman LJ, Hwang ES. Outcome of long term active surveillance for estrogen receptor-positive ductal carcinoma in-situ. Breast 2011;20:529-33.

268. Gierisch JM, Myers ER, Schmit KM, Crowley MJ, McCrory DC, Chatterjee R, Coeytaux RR, Kendrick A, Sanders GD. Prioritization of research addressing management strategies for ductal carcinoma in-situ. Ann Intern Med 2014;160:484-91.

269. Fallowfield L, Francis A, Catt S, Mackenzie M, Jenkins V. Time for a low-risk DCIS trial: harnessing public and patient involvement. Lancet Oncol 2012;13:1183-5.

270. Koka R, Ioffe OB. Breast carcinoma: is molecular evaluation a necessary part of current pathological analysis? Seminars in Diagnostic Pathology 2013;30:321-8.

# APPENDIX

**Table A1 South Australian Health Omnibus Survey Questions (2012)**

Breast cancer risk factor questions asked in the 2012 HOS pertaining to data used in Chapter 3

Survey questions were administered face to face by trained survey administrators

| **BREAST SCREENING** |
|---|
| *If male Go to L1*<br>*If female and 15-39 years of age Go to L1*<br>**We would now like to ask you some questions about womens' heath issues.**<br>**K1 Have you ever received a letter of invitation to be screened for breast cancer from BreastScreen?** *For a routine screen at BreastScreen only, not for a follow-up x-ray if there has been a problem or symptom.*<br>1 Yes<br>2 No<br>3 Unsure/Can't say |
| **A mammogram is an x-ray which is taken by a machine that presses against the breast. A screening mammogram is a routine check for breast cancer when you had no specific problems or symptoms.**<br>**K2 Have you ever had a screening mammogram at a BreastScreen clinic or in the BreastScreen mobile van?** *For a routine screen at BreastScreen only, not for a follow-up x-ray if there has been a problem or symptom.*<br>1 Yes<br>2 No<br>3 Unsure/Can't say *Go to K4* |
| **K3 How long ago did you have your last mammogram at breast screen?**<br>☐☐ *Enter years or D for Don't know.*<br>*Enter '1' if less than one year ago* |
| **K4 Have you ever had a screening mammogram at a private clinic or other facility that was NOT run by BreastScreen? Again this only refers to routine checks when you had no specific problems or symptoms.**<br>1 Yes<br>2 No<br>3 Unsure/Can't say |

*Table A1 (continued) 2012 HOS questions*

| | |
|---|---|
| **K5** | **Have you ever been diagnosed with breast cancer?**<br>1 Yes<br>2 No<br>3 Unsure/Can't say |
| **K6** | **Have you ever had a breast biopsy or other surgical procedure involving the breast, for any reason?**<br>1 Yes<br>2 No<br>3 Unsure/Can't say |
| **K7** | **How many of your female blood relatives have been diagnosed with breast cancer? (This includes your mother, and any sisters, daughters or any aunties who are genetically related to you)**<br>☐☐ *Enter number or D for Don't know.*<br>*If None or Don't know Go to K9* |
| **K8** | **Were any of these relatives less than 50 years old when they were diagnosed with breast cancer?**<br>1 Yes<br>2 No<br>3 Unsure/Can't say |
| **K9** | **Have you ever used hormone replacement therapy?** *Doctor prescribed HRT only. Do not include herbal/alternative/complimentary.*<br>1 Yes<br>2 No *Go to K13*<br>3 Unsure *Go to K13* |
| **K10** | **How many years (in total) have/did you use hormone replacement therapy for?**<br>☐☐ *Enter years or D for Don't know.*<br>*Enter '1' if less than one year* |
| **K11** | **How old were you when you last used hormone replacement therapy?**<br>1 Enter age ☐☐<br>2 Still using HRT<br>3 Don't know/can't say |

*Table A1 continued*

**K12 What was the main type of hormone therapy that you have used?** *Prompt for type used for longest period of time. If not sure which was main but both estrogen and progestin/progesterone are mentioned record estrogen and progesterone combined*

1 Oestrogen only
2 Estrogen And Progesterone combined
3 Other type (specify)……..
4 Don't Known /Can't Say/ Refused

**K13 How many children have you given birth to?**

☐☐ *Enter number or R for Refused*

*If None or Refused skip to K15*

**K14 How old were you when you first gave birth?**

☐☐ *Enter age or R for Refused*

**K15 At what age did you first start your period?**

☐☐ *Enter age or R for Refused/Don't know or Can't say*

**K16 At what age did you reach menopause (ie stop having periods)?**

1 Enter age ☐☐
2 Not reached menopause
3 Don't know/can't say

**WEIGHT/HEIGHT**

**Changing the subject.**

**F1 What is your height without shoes?**

☐☐☐ Centimetres *(OR)*

☐:☐☐ Feet/Inches

☐ (D)Don't' know

**F2 What is your weight (undressed in the morning)?**

☐☐☐ Kilograms *(OR)*

☐☐:☐☐ Stones/Pounds

☐ (D) Don't know

*Table A1 continued*

**ALCOHOL**

**Changing the subject. These next questions may seem similar to previous questions, however some of the number of standard drinks differ slightly on the show card.**

**P1 How often do you usually drink alcohol?**

1 Enter days per week (maximum =7) ☐
2 Don't drink alcohol *Go to P3*
3 Less than once a week

**This next question may seem similar to a previous question, however some of the number of standard drinks differ slightly on the show card.**

**P2 A standard drink is equivalent to one glass of full-strength beer, one glass of wine or one nip of spirits. On a day that you drink, how many standard drinks do you usually have?** *Show prompt card P1*

.☐☐.☐ *Enter number of drinks or (R) for refused*

**PHYSICAL ACTIVITY**

**G1 In an average week, how much time per week would you spend doing strenuous physical activity (that makes you sweat and breath hard)**

1 Enter minutes ☐☐
2 Enter hours ☐☐
3 Don't know/can't say
4 Refused

*Table A1 continued*

| | |
|---|---|
| **G2** | **How much time per week would you spend doing moderate physical activity (eg swimming, cycling).**<br>1 Enter minutes ☐☐<br>2 Enter hours ☐☐<br>3 Don't know/can't say<br>4 Refused |
| **G3** | **How much time per week would you spend walking for exercise, pleasure or to get from place to place?**<br>1 Enter minutes ☐☐<br>2 Enter hours ☐☐<br>3 Don't know/can't say<br>4 Refused |