This is the published version:


Available from Deakin Research Online:

http://hdl.handle.net/10536/DRO/DU:30046716

Reproduced with the kind permission of the copyright owner.

Copyright: 2009, Australian Government, Department of Health and Aging
BreastScreen Australia Evaluation

A Review of Methodological Options for Evaluating the Effect of BreastScreen Australia on Breast Cancer Mortality

December 2006
FOREWORD

This report provides evidence-based recommendations for appropriate and cost-effective methods that could be used to evaluate the impact of the national BreastScreen Australia population-based mammographic screening program on mortality from female breast cancer. The report represents a significant collaboration between the Australian Government, the National Breast Cancer Centre as well as Australian and international experts in mammography research and evaluation, epidemiology and health services research.

The recommendations are based on a review of national and international evidence on approaches used to assess the impact of mammography screening programs on breast cancer mortality in other settings. The review has used a systematic approach to assessing the strategic and methodological approaches taken in each of the studies identified and their potential limitations.

The national evaluation of the BreastScreen Australia Program aims to assess the appropriateness, efficiency and effectiveness of the BreastScreen Program. The completion of this report marks an important first step in that process. In addition, the review and recommendations in this report may have broader application at an international level.

Professor Dallas English
Chair, Project Steering Committee
Professor of Epidemiology & Biostatistics
The University of Melbourne

Dr Helen Zorbas
Director
National Breast Cancer Centre
# CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreword</td>
<td>i</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iv</td>
</tr>
<tr>
<td>Executive summary</td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>5</td>
</tr>
<tr>
<td>Objectives</td>
<td>7</td>
</tr>
<tr>
<td>Methodology</td>
<td>8</td>
</tr>
<tr>
<td>Literature search</td>
<td>8</td>
</tr>
<tr>
<td>Literature review</td>
<td>12</td>
</tr>
<tr>
<td>Developing evaluation design options</td>
<td>12</td>
</tr>
<tr>
<td>Rating the features of the design options</td>
<td>13</td>
</tr>
<tr>
<td>Ranking evaluation design options</td>
<td>13</td>
</tr>
<tr>
<td>Results of the literature search</td>
<td>14</td>
</tr>
<tr>
<td>Implications of inferring causality from epidemiological evidence</td>
<td>15</td>
</tr>
<tr>
<td>Causality</td>
<td>15</td>
</tr>
<tr>
<td>Epidemiological investigations of causality</td>
<td>16</td>
</tr>
<tr>
<td>Challenges for assessing causality in breast-cancer screening</td>
<td>20</td>
</tr>
<tr>
<td>Describing and ranking design options</td>
<td>23</td>
</tr>
<tr>
<td>Considering design options</td>
<td>23</td>
</tr>
<tr>
<td>Aims and process outlines by design option</td>
<td>27</td>
</tr>
<tr>
<td>Initial rating of design options</td>
<td>38</td>
</tr>
<tr>
<td>Final synthesis of merits of optional designs</td>
<td>50</td>
</tr>
<tr>
<td>Proposals</td>
<td>55</td>
</tr>
<tr>
<td>Proposed study designs for evaluating BreastScreen</td>
<td>55</td>
</tr>
<tr>
<td>Proposed ecological study</td>
<td>55</td>
</tr>
<tr>
<td>Proposed case-control study with mortality endpoint</td>
<td>58</td>
</tr>
<tr>
<td>Proposed risk-factor survey</td>
<td>62</td>
</tr>
</tbody>
</table>
A review of methodological options for evaluating the impact of BreastScreen Australia on breast cancer mortality

December 2006

Issues not addressed through recommended designs 64
Recommendations 65
References 66
Included publications 78
Appendices 72
Appendix A-1: Terms of reference of the Steering Committee 72
Appendix A-2: Terms of reference of the Expert Technical Group 72
Appendix A-3: Stated potential conflicts of interest of Steering Committee and Expert Technical Group members 73
Appendix A-4: International experts contacted for the review 74
Appendix A-5: Data extraction form 75
Appendix A-6: Prompters of selected study designs, outcome options and types of bias 77
Appendix A-7: Possible study designs for evaluating the effect of BreastScreen on breast cancer mortality 80
Appendix A-8: Overview of national data sources 81
Appendix A-9: Study design ranking form 85
Appendix A-10: Data extraction forms for included publications 86
Appendix A-11: Ranking of study designs by Expert Technical Group 150
Appendix B-1: Methodological publications relating to evaluating the impact of mammography service screening upon population mortality 151
Appendix B-2: Publications modelling impact of mammography service screening upon population mortality 156
Appendix B-3: Publications reporting cancer survival but not population mortality 157

Tables
Table 1: Results of literature search 9
Table 2: Number of included studies 10
Table 3: Reports obtained from Cochrane Library, HTA agencies and ‘grey’ literature 11
Table 4: Summary of design criteria for each considered design option 51
ACKNOWLEDGEMENTS

The National Breast Cancer Centre gratefully acknowledges the work of the members of the Steering Committee and Expert Technical Group who contributed enormously to the development of this report.

FUNDING

Funding for the development of this report was provided by the Australian Government Department of Health and Ageing.

STEERING COMMITTEE

Professor Dallas English (Chair)
Professor Robert Burton
Professor David Roder (co-opted)
Professor Richard Taylor
Dr Helen Zorbas
Expert Technical Group
Professor David Roder (Chair)
Professor Adrian Esterman
Associate Professor Dorota Gertig
Professor John Mathews
Dr Chris Stevenson

LITERATURE REVIEW TEAM

The literature review and data extraction for this report were conducted by Dr Adele R Weston, Health Technology Analysts Pty Ltd.

National Breast Cancer Centre Staff
Ms Fleur Webster
Dr Alison Evans
Dr Elmer Villanueva
Ms Holly Goodwin
Ms Rosemary Vagg
EXECUTIVE SUMMARY

OBJECTIVE

The aim of the Project was to review and assess the design options for evaluating the impact of the BreastScreen Australia Program (hereinafter, the ‘Program’) on breast-cancer mortality in the population, with due regard for feasibility and quality of evidence. This included effects on:

- all women eligible for screening,
- women offered screening, and
- women screened

when in the target age range of 50–69 years.

PROCESS

The Steering Committee, Expert Technical Group (ETG) and Literature Review Team established a context for project deliberations by reviewing the principles for inferring causality from epidemiological evidence. After scanning a total of 3,928 papers that met search criteria, 59 methodological papers on breast screening evaluation and a further 60 reports of screening program evaluation studies were identified from the international peer-reviewed literature. A group of international experts was consulted to nominate any papers that had been omitted from this list and any significant studies that may not have appeared in the peer-reviewed literature. Based on these papers, members of the ETG developed nine potential design options for studying potential effects of the BreastScreen Australia Program on mortality from breast cancer.

The ETG agreed on a set of merit criteria for a mortality study and used these criteria to derive an initial ranking of the potential design options. The criteria related to quality of scientific evidence as well as feasibility issues. The ETG discussed each option in detail, including the capacity of each design option to answer the principal evaluation questions, and developed a draft set of recommendations for the proposed study. These recommendations were considered and reviewed in detail at a combined meeting of the project Steering Committee and the ETG, where consensus was reached on a final set of recommendations.
CONSIDERATION OF DESIGN OPTIONS

Based on the literature review, it was evident that a sophisticated evaluation methodology would be necessary, particularly as the period since implementation of BreastScreen Australia had been relatively short and any effects of screening on mortality could be potentially confounded by changes in treatment and risk factors over the same period. A randomised controlled trial, the gold standard design for evaluating health interventions, was clearly not possible or ethical for the evaluation of BreastScreen, given the trial evidence already available and the advanced development of the BreastScreen Australia Program. The selected study design would have to be observational only, and careful consideration would need to be given to minimising the potential for bias from sources such as lead and length bias, and sample selection, to which these types of studies were often susceptible.

The ETG considered a number of design features in compiling a list of potential design options. They included:

- design types: including ecological, case-control, cohort, and case-cohort designs, and simulation studies
- study endpoints: including mortality, such as breast-cancer mortality, all-causes mortality and excess mortality, and mortality surrogate measures, such as tumour size, nodal status and grade at diagnosis, and progression to distant metastases
- sources of bias: including bias relating to subject definition, lead time, length bias, control selection, screening participation, length of follow-up time, effects of healthy screening participation, recall effects, cross-level bias occurring when imputing individual effects from population studies or the reverse, and misclassification of diagnostic and screening mammograms
- factors attenuating measured effects: including inadequate follow-up time, poor definition of the screened and unscreened, and inclusion of deaths from cancer diagnosed prior to screening commencement
- statistical issues: including choice of analytical method, meeting underlying statistical assumptions, and allowing for clustering.

It is important to note that none of the individual study design options considered by the ETG was completely free of potential for bias, as they were all non-experimental. However, the ETG believes that the design options chosen, when taken together, would provide the best estimate of effects of the BreastScreen Australia Program on breast-cancer mortality.
RECOMMENDATIONS

It is recommended that:

1. The effects of the BreastScreen Australia Program on mortality should be assessed in two separate studies: one that assesses the effect of the Program on sub-populations within Australia, and one that assesses the effect at the level of individual women who participated in, or were invited to BreastScreen.

2. The study of sub-populations should assess the association of BreastScreen participation with breast-cancer mortality by Statistical Local Area (SLA). The budget cost for this ecological study is estimated to be between $200,000 and $300,000, depending on the operational complexities and extent of inclusion of nested sub-studies of greater depth.

3. The study of effects on individual women would compare exposure to BreastScreen in women who died of breast cancer with exposure to BreastScreen in women who had not died of this disease. Two analyses would be undertaken in this case-control study, the first defining BreastScreen exposure to include receiving an invitation to BreastScreen screening, and the second restricting this definition to BreastScreen screening participation. The budget cost is estimated to be between $550,000 and $650,000, depending on the operational complexities and extent of inclusion of nested sub-studies of greater depth to support the main study.

4. Concurrently with the case-control study, a population survey should assess how the exposure of women to BreastScreen has been related to (‘diagnostic’) screening mammograms outside the Program and how differences in BreastScreen exposure are related to breast-cancer risk factors, medical care and other social and demographic factors. Because exposure to risk factors may vary over time, the aim would be to assess past as well as current exposures. Depending on the depth of survey questioning, the cost is estimated to be between $80,000 and $110,000.

Although outside the scope of this project, the results of all three studies could be used in statistical modeling to investigate whether any associations of BreastScreen with breast-cancer mortality have been modified by social, demographic, familial and behavioural risk factors, and by health care outside the BreastScreen system. The results would be of great value in planning how to increase the future impact of breast screening in Australia.

The proposed ecological study would be appropriate for evaluating the BreastScreen Program, because the associated group-level analyses would be directed at program outcomes. The primary interest in this project is at a program level, not at biological or behavioural factors at an individual level. Because an ecological study would be relatively easy to implement, the cost would be comparatively low, and the study could be completed within a short time frame. Data required for the study would be readily available and there is proof of concept, in that similar studies have been successfully implemented in NSW and overseas. While ecological studies have been criticised as being subject to bias from the ‘ecological fallacy’ when used to infer effects on individuals, this would not be the purpose in the present context. It is also proposed that the ecological study be complemented with a case-control study, where this potential bias would not apply, and which would more directly measure effects of BreastScreen on mortality among those individuals exposed to the BreastScreen Program.
The ETG believes that a well-designed case-control study would produce findings as valid as a retrospective cohort study, and in a much shorter period of time. Australia is well placed to undertake a valid case-control study, given the availability of near-to-complete registry data and screening records, and the opportunities that exist to link these records. While case-control designs can be vulnerable to recall and selection bias in many settings, these dangers can be minimised in Australia through the use of death registers, cancer registries, and population listings for gaining representative samples, and by linkage of these databases to screening records to avoid recall bias of screening exposure.

It is proposed that the endpoint for the case-control study be mortality from breast cancer. The ETG discussed the use of surrogate measures as endpoints, as these would occur ahead of deaths and would allow more timely results. Also, study participants would be available for interview on exposure to screening outside the Program, and on breast-cancer risk factors, to investigate potential confounding. Nonetheless, mortality was selected as the appropriate endpoint, in view of the uncertain link between some surrogate measures and breast-cancer mortality, and the possible biases introduced by using them.

Because mortality has been selected as the endpoint, deceased women in the study would not be able to provide self-reported data on any exposure to screening outside the Program and on breast-cancer risk factors. This would lead to bias, if there were non-comparability between cases and controls that was not addressed. Hence the ETG recommends a concurrent population survey to investigate exposure to de facto screening outside BreastScreen, and to breast-cancer risk factors. Taken together with existing data on the strength of the associations between the risk factors and breast-cancer mortality, this would allow a sensitivity analysis of the possible extent of confounding by performing external adjustment of odds ratios.

The selected studies would indicate whether exposure to BreastScreen is associated with a lower risk of breast-cancer mortality, and also the strength of such an association.

There are questions outside the scope of this project that would require separate study. For example, should exposure to BreastScreen be found to be associated with lower risk of breast-cancer mortality, further research would be needed to explore the extent to which this was due to BreastScreen screening per se, as compared with possible improvements in clinical care or other factors associated with BreastScreen implementation.
INTRODUCTION

Evidence indicates that the impact of a national mammographic screening program on deaths from breast cancer will not be evident for some years – possibly more than 10 years – following screening implementation. A priority is to develop a national strategy by which this impact can be measured when it is feasible to do so.

During the 1970s and 1980s, the results of a number of randomised trials were published which demonstrated that mammographic screening could reduce deaths from breast cancer among women aged 50–69 years. First-round participation rates in overseas randomised trials of mammographic screening ranged between 61% and 100% (IARC, 2002). In these trials, reductions in mortality of up to 30% were achieved among those populations offered screening, with larger reductions estimated among those being screened. There is also indicative evidence from the trials of benefits in the 40–49-year age range, although this evidence is much weaker and the benefits appear smaller and to emerge later.

In response to the results of these trials, it was decided in the 1980s to explore the feasibility of a national mammographic screening program in Australia. The program was piloted in 11 sites over a three-year period during 1987–1990, leading to the establishment of the BreastScreen Australia Program (hereinafter, the ‘Program’) in 1991. Principally, BreastScreen aims to reduce mortality and morbidity from breast cancer among women aged 50–69 years, although women 40 years of age and over are eligible to attend. Based on the evidence from the trials, the objective is to screen at least 70% of women aged 50–69 years every two years.

In Australia, the national participation rate for women aged 50–69 years in the two year period of 2002-2003 was 56.1% (AIHW and NBCC, 2006). BreastScreen monitors its performance using structure, process and intermediary outcome standards based on evidence from the original trials, international expert reports, additional research, and consensus judgements of Australian experts. The rationale is that results should be at least as favourable as trial results if mammography screening is performed in a manner, and at a standard, equivalent to or better than in the trials.

Jointly funded by the Australian and State and Territory governments, the BreastScreen Program operates within a national policy framework coordinated by the Australian Population Health Development Principal Committee, with State- and Territory-based policy and service implementation. Within each State and Territory, a State Coordination Unit has responsibility for implementation of the Program. State- and Territory-based Screening and Assessment Services have been established, covering rural, urban, and public and private health sectors. Approximately 34 Services are operating across Australia, each of them screening between 1,000 and 65,000 women per year.

This study is part of a larger evaluation of BreastScreen Australia, which aims to assess the appropriateness, efficiency and effectiveness of the BreastScreen Program. This larger evaluation will also assess and address ongoing and emerging issues that have an impact on the Program, and identify opportunities for overall improvement. To achieve its objectives, the evaluation will cover a range of projects aimed at assessing the health effects, processes and economic outcomes delivered by the Program. This will entail assessing benefits and risks associated with screening, including the impact of the Program on mortality.
and morbidity due to breast cancer, efficiency of the delivery model, trends in performance indicators, infrastructure and capacity, and economic outcomes. The impact of the Program on breast-cancer mortality is central to the larger evaluation, which is due for completion in October 2008.

The present the project has been overseen by a Steering Committee and implemented by an Expert Technical Group (ETG) in collaboration with an experienced Literature Review Team. Steering Committee members were selected to provide expert opinion and guidance on the directions and outcomes of the project. The terms of reference for the Steering Committee are outlined in Appendix A-1. The ETG comprised Australian leaders experienced in mammography screening research and evaluation, and with broader expertise in epidemiology and population health research. The terms of reference for the ETG are outlined in Appendix A-2. In addition, a literature review team with an extensive track record in systematic literature reviews in relation to population health research was commissioned to undertake the literature review.

At the commencement of the project, members of the Steering Committee and the ETG were asked to declare any potential conflicts of interest. A number of members declared such conflicts (Appendix A-3). However, members agreed that although it was prudent to record apparent conflicts of interest for the public record, none of these conflicts warranted disqualification from participating in the project.

Initially, basic principles for inferring causality from epidemiological evidence were reviewed, as a context for considering strategic options for the evaluation of the BreastScreen Australia program. National and international evidence on approaches used to assess effects of mammography screening programs on breast-cancer mortality in other settings was then assessed. While it is important to evaluate effects on breast-cancer mortality directly, it was evident that study designs by necessity would be observational and results of any single study could only be regarded as indicative.

The possible sources of bias in observational designs are greater than for randomised trials and careful attention must be paid in their design to reducing potential for bias. The results of observational studies of screening programs should be interpreted with care and ideally confirmed through multiple investigations.

In combination with other studies around the world, an Australian study could make an important contribution to the evidence available of effects of mammography screening on mortality outcomes.
OBJECTIVES

The purpose of the project was to develop and rank design options for evaluating effects of BreastScreen Australia on mortality from female breast cancer. The project set out to address respectively the effects on all women eligible for screening, those offered screening, and those screened when in the target age range of 50–69 years.

An objective was to determine the best outcome measure. While mortality from breast cancer ideally would be measured directly, indirect measures were considered, including use of surrogate markers such as tumour size, grade and nodal status at diagnosis, and the occurrence of distant metastases. Alternatively, effects on breast-cancer mortality may be inferred from effects on mortality from all causes combined, or from the excess mortality experienced by women with a breast-cancer diagnosis.

The main focus was on screening effects on all women in the target age range, which determined the scope and design of the recommended evaluation. A secondary aim was to explore differences in effect across population sub-groups, including sub-groups classified by service access, socio-economic status, and metropolitan, regional urban or rural place of residence, and if feasible, by exposure to individual risk factors associated with breast cancer, such as family history, genetic predisposition, reproductive history, exposure to hormone replacement therapy, and history of benign breast disease.

The objective was to review, for the period since 1980: (1) the peer-reviewed literature on screening mammography evaluation methodology and design aspects of individual mammography screening program evaluations (as opposed to trials); and (2) additional reports on these topics in the 'grey' literature recommended by a panel of international experts. This review informed the development and ranking of design options.

The aim of the project was to rank design options by merit according to likely value of evidence, ease of implementation, risk of unsuccessful implementation, comparative cost, and time required for study completion.
METHODODOLOGY

LITERATURE SEARCH

To ensure that all relevant studies were identified, a comprehensive search of the published literature was conducted. Multiple approaches were used to search for relevant data, namely:

1. a search of published literature using the electronic database EMBASE.com, which includes EMBASE and MEDLINE, employing terms for breast cancer, mammography screening and mortality (see below)

2. a search of published literature using the electronic database EMBASE.com, which includes EMBASE and MEDLINE, employing author details for 167 known mammography screening researchers (see below)

3. a search of the Cochrane Library

4. a review of bibliographies of included studies (‘snowballing’)

5. a request to international researchers for identification of additional relevant studies

6. an examination of references provided in study reports included an existing systematic review to ensure all relevant ones were captured.

A summary of the literature search strategies that were employed is presented in Table 1. MEDLINE and EMBASE were searched concurrently using EMBASE.com. The search was conducted in August 2006. The search was limited to papers published after 1980, because the routine use of mammography screening as a community service program was limited before that time. The search was restricted to publications in the English language.

After removing duplicates, a total of 3,928 unique citations remained. Titles and abstracts of these 3,928 publications were reviewed in accordance with the exclusion criteria listed below. A random sample of excluded citations, comprising 10% of all those excluded, was checked by a member of the ETG as a quality control procedure. Following this initial exclusion process, 285 publications were retrieved as full-length papers and reviewed in detail to determine final inclusion or exclusion.
Table 1: Results of literature search

<table>
<thead>
<tr>
<th>EMBASE.com (includes Medline and EMBASE 1980 to 01 August 2006)</th>
<th>#1</th>
<th>3,109</th>
</tr>
</thead>
<tbody>
<tr>
<td>#2</td>
<td>1,323</td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td>Citations included in bibliographies provided by Expert Technical Group and snowballing</td>
<td>54</td>
</tr>
<tr>
<td>#4</td>
<td>Citations included in bibliographies provided by overseas experts</td>
<td>32</td>
</tr>
</tbody>
</table>

Total unique citations: 3,928
The exclusion criteria are listed below. They were not applied in any particular order. It is possible that a single publication would have met more than one exclusion criterion, but only one was noted. The criteria were:

1. not a paper reporting the evaluation of mammography as a service screening program
2. unsuitable outcomes (i.e., not population mortality or a surrogate marker directly modelled to it)
3. narrative reviews/opinions only
4. study design not suitable for retrospective evaluation of an existing program (i.e., prospective randomised controlled trial)
5. patient information material only
6. unsuitable population, unsuitable cancer, or unsuitable intervention
7. exclusion by ETG, as described below.

The outcomes from applying these criteria to the results of the literature search are summarised in Table 2. Members of the ETG then culled papers that were selected by the Literature Review Team. Excluded studies, complete with their reason for exclusion, are enumerated in Volume 2 of this report.

**Table 2:** Number of included studies

<table>
<thead>
<tr>
<th>Description</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full length papers reviewed</td>
<td>285</td>
</tr>
<tr>
<td>Excluded on basis of reasons listed above (Volume 2), together with those</td>
<td>134</td>
</tr>
<tr>
<td>excluded on basis of title or abstract</td>
<td></td>
</tr>
<tr>
<td>Methodological publications (Appendix B-1)</td>
<td>59a</td>
</tr>
<tr>
<td>Publications modelling (not measuring) the impact of mammography service</td>
<td>10</td>
</tr>
<tr>
<td>screening upon population mortality (Appendix B-2)</td>
<td></td>
</tr>
<tr>
<td>Publications reporting impact of mammography service screening upon survival</td>
<td>25</td>
</tr>
<tr>
<td>but not population mortality (Appendix B-3)</td>
<td></td>
</tr>
<tr>
<td>&quot;Included&quot; studies reporting impact of mammography service screening upon</td>
<td>60b</td>
</tr>
<tr>
<td>population mortality (see References: Included studies)</td>
<td></td>
</tr>
</tbody>
</table>

a Three methodological papers were also included amongst those reporting population mortality (Baker et al, 2004; Chen et al, 1998; Hakama et al, 1999) and one was included amongst those reporting survival (Michaelson et al, 2003)
b Not including systematic review (Gabe & Duffy, 2005)

The present review then focused upon study designs that were regarded as having the ability to evaluate the impact of mammography screening upon population breast-cancer mortality (or all-causes mortality). While the review was not designed to include studies that only reported the impact of mammography screening on cancer survival, because such studies were regarded as particularly vulnerable to length bias and lead-time bias, some such publications were identified and were also made available to the ETG. In addition, papers that were purely methodological but potentially relevant to the study design in question were identified. This range of papers informed deliberations of the ETG with respect to potentially useful study designs and methodological considerations.
A recent systematic review was identified that specifically reported the evaluation of effects of mammography screening programs on mortality (Gabe and Duffy, 2005). All publications included in that review were also considered for inclusion in the present review. In addition to the published literature indexed in EMBASE and MEDLINE, a search of the Cochrane Library, Health Technology Assessment agencies and other ‘grey’ literature was performed. A total of 14 reports were obtained through these means, including three Cochrane systematic reviews (Table 3). However, these reports primarily fell into two categories: i) reports of process and performance indicators of regional or national mammography screening programs, eg, invitation rates, participation rates, diagnostic performance, and detection rates; and ii) systematic reviews of randomised controlled trials. None of these reports provided information on study designs for evaluating the impact of mammography on mortality that was not reported in the indexed published literature above. Nonetheless, they were made available to inform deliberations of the ETG.

Table 3: Reports obtained from Cochrane Library, HTA agencies and ‘grey’ literature

<table>
<thead>
<tr>
<th>Report title</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>The benefit of population screening for breast cancer with mammography (2002)</td>
<td>Health Council of the Netherlands</td>
</tr>
<tr>
<td>Interventions for relieving the pain and discomfort of screening mammography (2002)</td>
<td>Cochrane Database of Systematic Reviews</td>
</tr>
<tr>
<td>Public health focus: mammography (1992)</td>
<td>National Centre for Chronic Disease Prevention and Health Promotion (USA)</td>
</tr>
<tr>
<td>Report from the evaluation indicators working group: guidelines for monitoring breast screening program performance</td>
<td>Health Canada</td>
</tr>
<tr>
<td>Review of the evidence about the value of mammographic screening in 40-49 year old women (1997)</td>
<td>NHMRC National Breast Cancer Centre</td>
</tr>
<tr>
<td>Screening for breast cancer with mammography (2001)</td>
<td>Cochrane Database of Systematic Reviews</td>
</tr>
<tr>
<td>Strategies for increasing the participation of women in breast cancer screening (2001)</td>
<td>Cochrane Database of Systematic Reviews</td>
</tr>
</tbody>
</table>

In addition, a sample of international experts (Appendix A-4) with backgrounds in mammography screening evaluation was contacted and asked to review the list of included publications and cite any additional ones of relevance known to them in the peer-reviewed or ‘grey’ literature that had been omitted. In addition, they were canvassed about current or planned projects in the area of mammography evaluation that had not yet been reported.
LITERATURE REVIEW

A standardised data extraction form was used to record a synopsis of study design features for each of the 'included' publications. This form was designed by members of the ETG and Literature Review Team (see Appendix A-5).

The form was fine-tuned by the Literature Review Team after a pilot run on 10 publications. Only information relating to the process of assessing the mammography screening effect on mortality was extracted (i.e., information relating to the actual screening outcomes or other aspects not relevant to the primary objective of the review was not extracted).

In addition to information relating to the study design and analyses, the Literature Review Team performed an initial assessment of the strengths and weaknesses of the study design and analyses. These strengths and weaknesses were then considered further and in more detail by the ETG.

In several instances, a research group had published the same methodology several times (e.g., in response to a longer follow-up of patients and for updating results over time). In these instances, all publications remained within the 'included' group of publications, although information on the design and analyses was extracted for the most recent publication only. On the other hand, where the same research group had subsequently used a different methodology, a separate data extraction process was performed for each publication.

DEVELOPING EVALUATION DESIGN OPTIONS

Members of the ETG developed a list of study design options for evaluating BreastScreen effects on mortality, drawing on designs described in the initial methodology papers, on the data extraction forms, in copies of full reports of evaluation studies provided by the Literature Review Team, and in reports cited by the international experts.

A summary of key study designs, possible outcome measures, and possible threats to validity were circulated to ETG members to assist this process (Appendix A-6). Design types covered theoretical simulation models, ecological study designs, case series, case-control studies, cohort designs, and hybrid cohort/case-control designs. Common biases were outlined, including those associated with lead-time, length-time effects, over-diagnosis, self-selection, healthy screening participation effect, attenuation through case/control crossover, inadequate duration of follow-up, and misclassification of screening and diagnostic mammograms.

The process of developing design options involved exchanging lists of options and critical comments about each option by email (Appendix A-7). For each option, statements were provided concerning the study aim, the implementation process outline, and perceived advantages, disadvantages and risks. Likely data availability was considered in developing design options, using the BreastScreen Manual of Data Items and Dictionary Definitions, plus an overview of national data sources (Appendix A-8). A teleconference was held at the end of this process to finalise the list of possible designs.
RATING THE FEATURES OF THE DESIGN OPTIONS

Members of the ETG rated individual design options independently, indicating their assessment on a five-point ordinal scale for five features of design and feasibility, namely: likely (1) value of evidence; (2) ease of implementation; (3) risk of unsuccessful implementation; (4) cost; and (5) time to study completion. A standard scoring sheet was employed to rank the five features of each design option independently, with space provided alongside each feature to record narrative comments (Appendix A-9).

Value of evidence was assessed by considering the potential for bias associated with recall, lead time, length bias, over-diagnosis, self selection, the healthy screening participation effect, attenuation through case/control crossover, inadequate duration of follow-up, misclassification of screening and diagnostic mammograms, and other factors; the potential for confounding; the strength and appropriateness of statistical analyses; and the statistical power available. Ease of implementation was assessed by considering: likely data availability and access; likely obstacles in gaining human research ethics committee approvals; and the potential for administrative barriers.

Differences between members in rating scores of two or more were discussed by teleconference and reviewed, prior to aggregation of these ratings into a single ETG mean rating. As a result, each design option ended up with five ETG mean ratings, one for each design feature.

RANKING EVALUATION DESIGN OPTIONS

A face-to-face meeting of ETG members was scheduled to rank overall design options by merit. Initially this entailed averaging the mean ratings for features of each design option to gain an aggregate score for that option. This was undertaken, giving a weighting of 0.4 to value of evidence, 0.3 to time to complete study, 0.1 to ease of implementation, 0.1 to risk of unsuccessful implementation, and 0.1 to cost. This enabled an initial rank ordering of design options by merit, which then became the basis for subsequent discussion, debate, and finalisation.
RESULTS OF THE LITERATURE SEARCH

In total 60 original publications were identified that evaluated the impact of mammography screening on population breast-cancer mortality using a study design that was potentially useful for the evaluation of an existing service screening program. They included all relevant publications listed in the recent systematic review by Gabe and Duffy (2005). The literature review also identified a further 59 methodological papers on breast-screening evaluation.

The data extraction forms for all ‘included’ publications are presented in alphabetical order in Appendix A-10. Where a publication was superseded by a more recent publication from the same research group with a similar methodology, this has been noted.
IMPLICATIONS OF INFERRING CAUSALITY FROM EPIDEMIOLOGICAL EVIDENCE

CAUSALITY

Causality research seeks to determine whether a causal relationship exists, for example, between an intervention and an effect, or between a risk factor and a disease. In the evaluation of mammography screening, investigators often wish to determine whether the screening was the cause of an observed decline in breast-cancer mortality. Typically the presence of an association is hypothesised and tested in the first instance, with assessment of causality being addressed subsequently. An association is a necessary criterion for inferring causality, but the presence of an association is by no means sufficient in itself to establish proof of causality.

Several factors must be considered before an observed association can be considered to be causal. Most obviously, one must consider the temporal order of the intervention or exposure and the effect with which it is associated. A causal relationship can be ruled out if the outcome is present before the intervention or exposure occurs, or before it is sufficiently matured to impose an effect. In the case of screening mammography, there is debate for example about whether changes in breast-cancer mortality in the early phases of screening program implementation can be attributed to screening per se.

Secondly, the impact of the intervention must be disentangled from other possible explanations of the association. This is often the greatest challenge in the determination of causality, made more difficult when possible alternative explanations have not been subject to measurement during the period of interest.

Of particular relevance to screening mammography is whether the declines in breast-cancer mortality that have been observed in other countries have been linked to screening or advances in treatment. Other factors that may influence or confound an observed association between screening and mortality are participation, as in whether women who undergo screening are healthier than those who don’t (the so-called ‘healthy screenee’ or ‘healthy screening participation’ effect), and other risk factors for breast cancer, such as hormone replacement therapy, that might be related to screening participation.

Additional features of observed associations that may strengthen the likelihood of causality are: i) a plausible biological explanation; ii) reversibility of the association when the intervention is removed; and iii) evidence of a dose–response relationship between the intervention and the outcome (even if this ultimately displays a plateau). In relation to dose–response, evaluations of screening may compare effects of different screening regimes on mortality. In other words, it may be expected that women who were screened more frequently would have lower mortality than those having infrequent mammograms.
EPIDEMIOLOGICAL INVESTIGATIONS OF CAUSALITY

The broader discipline of epidemiology uses several different approaches to investigating causality in the face of possible confounding factors. These approaches are discussed by study design type.

1. ECOLOGICAL INVESTIGATIONS

These are studies of sub-populations rather than of individual people. Often they do not entail the collection of new customised data, but rather an analysis of existing health statistics (e.g., mortality statistics). Ecological investigations are generally not suitable for demonstrating causality at an individual level, but they contribute relevant evidence of an association, or lack of association. It is often through ecological research studies that an association is first observed, prompting further empirical study of the causal significance. On the other hand, ecological investigations are often the most appropriate designs for evaluating program effects at a population or group level. Ecological investigations include surveillance of populations, using routinely collected health statistics. They may comprise: i) mortality statistics classified by age, sex, locality, and cause of death; or ii) disease notifications or recordings, as applying for example to infectious diseases or cancers. Regular, ongoing national health surveys may also provide ecological data. Serial measurements over time are often a feature of ecological research.

The main point of differentiation between an ecological investigation and other observational studies is that individual person-level data are not used in an ecological investigation. Instead, analyses are applied to routinely collected group or population statistics.

2. INDIVIDUAL-LEVEL OBSERVATIONAL STUDIES

The main categories of observational study designs that make use of information at the level of the individual are the case-control and cohort designs. However, a simple cross-sectional survey with data available at the individual level would also fall into this category.

2.1 CASE-CONTROL STUDY

Case-control studies compare the medical history and past experiences of persons who have the outcome of interest (cases) with those of persons who do not have the outcome of interest (controls). The outcome of interest is typically the presence of a particular condition, or death from this condition. To minimise selection bias, cases and controls should come from the same population.

Statistical analysis is used to determine whether there is an association of past exposure to the putative risk/protective factor with the outcome of interest. A case-control study has the ability to investigate more than one risk or protective factor at a time, but not more than one outcome of interest. A case-control study is almost always retrospective, dealing with historical risk/protective factors.

Although case-control studies can be flawed by biases that may be difficult or even impossible to eliminate, they are a valuable design option because they can give indicative results rapidly and at relatively low expense. The advantages of the case-control method are that: i) it is an excellent way to
study rare diseases where it would be otherwise difficult to recruit adequate numbers of cases; ii) it is an excellent way to study diseases with a long latency; iii) because it is retrospective, it provides a relatively quick answer; iv) it is comparatively cheap; v) it usually requires a smaller sample size than an equivalent cohort study; vi) it can often be based on existing records; vii) it may be better placed to investigate confounding and interaction precisely than a cohort study, due to more manageable numbers of study participants; and viii) it can be used to study several possible causes or exposures to risk simultaneously, including short-term risk factors.

The disadvantages of the case-control method are that: i) it often relies on subjects’ recall and/or completeness of existing records; ii) it may be difficult or impossible to validate this information; iii) there may be incomplete allowance for extraneous factors; iv) the selection of a suitable comparison (control) group may be difficult; v) rates cannot always be calculated; vi) it is often difficult to obtain sufficiently detailed and accurate information about the timing of the exposure and the outcome to investigate temporal sequence; and vii) only one outcome can be investigated. Case-control studies are often preferred for reasons of convenience or pragmatism. It is essential that adequate consideration be given to sources of bias, particularly in case and control selection, and in relation to validity of recalled information. The use of recorded information, including that obtained through data linkage, often would be more valid than that gained through self-reported recall. In addition, concurrently collected information on historical exposure to possible confounding factors, if available, would generally strengthen the research evidence.

Nested case-control and case-cohort designs are variations of the case-control study design, which are set within a broader cohort study, thus facilitating avoidance of selection bias. Such studies present efficiency advantages over the traditional cohort approach.

2.2 COHORT STUDY

In a cohort study, a defined population (‘cohort’) is identified and classified on the basis of exposure to known levels of putative risk/protective factors for the outcome of interest (eg, diagnosis with a disease or death). The numbers of persons in the cohort and the numbers in each subset are known. All are followed over a period, often for years or even decades, and the disease outcomes recorded and counted at specified intervals. Hypotheses about effects of putative risk/protective factors are tested by comparing rates of disease in different exposure categories.

Prospective cohort studies are generally a more valid study design than case-control studies, but require a substantial number of subjects, especially if the outcome of interest is rare. Cohort studies in the cancer field often require serial measurements over prolonged periods of time. It is important to recognise that a cohort study, like a case-control study, is not an experiment, as the risk/protective factors are not assigned by the investigators (i.e., they are simply observed). Therefore, inferring causality is more problematic than for an experimental design where the factor/intervention of interest can be varied (preferably in one study arm only) in isolation from other possible confounding factors.

It is possible to get results from a cohort study without a prolonged period of study, if detailed information about exposure to risk factors at some time in the past is available in sufficient detail for a population of sufficient size (i.e., the study can be undertaken retrospectively – a ‘historical cohort’ study). A method that permits reliable linking of past and present medical and other relevant records, such as a record linkage system, would facilitate this approach. Record linkage is the process of relating information
from two or more sets of records — often compiled years apart and/or by different agencies — about the same individuals. A prerequisite is a way to identify individuals with a high degree of precision, such as a unique numbering system or ID number. Although retrospective cohort studies can be undertaken much faster than prospective studies, and provide useful indicative results, the most powerful cohort studies generally are prospective.

Advantages of cohort studies include that: i) it may be possible, with good monitoring, to obtain complete data on putative risk/protective factors, cases and stages; ii) the timing or order of exposures and outcomes is unambiguous; iii) more than one outcome can be studied simultaneously; iv) it is possible to calculate and compare rates and risks; and v) for prospective designs, there can be better quality control of data collection.

Weaknesses that may result in biased or spurious results can occur from the following factors: i) researchers may study sub-optimal numbers, particularly when the outcome is rare; ii) subjects are not assigned randomly to a group, as in an experimental study, and therefore confounding from inter-related or extraneous factors can be present; iii) studies may need to take many years, even decades; iv) circumstances (including the presence/absence/nature of the risk/protective factor) may change during the study; v) study questions may become redundant or change during course of the investigation; and vi) these studies are typically expensive, leading to compromises in data collection.

Analyses of data from cohort studies generally involve calculating and comparing rates of the outcome of interest, which are commonly expressed as person-years of observation. Adjustment can be made for potentially confounding factors, as long as sufficient data have been collected on these factors during the study.

3. EXPERIMENTAL STUDIES

These studies often provide the underpinning evidence for introducing a service. They also may be used in established services to trial potential enhancements. They would not be applicable, however, for evaluation of services that are already evidence-based. Ethical problems generally would prevent withholding such services to establish a control group, once initial trials have proven benefit and the services have been established.

Although not applicable for evaluating BreastScreen, experimental designs are discussed now for reference purposes and to provide a context for assessing the observational designs that are more directly applicable.

An experimental study entails the prospective and active intervention of the investigators, preferably with a control arm of subjects who receive identical management except for the intervention of interest. In some way, the treatment (or other intervention) is ‘assigned’ by the investigator in a random or non-random fashion. The purpose of random allocation is to eliminate or minimise bias in the selection of subjects for treatment. This greatly enhances the validity of the results. Preferably, the subjects and those
observing trial results should be unaware of which subjects are receiving the experimental and control regimens, thus eliminating any influence upon the individual’s response or upon the measurement of the outcome. The impact of the intervention should be systematically and proactively measured.

An advantage of experimental studies is that the nature of the intervention can be precisely defined and controlled. This makes it possible to control both the nature and the magnitude of the exposure or dose. One can also more carefully control background clinical management, or at least record this management to ensure that it is equally applied across the treatment arms. It is possible to assign subjects to treatment arms by stratification, to ensure that other factors are comparable (usually allocating an equal number of men and women to each arm). It would be expected that other potentially confounding factors, which have not been allocated by design, would be averaged out by the randomisation.

Despite randomisation, residual confounding may still occur, as for example from subtle differences between comparison arms. It is important, therefore, not to rely entirely on randomisation to ensure balance of measured risk factors between comparison arms, even when the study seems to be sufficiently large to ensure that randomisation has had the opportunity to achieve balance. Notably, reliance on randomisation is more important for achieving a balance of unmeasured factors that may confound results.

The measurement of outcomes should be made at baseline, if possible at intermediate time-points (not relevant when the outcome is death), and at the end of study for all participants. Subjects who are loss to follow-up have the potential to bias the results as: i) they may not be comparable to the remainder of the study population; and ii) they often are more likely to have not responded to the intervention. A conservative analysis is to treat all those lost to follow-up as treatment failures.

A disadvantage of an experimental study is that it may not reflect how, or in whom, the intervention may be used in practice. Furthermore, the more regular follow-up visits included in the study may in themselves have influence upon the results. One must also take care not to extrapolate the results of experimental studies beyond the population in which they were conducted. Many subjects are excluded from experimental studies who may be present in the broader community (e.g., children, the elderly, pregnant women, persons with co-morbidity, etc.), and therefore one shouldn’t presume that the results of an experiment study necessarily can be generalised to the broader population. Furthermore, the intervention itself may not be used in the same manner or with the same skill, when applied in the broader community. Once again this has the potential to impact on the general applicability of the findings.
CHALLENGES FOR ASSESSING CAUSALITY IN BREAST-CANCER SCREENING

UNIQUE FEATURES OF CANCER

There are several features of cancer incidence or mortality as outcomes of interest that introduce additional challenges for assessing causality from epidemiological studies. Cancer is generally a slowly progressing disease with a long period of latency before it is able to be detected through screening. Other complications can arise from a variable, and often substantial, time between that point when the cancer is able to be detected by screening and the time when it is clinically detectable (‘sojourn time’). A sub-period of the sojourn time is the ‘lead time’, this being the period of time gained by bringing forward the diagnosis by detecting the cancer at a screening visit (see below for the implication of lead-time bias upon screening studies).

Secondly, tumours behave variably. Even within one cancer type, some tumours are slow-growing while others are fast-growing. Genetic differences between tumours also are becoming more apparent. It is possible that cancers detected by screening include some that may never have become problematic within the lifetime of the individual.

When considering the heterogeneity of tumours, it is important to recognise that there may be a relationship between the nature of the tumour, the length of the preclinical sojourn time and the survival time. This has implications for length bias (see below).

UNIQUE FEATURES OF SCREENING

There are several features of screening interventions that present additional challenges for the assessment of causality through epidemiological studies. Some people with positive tests will ultimately be found to be disease-negative (false positives) and some people with negative tests will be found to be disease-positive (false negatives), as seen in diagnostic testing. In the screening setting, false-negative results may not be detected for a considerable period of time while the person remains asymptomatic. Adequate follow-up therefore can be important over a prolonged period to assess false negativity.

The main rationale for a screening program is that earlier diagnosis will lead to an improvement in outcomes, such as prevention of advanced disease or of death. In other words, if the lead-time enables a curative treatment to be offered rather than a palliative one, then screening can be justified (Warwick & Duffy, 2005). However, if survival time (measured from detection to death) is only lengthened by the duration of the lead time itself, then there would be little if anything gained, i.e., the patient would still die at the same time as occurring without screening, but with the added burden of being aware of the disease for longer. This phenomenon can be responsible for ‘lead-time bias’ in poorly designed survival studies. For this reason, disease-specific mortality, rather than disease-specific survival, is generally chosen to be the primary endpoint in a screening study.
A related problem is ‘length bias’, which occurs because of the positive correlation between the length of the preclinical period of disease and survival (i.e., slow-growing tumours result in better survival). Somewhat counter-intuitively, slow-growing tumours are more likely to be detected by screening because their preclinical phase is longer (Warwick & Duffy, 2005). Therefore it is possible that screen-detected tumours may artificially have a more favourable prognosis than clinically detected tumours, even if the tumours are of a similar size. As a result, survival analyses may again be biased.

Controlling for confounding factors in screening studies is problematic. In the case of whole population screening, the population is highly heterogeneous with respect to demographic characteristics, such as age, gender, race, and socio-economic status, and in risk profile. Many of these factors would be potential confounders that may not have been recorded in epidemiological screening studies. Furthermore, healthy screening participation bias is a potential problem if health conscious persons are more likely to attend screening. Conversely, persons who may be aware that they are at increased risk (e.g., as may apply from family history or a known high-risk genotype) may be more likely to attend.

In the analyses of screening studies, investigators must be clear whether the findings relate to the effect of being screened (i.e., actually attending screening), or the availability of a screening program (e.g., being invited to attend screening). The latter is inclusive of persons who do and do not attend screening.

**UNIQUE FEATURES OF EVALUATING ESTABLISHED SERVICE PROGRAMS**

There are considerable challenges when evaluating an existing service screening program. In such instances, it is often desirable to be able to evaluate the program retrospectively, although this can introduce considerable complexity. On the other hand, an evaluation of service screening aims to captures how the program works in practice, with more realistic invitation rates, reminder rates, delays in time to attendance, operator skill, and rates of subsequent referral than may apply in an experimental setting.

The quality of records relating to screening invitation and attendance, and cancer diagnosis and mortality may be less than ideal for an evaluation study. Details of potentially confounding risk factors may not have been recorded at all, and it may not be possible to obtain such information retrospectively. It is also possible that the quality of records will not be consistent across the entire population, with variations between geographical, cultural or clinical settings.

The usefulness of records and ability to minimise bias can be very much enhanced by data linkage at the individual level between screening, cancer and mortality registries. Typically this calls for a unique identification number. Where screening has been offered to a particular age cohort, information on age at screening, age at diagnosis, and age at death are important requirements for appropriate matching. Ideally, the outcome of interest (e.g., cause of death) can be independently verified, as for example by a panel of experts who are blinded to the patient’s screening history.

When evaluating an existing screening service, one must be cognisant of the possible dilution of screening with diagnostic tests. Persons who have suspicious symptoms may be more likely to attend for screening. In contrast, screening may also occur by proxy in the form of a diagnostic test, particularly if this test is reimbursed by the health system where screening is not. The timing and number of screens should also be considered, preferably at the individual level.
It is not always possible to disaggregate the impact of a screening program per se from other forms of screening that may occur in parallel. Women who have not participated in BreastScreen, for example, may have undergone screening outside the Program. In Australia, screening mammograms are not rebated through Medicare, but substantial numbers of women are thought to be screened under the guise of diagnostic testing. Consequently, the effect of screening mammography on mortality may be diluted. In addition, an increase in the rate of physician physical examinations or self-examinations may occur in parallel with this screening. Increased awareness from advertising and health promotion activities could itself impact upon health outcomes. This has the potential to confound estimates of effects of screening mammography.

An overarching problem with the evaluation of an existing screening service, particularly a national program available to all, is the absence of a comparable control group, although such a group may exist for a time during the roll-out phase. Without a comparable control group, selection bias may be impossible to negate. Omitting a contemporaneous control group and simply comparing mortality before and after the introduction of the screening program would not allow account to be taken of factors other than screening that have changed within the same time-frame, such as the introduction of new treatments.

Because of the prolonged periods that can occur between cancer diagnosis and cancer death, a cancer-screening program can take a considerable period time to have a demonstrable mortality effect. Evaluation of screening effects should allow for sufficient follow-up time to detect decreases in mortality. This may exceed 10 years from screening implementation.

CONCLUDING COMMENT

Together the factors outlined in this section of the report create challenges in the determination of causality from epidemiological evidence, particularly in the evaluation of cancer screening services.
DESCRIBING AND RANKING DESIGN OPTIONS

CONSIDERING DESIGN OPTIONS

METHODOLOGY REVIEW

ETG members were aware from the literature search that a sophisticated evaluation methodology probably would be necessary to identify a BreastScreen effect, especially as the period since rollout of BreastScreen had been comparatively short. This was underscored in the review by Tornberg, et al. (2006) of data from the Nordic capitals, which showed that population-based breast-cancer mortality trends were too crude a measure to detect screening effects in the initial 7–12 years following screening introduction. These trends did not reveal a mortality effect, even in settings where significant effects already had been demonstrated using a more valid approach. The researchers recommended that a meticulous approach to evaluation be taken, ideally involving the linking of screening, cancer incidence and mortality records at an individual person level. Additional methodological considerations are presented in the original report.

Design options and issues for BreastScreen evaluation were considered in detail in a review of methodological papers on breast screening evaluation (Appendix B-1). These papers addressed aspects such as:

- **design types**: i.e., ecological, case-control, paired availability, cohort, case-cohort, simulation models, and population-mortality trend analyses and forecasting – Alibhai (2006); Baker, Kramer and Prorok (2004); Berry et al. (2005); Blanks (1996); Boer, Plevritis and Clarke (2004); Chen, et al. (1998); Demissie, Mills and Rhoads (1998); Duffy (2005); Eddy (1980); Erbas, Hyndman and Gertig (2005); Eyre et al. (1995); Feig (1996); Feuer et al. (2004); Hakama et al. (1999); Hayasaka (2005); Joffe (2003); Michaelson et al. (2003a); Michielutte et al. (2000); Moss (1991); Olsen et al. (2005); Prorok (1984); Sasco (1986); Szeto and Devlin (1996); Tornberg et al. (1994); Walter (2003); Warwick and Duffy (2005); Weiss (1983)

- **different endpoints**: i.e., breast-cancer mortality, all-causes mortality, excess mortality, selective mortality, and mortality surrogate measures such as tumour size, nodal status, grade, and screening parameters such as lead time, interval cancer rates, and screening sensitivity – Alibhai (2006); Benichou (2001); Day et al. (1995); Day, Williams and Khaw (1989); Day (1989); Day (1985); Eyre et al. (1995); Feig (1996); Lenner (1990); Mackay et al. (2001); Michaelson et al. (2003a); Michaelson et al. (2003b); Myles et al. (2003); Prorok (1984); Sasco (1986); Tabar et al. (1999); Walter (2003); Warwick and Duffy (2005); Weiss (1983)

- **sources of bias**: i.e., biases related to subject definition, lead time, length time, control selection, screening participation, length of follow-up time, effects of healthy screening participation, recall effects, and misclassification of diagnostic and screening mammograms – Alibhai (2006); Day (1981); Eyre et al. (1995); Feig (2002); Gullberg (1991); Joffe (2003); Miettinen, Yankelevitz and Henschke (2003); Miettinen et al. (2002); Raffle (2003); Sasco (1986); Schmidt (1990); Smith-Bindman et al. (2006); Staatsman et al. (2004); Walter (2003); Warwick and Duffy (2005); Weiss (1983)
• factors attenuating measured effects: i.e., inadequate follow-up time, poor definition of the screened and unscreened, and inclusion of deaths from cancers diagnosed prior to screening commencement – Demissie, Mills and Rhoads (1998); Glasziou (1992)

• statistical issues: i.e., as relating to choice of test, meeting underlying statistical assumptions, and allowing for clustering – Tarone (1995), Tarone and Gart (1989); Warwick et al. (2004).

EVALUATION DESIGNS USED PREVIOUSLY

The publications reviewed for this purpose are listed in References: Included Studies. The most commonly used designs for evaluating other screening programs have included:

• retrospective cohort designs – which applied to 24 publications and in general produced indicative results: Anttila, Koskela and Hakama (2002); Barchielli et al. (1994); Blanks et al. (2000); Broeders et al. (2001); Duffy (2006a, 2006b); Duffy et al. (2002); Gorini et al. (2004); Hakama et al. (1999); Jonsson et al. (2000, 2001, 2003a, 2003b); Lenner and Jonsson (1997); McCann, Duffy and Day (2001); Olsen et al. (2005); Paci et al. (2002a, 2002b); Parvinen et al. (2006); Peer et al. (2240); Sasieni (2003); Tabar et al. (2001, 2003); Van Dijck et al. (1996)

• case-control designs – which applied to 13 publications and in general produced indicative results: Broeders et al. (2002); Collette et al. (1992, 1984); Elmore et al. (2005); Fiedler et al. (2004); Miltenberg et al. (1998); Moss et al. (1992); Palli, Del Turco and Buiatti (1986); Palli et al. (1989); Van Dijk et al. (1996, 1994); Verbeek, Hendriks and Holland (1984, 1985)

• population time-trend analyses – which applied to 10 publications and in general produced trend results that were difficult to interpret due to potential for confounding from treatment and other secular changes: Barchielli and Paci (2001); Chu et al. (1996); Coburn et al. (2004); Garne et al. (1997); Giles and Amos (2003); Kricker et al. (1999); Quinn and Allen (1995); Otto et al. (2003); Sigurdsson, Adalsteinsson and Ragnarsson (1991); Smith, Kricker and Armstrong (1998)

• prospective cohort designs – which applied to two publications and produced good-quality evidence: Ellman et al. (1993); Moss (1999)

• ecological designs (geographic comparisons) – which applied to two publications and produced results indicative of screening effects: Das, Feuer and Mariotto (2005); Taylor et al. (2004)

• case-cohort designs – which applied to one publication and produced good-quality evidence: Thompson et al. (1994)

• Markov simulations – which applied to one publication (in addition to publications in the methodology section) and produced results suggestive of screening effects: Chen et al. (1998)

• paired availability designs – which applied to one publication and produced indicative results: Baker, Kramer and Prorok (2004)

Further details of these designs, the associated statistical analyses, and study assumptions, strengths and weaknesses are recorded on the data extraction forms (Appendix A-5). All designs were regarded as having some merit, although ETG members considered population-trend analyses to be too crude an approach for the present purpose. The paired availability design had useful characteristics regarding the use of historical controls that members felt should be considered when deciding on preferred designs.
MESSAGES FROM THE REVIEW

The review demonstrated the vulnerability of all observational designs to bias. However well designed, it is inevitable that evidence from these studies would lack the validity of randomised trials. While case-control studies were frequently undertaken in preference to prospective studies for efficiency purposes, they were generally more vulnerable to bias, unless designed and undertaken with great care. Concern often presented about potential selection bias in case-control studies, especially when there was limited access to risk factor and treatment data to adjust for potential confounding, and the availability of only poor quality data on screening exposure such that cases and controls were poorly defined. Depending on the design, bias from a healthy screening participation effect also was possible.

By comparison, prospective cohort studies generally were not a viable option due to the long study durations required. Prospective case-cohort options, while reducing the scale of the study, also would take too long to complete. A deficiency of many evaluation studies, irrespective of design, was implementation very early after screening commencement when little effect would be expected.

Despite misgivings about results of individual case-control studies, pooled analyses indicated that their results were generally very similar to estimates drawn from the randomised trials (i.e., estimates adjusted for sub-optimum trial participation). In other words, estimated mortality effects were similar, irrespective of study design, suggesting that biases in individual case-control studies occurred in different directions and tended to cancel out in pooled analyses.

Retrospective designs, whether cohort or case control, sometimes lacked comprehensive data on screening exposure, and more frequently, on potential confounders such as differences in treatment effect and in family history of breast cancer, genetic predisposition, reproductive history, use of hormone replacement therapy, history of breast surgery for non-malignant conditions, and other risk indicators. Results were therefore more reliant on questionable assumptions of equivalence.

When surrogate measures of breast-cancer mortality, such as tumour size, nodal status and grade, were used in a case-control design, self-reporting of family histories and other risk factors often took place. In addition, the evaluation was able to occur earlier (i.e., before waiting for deaths) and there were opportunities for greater statistical power. However, surrogate measures were imperfect measures of mortality and were potentially vulnerable to lead-time, length-time and over-diagnosis biases.

The review showed that simulation models had been used on many occasions to assess effects of screening mammography, but results varied considerably, albeit tending to point in the same direction. Results seemed to be very sensitive to small changes in model structure and assumptions. The publications reviewed in this regard included those listed in Appendix B-2: Borio et al. (1980); Duffy et al. (1980); Fett (2001); Knox (1988); Paci et al. (1995); Rijnsburger et al. (2004); Van den Akker-Van Marle et al. (1999, 1997); Van der Maas et al. (1989); Wang et al. (2001) and the following references in Appendix B-1: Berry et al. (2005); Boer, Plevritis and Clarke (2004); Chen et al. (1998); Eddy (1980); Myles et al. (2003); Szeto and Devlin (1996).

Studies of case survivals were considered to be of secondary importance in the present review, since the primary objective was to review design options for assessing screening effects on population-based mortality. In other words, the denominator of interest was the population, not people with diagnosed cancers. Nonetheless, because ETG members felt that survivals might permit inferences of population-based mortality effects, 25 survival publications were included in the review (Appendix B-3).
ETG members considered that studies comparing survivals for screen-detected and symptomatic cancers were vulnerable to bias from lead-time, length-time, and potentially over-diagnosis. Where interval cancers were excluded, comparisons would be subject to additional bias. As an indicator of population-based mortality effects, results of survival analyses were generally very difficult to interpret.

Further examination of ‘grey’ literature publications gave useful contextual information for the present review, although ETG members felt that few insights were gained from these publications that were additional to those provided through the peer-reviewed literature.

**DEVELOPING DESIGN OPTIONS**

After considering the methodology papers and the evaluation designs used for other screening services ETG members nominated nine design options for further consideration:

- **option 1**: a MISCAN micro-simulation, incorporating Australian population data, screening data, and cancer registry data, and calibrated to Australian breast-cancer incidence and mortality rates
- **option 2**: an ecological design, in which the association of BreastScreen participation with trends in breast-cancer mortality was investigated by SLA
- **option 3**: a case-control design, to investigate associations of exposure of individual women to a BreastScreen invitation for screening, and to participation in BreastScreen, with unfavourable breast-cancer prognosis at diagnosis
- **option 4**: a case-control design, to investigate associations of exposure of individual women to a BreastScreen invitation for screening, and to participation in BreastScreen, with risk of distant metastatic breast-cancer recurrence
- **option 5**: a case-control design, to investigate associations of exposure of individual women to a BreastScreen invitation for screening, and to participation in BreastScreen, with risk of death from breast cancer
- **option 6**: a retrospective cohort design, to:
  1. compare death rates from breast cancer in women who were invited by BreastScreen to screening with those not invited to this screening
  2. compare death rates from breast cancer in women who participated in BreastScreen with those who did not participate in BreastScreen
- **option 7**: a prospective cohort design, to:
  1. compare death rates from breast cancer in women who were invited by BreastScreen to screening with those not invited to this screening
  2. compare death rates from breast cancer in women who participated in BreastScreen with those who did not participate in BreastScreen
- **option 8**: a case-cohort design, to investigate associations of exposure of individual women to BreastScreen invitations for screening, and to participation in BreastScreen, with risk of death from breast cancer
- **option 9**: A comparison of prognostic characteristics and survivals for female breast cancers according to whether detected by screening mammography.

While other options also were considered, including a number with changes in survivals as endpoints, they were excluded from the ranking because of serious concerns about validity and/or feasibility.
AIMS AND PROCESS OUTLINES BY DESIGN OPTION

OPTION 1: MISCAN MICRO-SIMULATION OF BREASTSCREEN EFFECT ON BREAST-CANCER MORTALITY

Aim

To estimate the effect of BreastScreen screening participation on breast-cancer mortality, using a MISCAN micro-simulation model, with Markov transitions between states, which incorporated Australian population data, breast-cancer incidence prior to the introduction of BreastScreen, staging, and stage-specific survival data, and BreastScreen cancer detection and interval cancer rates, and with calibration to Australian breast-cancer incidence and mortality.

Process outline

- The MISCAN model would be adapted and calibrated to the Australian setting, using local population data on births, deaths from causes other than breast cancer, and life tables; pre-screening incidence, stage and stage-specific survivals; differences in incidence and mortality by screening exposure; and estimated screening sensitivity, specificity and related characteristics.
- The model would be used to estimate effects of BreastScreen participation on breast-cancer mortality by entering relevant inputs like the screening age range, screening interval, and screening participation, and comparing model outputs according to BreastScreen exposure.
- The model also could be used, subject to gaining access to relevant treatment data, to estimate effects of accompanying treatment changes on breast-cancer mortality.
OPTION 2: ECOLOGICAL STUDY OF ASSOCIATION OF BREASTSCREEN PARTICIPATION WITH BREAST-CANCER MORTALITY BY SLA

Aim
To determine associations by residential SLA of BreastScreen participation with: (1) time-lagged breast-cancer mortality; or (2) time-lagged annual changes in breast-cancer mortality.

The time lag either would be: set a priori from mean survival periods available from the National Cancer Statistics Clearing House (NCSCH) or obtained through modelling these outcomes on participation rates to identify the time lag that gave the best fit. There would be adjustment in the analysis for numbers of mammograms provided outside BreastScreen, relative socio-economic disadvantage of residential area (e.g., SEIFA index), health-service accessibility (e.g., ARIA index), parity estimates using perinatal statistics, other socio-demographic descriptors such as ethnicity mix, State/Territory jurisdiction, and breast-cancer mortality rates for cancers diagnosed in the 1980s before the advent of screening mammography.

Process outline
• SLA-specific data would be retrieved for the 1990–2005 period for 50–69 year olds by age, as relating to:
  – BreastScreen participation (data from BreastScreen)
  – population data and socio-demographic descriptors (interpolations from ABS census data)
  – privately funded mammograms (data from Medicare)
  – mammograms provided to public hospital inpatients (data from hospital separations).

• Annual SLA-specific breast-cancer death data would be obtained from the NCSCH (or State/Territory cancer registries) by age for diagnoses occurring in 1990–2004 among 50–69 year olds resident in the respective SLA at diagnosis.

• Breast-cancer mortality rates would be obtained for diagnoses occurring in 1983–89 among 50–69 year olds resident in the respective SLA at diagnosis (data from the NCSCH).

• Breast-cancer death rates or annual reductions in breast-cancer mortality would be modelled by SLA according to BreastScreen participation, with adjustment for other characteristics that condition BreastScreen risk estimates, such as SEIFA index, ARIA index, State/Territory jurisdiction, other socio-demographic descriptors, mortality rates for breast cancers diagnosed in 1983–89, and mammogram exposure outside BreastScreen. This could involve a weighted regression analysis, using count data as the outcome with a population offset.
OPTION 3: CASE-CONTROL STUDY OF ASSOCIATIONS OF BREASTSCREEN INVITATION FOR SCREENING, AND OF BREASTSCREEN PARTICIPATION, WITH A SURROGATE DIAGNOSTIC MARKER OF BREAST-CANCER MORTALITY

Aim

To determine associations of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with a surrogate measure of breast cancer mortality (e.g., most severe category of Nottingham Prognostic Index (NPI)) in women diagnosed with breast cancer from 2007 at 50–69 years of age.

Adjustment would be made for exposure to other factors that conditioned risk estimates for BreastScreen invitation/participation, such as exposure to Medicare-funded and other externally funded mammograms, SEIFA index, ARIA index, State/Territory jurisdiction, other socio-demographic markers, and self-reported risk factors, as related to family history of breast cancer, genetic predisposition, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions.

Process outline

- Cases would be identified prospectively from notifications of breast cancer diagnoses from 2007 to State/Territorial cancer registries where:
  - the diameter, nodal and grade signified the most severe NPI category (i.e., 15-year survival indicated to be around 15% by historic data)
  - the person was aged 50–69 years at diagnosis
  - the person was recorded on the Electoral Rolls.

- To enhance statistical power, three controls would be selected at random per case, from among women with the same date of birth on the Electoral Rolls, who were still alive at the time of diagnosis of the case, and where there was not evidence from the State/Territorial cancer registry of a breast cancer diagnosis before the diagnosis date of the respective case.

- State/Territory BreastScreen services would search their records and classify cases according to whether they had been: i) invited for screening (y/n); and ii) screened (y/n), in the 3 years prior to diagnosis, or earlier. Controls would be similarly classified for the time ‘window’ applying to their matched cases.

- Cases and controls would be surveyed:
  - regarding self-reported histories of exposure to screening mammography outside BreastScreen
  - to confirm histories of BreastScreen screening invitations, and BreastScreen screening participation, as obtained from BreastScreen records
  - regarding risk factors for breast cancer, as related to family history of breast cancer, genetic predisposition, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions.
A standard case-control analysis would be undertaken to determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with being a case as opposed to a control (i.e., two separate analyses). In particular, the relative odds of being a case would be determined according to BreastScreen invitation/participation, adjusting for other characteristics that conditioned risk estimates, such as exposure to mammograms outside BreastScreen, SEIFA and ARIA indices (drawn from residential postcodes), State/Territory jurisdiction, and risk factor exposure (e.g., as related to family history, genetic predisposition, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions). Conditional logistic regression analysis could be used for this purpose.
OPTION 4: CASE-CONTROL STUDY OF ASSOCIATIONS OF EXPOSURE OF INDIVIDUAL WOMEN TO A BREASTSCREEN INVITATION FOR SCREENING, AND TO BREASTSCREEN PARTICIPATION, WITH RISK OF DISTANT METASTATIC RECURRENCE OF BREAST CANCER AS A SURROGATE MORTALITY MARKER

Aim

To determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with risk of a distant metastatic breast-cancer recurrence in women diagnosed with breast cancer when 50–69 years of age in 1994–2005.

Adjustment would be made for other characteristics that condition risk estimates for BreastScreen invitation/participation, such as exposure to Medicare-funded and other externally funded mammograms, SEIFA index, ARIA index, other socio-demographic descriptors, State/Territory jurisdiction, and self-reported risk factors (i.e., as related to family history of breast cancer, genetic predisposition, adverse reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions).

Process outline

- Medical oncologists would be recruited to identify women with a recurrence from 2007, so that these women could be invited to participate in the study.
- Participants would be restricted to women aged 50–69 years at diagnosis who were listed on the Electoral Rolls and had been diagnosed since 1994 (this would allow some time for initial screening roll-out).
- Three controls would be selected at random per case from the Electoral Rolls. They would be women with the same date of birth as the respective case who were still alive at the time of recurrence of the case, and where there was not evidence from State/Territorial cancer registries of a breast-cancer diagnosis before the diagnosis date of the case.
- State/Territory BreastScreen services would search their records and classify cases according to whether they had been: i) invited for screening (y/n); and ii) had been screened (y/n), in the 3 years prior to diagnosis, or earlier. Controls would be similarly classified for the same time ‘windows’ as their respective cases.
- Cases and controls would be surveyed regarding:
  - exposures to mammograms for screening purposes outside BreastScreen, and also to confirm BreastScreen screening invitation and BreastScreen screening histories
  - risk-factor profiles, as related to family history of breast cancer, genetic predisposition to breast cancer, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions.
A standard case-control analysis would be performed to determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with being a recurrence case as opposed to a control (i.e., two separate analyses). The relative odds of being a recurrence case would be determined for BreastScreen invitation/participation, adjusting for other characteristics that condition risk estimates, such as exposure to mammograms outside BreastScreen, SEIFA and ARIA indices (drawn from residential postcodes), State/Territory jurisdiction, and risk factor exposure (e.g., as related to family history of breast cancer, genetic predisposition, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions). Conditional logistic regression analysis could be used for the analysis.
OPTION 5: CASE-CONTROL STUDY OF ASSOCIATIONS OF EXPOSURE TO A BREASTSCREEN INVITATION FOR SCREENING, AND TO BREASTSCREEN PARTICIPATION, WITH RISK OF BREAST-CANCER MORTALITY

Aim
To determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with breast cancer mortality in women diagnosed with breast cancer when 50–69 years of age in 1994–2005. Adjustment would be made for exposure to other characteristics that conditioned risk estimates for BreastScreen invitation/participation, such as exposure to Medicare-funded and other externally funded mammograms, SEIFA index, ARIA index, other socio-demographic descriptors, and State/Territory jurisdiction.

Process outline
• State/Territorial cancer registries would identify breast-cancer deaths occurring in 1994-2005 where:
  – the person was diagnosed in 1994 or later
  – age at diagnosis was 50–69 years
  – the person was recorded on the Electoral Rolls.
• Three controls would be selected at random per case from among women with the same date of birth on the Electoral Rolls, who were still alive at the time of death of the case, and where there was no evidence from State/Territorial cancer registries of a breast cancer diagnosis before the date of diagnosis of the case.
• State/Territory BreastScreen services would search their records and classify cases according to whether they had been: i) invited for screening (y/n); and ii) screened (y/n), in the 3 years prior to diagnosis, or earlier. Controls would be similarly classified for the time ‘window’ as their respective cases.
• Exposures to mammograms outside BreastScreen would be investigated using Medicare and if feasible, hospital separation data.
• A standard case-control analysis would be performed to determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with being a case as opposed to a control (i.e., two separate analyses). The relative odds of being a case would be determined for these predictors, adjusting for other characteristics that condition risk estimates for BreastScreen invitation/participation, such as exposure to mammograms outside BreastScreen, SEIFA and ARIA indices (drawn from residential postcodes), and State/Territory jurisdiction. Conditional logistic regression analysis could be used for this purpose.
OPTION 6: RETROSPECTIVE COHORT STUDY TO: (1) COMPARE DEATH RATES FROM BREAST CANCER IN WOMEN WHO WERE INVITED BY BREASTSCREEN TO SCREENING WITH THOSE NOT INVITED TO SCREENING; AND (2) COMPARE DEATH RATES FROM BREAST CANCER IN WOMEN WHO PARTICIPATED IN BREASTSCREEN WITH THOSE WHO DID NOT PARTICIPATE IN BREASTSCREEN

Aim
To determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with breast-cancer mortality in women diagnosed with breast cancer when 50–69 years of age in 1994–2005.

Adjustment would be made for other characteristics that conditioned risk estimates for BreastScreen invitation/participation, such as exposure to Medicare funded and other externally funded mammograms, SEIFA index, ARIA index, other socio-demographic descriptors, and State/Territory jurisdiction.

Process outline
- Electoral Rolls would be used to construct a register of women aged 50–69 in the period since 1990 for follow-up.
- Records of State/Territorial BreastScreen services would be used to identify whether these women were invited/screened through BreastScreen and to determine dates of invitations and screening rounds.
- Exposure to mammograms outside BreastScreen would be investigated using Medicare and if feasible, hospital separation data.
- The study group would be linked to the National Death Index to identify deaths from breast cancer and other causes, and to cancer registries to determine rates of diagnoses (see later section on sub-analyses restricted to diagnosed cancers).
- A standard cohort analysis would be performed to determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen participation, with breast cancer mortality (i.e., two separate analyses). Relative risks of breast-cancer mortality would be determined for these predictors, adjusting for other characteristics that conditioned risk estimates for BreastScreen invitation/participation, such as exposure to mammograms outside BreastScreen, SEIFA and ARIA indices (drawn from residential postcodes), and State/Territory jurisdiction. Cox or Poisson regression may be used for this purpose. Person years of follow-up would be censored for different components of the analysis at date of diagnosis of breast tumour, at the end of 2005 for those not diagnosed with breast cancer, or at time of death from other causes if occurring prior to the end of 2005.
- A further sub-analysis could be restricted to women diagnosed with breast cancer from 50 years of age to assess the relative risk of breast cancer death according to: i) whether screened in the three years prior to diagnosis, or earlier; ii) numbers of pre-diagnostic screens; and iii) period since last pre-diagnostic screen. A further analysis again could be restricted to women diagnosed with breast cancer from 50 years of age who had a BreastScreen history to assess the relative odds of breast cancer death according to: i) whether screened in the three years prior to diagnosis, or earlier; ii) numbers of pre-diagnostic screens; and iii) period since last pre-diagnostic screen, adjusting for family history of breast cancer and any other risk-factor data available from BreastScreen records.
OPTION 7: PROSPECTIVE COHORT STUDY TO: (1) COMPARE DEATH RATES FROM BREAST CANCER IN WOMEN WHO WERE INVITED BY BREASTSCREEN TO SCREENING WITH THOSE NOT INVITED TO THIS SCREENING; AND (2) COMPARE DEATH RATES FROM BREAST CANCER IN WOMEN WHO PARTICIPATED IN BREASTSCREEN WITH THOSE WHO DID NOT PARTICIPATE IN BREASTSCREEN

Aim

To determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with breast-cancer mortality in women diagnosed with breast cancer when 50–69 years of age from 2007.

Adjustment would be made for other factors that conditioned risk estimates for BreastScreen invitation/participation, such as exposure to Medicare-funded and other externally funded mammograms, SEIFA and ARIA indices, other socio-demographic descriptors, State/Territory jurisdiction, and risk factors assessed prospectively for all women (i.e., as related to family history of breast cancer, genetic predisposition, pregnancy history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions).

Process outline

- Women who were recorded on the Electoral Rolls and aged 50–69 years would be enrolled in the study prospectively with consent.
- Exposure to BreastScreen would be self-reported and verified through BreastScreen records.
- Exposure to mammograms outside BreastScreen would be investigated prospectively through self report, Medicare, and if feasible, through hospital separation data.
- The study group would be linked prospectively, with consent, to the National Death Index to identify deaths from breast cancer and other causes. There would be corresponding linkage to the NCSCH, and surrogate data for breast-cancer mortality would be collected through BreastScreen and State/Territorial cancer registries (e.g., diameter, nodal status and grade), with consent.
- A standard cohort analysis would be performed to determine the association of BreastScreen invitation/participation with risk of breast-cancer mortality (i.e., initially using surrogate measures and then breast-cancer mortality itself as the endpoint). The relative risk of breast-cancer mortality would be determined according to screening history, adjusting for other factors that condition risk estimates for BreastScreen invitation/participation, such as exposure to mammograms outside BreastScreen, SEIFA and ARIA indices, (drawn from residential postcodes), State/Territory jurisdiction, and breast-cancer risk factors (e.g., as related to family history of breast cancer, genetic predisposition, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions). Proportional hazards regression could be used for this purpose, as described for the retrospective cohort analyses in Option 6.
OPTION 8: CASE-COHORT STUDY OF ASSOCIATION OF EXPOSURE OF INDIVIDUAL WOMEN TO A BREASTSCREEN INVITATION FOR SCREENING, AND TO BREASTSCREEN SCREENING PARTICIPATION, WITH RISK OF BREAST-CANCER MORTALITY

Aim
To determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with breast cancer mortality in women diagnosed with breast cancer when 50–69 years of age.

Adjustment would be made for other characteristics that condition risk estimates for BreastScreen invitation/participation, such as exposure to Medicare-funded and other externally funded mammograms, SEIFA index, ARIA index, other socio-demographic descriptors, State/Territory jurisdiction, and risk factors assessed prospectively for all women in the study (e.g., family history of breast cancer, genetic predisposition, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions).

Process outline
- Women aged 50–69 years at diagnosis of breast cancer would be identified prospectively from 2007 through State/Territorial cancer registries for enrolment in the study, with consent. Only women recorded on the Electoral Rolls would be included.
- At that time, three women would be selected at random per cancer case from the baseline cohort on the Electoral Rolls, selecting women with the same date of birth, who were still alive at the time of diagnosis of the respective case.
- Exposure to mammograms inside and outside BreastScreen would be investigated, especially for the three-year period preceding case diagnosis, and for controls, prospectively through self-report, and through BreastScreen, Medicare and if feasible, hospital separation data.
- Exposure to other risk factors also would be investigated through self-report, as related to family history of breast cancer, genetic predisposition, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions.
- The study group would be linked prospectively, with consent, to the National Death Index to identify deaths from breast cancer and other causes. There would be corresponding linkage of controls to the NCSCH and surrogate data for breast cancer mortality would be collected through State/Territorial cancer registries (e.g., diameter, nodal status and grade), with consent.
- A regression analysis would be performed to determine the association of BreastScreen screening invitation/participation with risk of breast cancer mortality. The women diagnosed with breast cancer who die from it would be the cases. The women who do not die from their breast cancer would not be included in the analysis unless they were selected randomly for the sub-cohort. The sub-cohort would give the person time. Cox regression with appropriate allowance for the sampling (Prentice method) would be used for the analysis.
- In the analysis, the relative risk of breast-cancer mortality would be determined according to BreastScreen invitation/participation, adjusting for other characteristics that condition risk estimates for BreastScreen invitation/participation, such as exposure to pre-diagnostic mammograms outside BreastScreen, SEIFA and ARIA indices (drawn from residential postcodes), and State/Territory jurisdiction. Adjustments also could be made for self-reported breast-cancer risk factors.
OPTION 9: A COMPARISON OF PROGNOSTIC CHARACTERISTICS AND SURVIVAL FOR FEMALE-BREAST CANCERS ACCORDING TO WHETHER DETECTED BY SCREENING MAMMOGRAPHY

Aim

To determine the association of detection through BreastScreen screening with prognostic features of breast cancers and case survival among females aged 50–69 years at diagnosis.

Process outline

- Women diagnosed with invasive breast cancer at age 50–69 years during 1990–96 would be identified through State/Territorial cancer registries.
- BreastScreen records would be searched to assess whether these cancers were BreastScreen-detected.
- Cancer registry and case records would be searched retrospectively to determine diameter, nodal status and grade at diagnosis, and potentially TNM stage, and primary course of treatment.
- Exposure to externally funded screening would be investigated through Medicare and if feasible, hospital separation data.
- Effects of BreastScreen screening would be inferred from differences in staging characteristics and evidence from the literature of the association of these characteristics with long-term survival. In addition, early differences in survival would be assessed directly.
INITIAL RATING OF DESIGN OPTIONS

Members of the ETG independently rated the five design features of individual design options on a five-point ordinal scale, viz: likely: i) value of evidence; ii) ease of implementation; iii) risk of unsuccessful implementation; iv) cost; and v) time to study completion, using the standard scoring sheet (Appendix A-9). Differences were discussed by teleconference, and members could elect to modify their scores as a result.

As indicated in the Methods section, an initial rating of the overall merit score for each design was a weighted average of individual feature scores. This involved weighting the value of evidence by 0.4, ease of implementation by 0.1, risk of unsuccessful implementation by 0.1, cost by 0.1 and time to completion by 0.3. Rating scores and associated commentary are presented in Appendix A-11.

OPTION 1: MISCAN MICRO-SIMULATION OF BREASTSCREEN EFFECT ON BREAST-CANCER MORTALITY

Scores

- Value of evidence: Mean 2.4 (members’ scores: 3, 2, 1, 2, 4)
- Ease of implementation: Mean 4.5 (members’ scores: 5, 5, 4, 4.5, 4)
- Risk of unsuccessful implementation: Mean 4.5 (members’ scores: 4, 5, 5, 4.5, 4)
- Cost: Mean 4.6 (members’ scores: 4, 5, 5, 4, 4)
- Time to study completion: Mean 4.4 (members’ scores: 5, 5, 5, 3, 4)

OVERALL MERIT SCORE: 3.6

Commentary

The MISCAN model scored relatively low on value of evidence, but favourably on ease of implementation, cost, and time to study completion. Although the possibility existed that the output would not calibrate well with observed mortality, this risk was judged to be low.

Members felt that the results would not be as credible as direct measures of mortality outcome, however, since estimates would rely largely on international evidence that may not be applicable to Australia. There is scepticism among many researchers about results of such models, since it appears that results are very sensitive to small differences in model structure and assumptions, and are prone to differ widely.

Members thought it important, if a MISCAN model were applied, to re-estimate as many model parameters as possible, using Australian population data, breast-cancer incidence, staging, stage-specific survivals, screening coverage, and cancer detection and interval cancer rates, to increase local applicability. Although local treatment data would be important for assessing treatment effects, it was doubtful whether data of adequate quality would be available. Sensitivity analyses would be an important component of a MISCAN approach, to test reliability.
Members felt that despite limitations, MISCAN could have a useful adjunctive role in the BreastScreen evaluation by providing quick interim results at low cost. Most members estimated results within 12–18 months. MISCAN models have been widely used to indicate screening effects in other settings and in general have produced results broadly similar to results of screening trials.

Mostly the source data would be available from the ABS, NCSCH, clinical cancer registries and BreastScreen, for calibration to the Australian setting. Person-identified data would not be needed, which would simplify HREC and administrative approvals for data access.
OPTION 2: ECOLOGICAL STUDY OF ASSOCIATION OF BREASTSCREEN PARTICIPATION WITH BREAST-CANCER MORTALITY BY SLA

Scores

- Value of evidence: Mean 3.3 (members’ scores: 3, 3, 3, 3.5, 4)
- Ease of implementation: Mean 4.8 (members’ scores: 5, 5, 4, 5, 5)
- Risk of unsuccessful implementation: Mean 4.4 (members’ scores: 5, 5, 3, 5, 4)
- Cost: Mean 5.0 (members’ scores: 5, 5, 5, 5, 5)
- Time to study completion: Mean 4.5 (members’ scores: 5, 5, 5, 3.5, 4)

OVERALL MERIT SCORE: 4.1

Commentary

This ecological design option scored better than the MISCAN option on value of evidence. Favourable scores applied to ease of implementation, risk of unsuccessful implementation, cost, and time to study completion, with most members considering that study completion could be achieved in one year or less, or 18 months at most, so long as there were not undue delays gaining HREC and administrative approvals.

Proof-of-concept already has been established in NSW, the USA, and other populations. Most source data would be available and would not be required in person-identified form, facilitating HREC and administrative approvals. In addition, the study outcome may capture indirect effects (i.e., ‘externalities’) as well as direct screening effects, at least in part. In other words, (any) effects on outcomes of BreastScreen influences on women’s responses to breast symptoms or on the performance of other diagnostic and treatment services might be captured by this design.

There are risks, however, in that variation in BreastScreen participation may be smaller by SLA than ideal for study purposes, although this appeared not to be a problem in an earlier NSW study. In addition, there may be difficulties identifying historic BreastScreen screening coverage in jurisdictions without centralised record systems.

Some experts are negatively disposed towards ecological study designs, which would detract from the credibility of results. Another disadvantage is that much treatment data would not be available by SLA. Depending on the treatment data available by State/Territory, however, it may be possible to undertake supplementary sub-studies in greater depth in jurisdictions where there are more complete data. A further disadvantage is that personal risk-factor data, such as family history, genetic predisposition, exposure to hormone replacement therapy, and history of benign breast disease, would not be available by SLA level, such that equivalence by SLA would have to be assumed.

Another uncertainty was the appropriate time lag to allow between screening and mortality. This would require modelling or an a priori assumption (e.g., assuming that separately calculated mean case survival periods would apply). Standard ecological assumptions would need to be made, despite the vulnerability of these types of designs to so-called ecological fallacies.

Despite these limitations, members felt that the ecological design option could have a useful complementary role to other studies and would provide quick results at low cost. In general, this ecological design was regarded as a preferable option to a MISCAN model, because direct measures of mortality would be more credible than estimates from simulations relying substantially on overseas evidence.
OPTION 3: CASE-CONTROL STUDY OF ASSOCIATION OF BREASTSCREEN EXPOSURE WITH RISK OF POOR PROGNOSIS AT BREAST-CANCER DIAGNOSIS

Scores

- Value of evidence: Mean 3.1 (members’ scores: 4, 3, 2, 3.5, 3)
- Ease of implementation: Mean 3.8 (members’ scores: 4, 4, 4, 4, 3)
- Risk of unsuccessful implementation: Mean 3.7 (members’ scores: 4, 4, 4, 4, 2.5)
- Cost: Mean 2.4 (members’ scores: 2, 3, 3, 2, 2)
- Time to study completion: Mean 2.6 (members’ scores: 2, 3, 4, 2, 2)

OVERALL MERIT SCORE: 3.0

Commentary

This case-control option scored better than the MISCAN option on value of evidence, although slightly lower in this regard than for the ecological design. Also, ease of implementation and risk of unsuccessful implementation scored lower than for the MISCAN or ecological design. The least favourable scores were for cost, which were at the costly end of the scale, and for time expected to undertake the study. Mostly, members considered that the study would take two or three years to complete, so long as there were not undue delays gaining HREC and administrative approvals. Time would be needed to accrue enough study participants and undertake the required interviewing and data linkage with BreastScreen, Medicare and other records.

An advantage perceived with this design was that the surrogate endpoint would occur earlier than deaths, increasing the probability of finding screening effects early. Risk factor data, as related to family history, genetic predisposition, reproductive history, exposure to hormone replacement therapy, and history of benign breast disease, could be obtained through self-report, with confirmation of self-reported mammography data through Medicare and BreastScreen records. This would facilitate adjustment for confounding in the analysis.

Because consent could be sought to check Medicare and other records, HREC and administrative approvals would be facilitated. Approvals could also be sought to search treatment records such that confounding from treatment differences could be addressed. Another advantage was that where BreastScreen services had records of BreastScreen screening invitations, self-reporting of this exposure to invitations could be validated, and the analysis could be by intention to screen, as well as by screening participation, enabling assessment of, and allowance in the analysis for participation bias.

This design also was seen to have a number of disadvantages. Although a strong correlation of most severe prognostic category (eg, least favourable Nottingham Prognostic Index score) with risk of breast-cancer death would be expected, this would be an imperfect measure of mortality, in that not all deaths would be encompassed by the surrogate measure, nor would all cases identified by the surrogate measure end up dying from breast cancer.
Because cancer registries do not routinely collect data on tumour diameter, grade or nodal status, this collection would need to be negotiated and funded, along with the ‘fast-tracking’ needed of data collection to meet study timelines.

It was also evident that biases from differences in recall, length and lead time could not be ruled out. If participation rates were not high, there would be questions about the potential for selection bias. As data on risk factors would be ascertained at time of diagnosis, recall bias may be introduced. Externality effects, as described in the section on the ecological design, may not be addressed, which could be a disadvantage of all analytic designs. A key concern was that due to prospective collection of prognostic data, this case-control option would not address effects of BreastScreen screening on breast-cancer outcomes for cancers diagnosed prior to 2007.
OPTION 4: CASE-CONTROL STUDY OF ASSOCIATION OF BREASTSCREEN EXPOSURE WITH RISK OF DISTANT METASTATIC RECURRENCE OF BREAST CANCER

Scores

- Value of evidence: Mean 3.6 (members’ scores: 4, 3, 4, 3.5, 3.5)
- Ease of implementation: Mean 4.2 (members’ scores: 5, 4, 4, 4, 4)
- Risk of unsuccessful implementation: Mean 3.5 (members’ scores: 4, 3, 4, 2.5, 4)
- Cost: Mean 2.4 (members’ scores: 2, 3, 3, 1.5, 2.5)
- Time to study completion: Mean 2.6 (members’ scores: 2, 3, 4, 2, 2)

OVERALL MERIT SCORE: 3.2

Commentary

This case-control option was rated similarly to the earlier option based on prognosis at diagnosis, except that the value of evidence was rated more favourably. Again, the least favourable scores were for cost and time expected to undertake the study. Mostly, members considered that the study would take two to three years to complete, so long as there were not undue delays gaining HREC and administrative approvals. Time would be needed to accrue enough study participants and undertake associated record linkages.

As for the previous design option, an advantage would be that the surrogate endpoint would occur earlier than deaths, increasing the probability of finding screening effects early. However, this endpoint may be less subject to lead-time and length-time bias than a surrogate marker obtained at diagnosis.

Again, risk-factor data, as related to family history, genetic predisposition, reproductive history, exposure to hormone replacement therapy, and history of benign breast disease, could be obtained through self-report, and self-reported screening mammography invitations and exposure confirmed through BreastScreen records and Medicare. As consent could be obtained to check Medicare records for past exposure to screening mammography, HREC and administrative approvals would be facilitated. Consent also could be sought to search treatment records, such that it may be possible to address the potential for confounding from treatment differences.

Where BreastScreen had a record of screening invitations, self-reporting of these invitations could be validated, and the analysis could be by intention to screen, as well as by screening participation, enabling assessment of, and allowance for participation bias.

A major disadvantage could be a high risk of under-ascertainment of recurrence cases, with the potential for statistical power to be reduced and selection bias to occur. As in the previous case-control design, ‘externality’ effects may not be addressed and recall bias may be introduced.
OPTION 5: CASE-CONTROL STUDY OF ASSOCIATION OF BREASTSCREEN EXPOSURE WITH RISK OF BREAST-CANCER MORTALITY

Scores

- Value of evidence: Mean 4.0 (members’ scores: 5, 3, 4, 4, 4)
- Ease of implementation: Mean 3.9 (members’ scores: 4, 3, 5, 4, 3.5)
- Risk of unsuccessful implementation: Mean 3.8 (members’ scores: 4, 3, 4, 4, 4)
- Cost: Mean 2.9 (members’ scores: 3, 3, 4, 1.5, 3)
- Time to study completion: Mean 3.0 (members’ scores: 3, 3, 4, 2, 3)

OVERALL MERIT SCORE: 3.6

Commentary

This case-control option was rated higher on value of evidence than the preceding designs because mortality would be measured directly without the need for the underlying assumptions associated with intermediate endpoints. In other respects, ratings were similarly to those for other case-control designs but lower than for the MISCAN or ecological approach. Most members felt that this study design could be implemented in about two years, so long as there were not undue delays gaining HREC and administrative approvals, although there was the potential for it to extend to three years. A mid-range cost was estimated. Since interviews of study participants would not apply, the completion date may be earlier than for the other case-control designs.

Opportunity for approval to use Medicare records may be reasonable, since there would not be interviews of the study population, and analysis of a de-identified data file would be possible. Where BreastScreen had a record of BreastScreen invitations, the analysis could be by intention to screen, as well as by screening participation, enabling assessment of, and allowance for screening participation bias.

Perceived limitations included the later occurrence of the mortality endpoint than of surrogate measures, reducing the probability of finding screening effects early. Also risk factor and mammography exposure data could not be obtained from cases through self-report, increasing the potential for bias from non-equivalence of cases and controls. Treatment effects could not be measured, unless approval could be gained to search treatment files without consent. Potentially, opportunity to search treatment files may be obtained in some but not all jurisdictions, in which case supplementary analyses of treatment data could be restricted to the former jurisdictions.

Consent could not be sought for checking Medicare records of deceased people for mammography exposure, increasing the potential for difficulties in gaining HREC and administrative approvals. To facilitate linking with Medicare records, members felt that it would be possible for the Australian Institute of Health and Welfare (AIHW) to gain a linkage agreement along the same lines as gained by the WA linkage facility. The AIHW could then produce a de-identified linked file for analysis. Alternatively, the WA linkage facility, which already has a memorandum of understanding to link to Medicare data, might be engaged for this purpose.
OPTION 6: RETROSPECTIVE COHORT STUDY OF ASSOCIATION OF BREASTSCREEN EXPOSURE WITH RISK OF BREAST-CANCER MORTALITY

Scores

- Value of evidence: Mean 3.9 (members’ scores: 4, 3, 4, 4, 4.5)
- Ease of implementation: Mean 3.1 (members’ scores: 2, 3, 3, 4, 3.5)
- Risk of unsuccessful implementation: Mean 2.4 (members’ scores: 2, 2, 2, 2.5, 3.5)
- Cost: Mean 2.0 (members’ scores: 2, 2, 2, 2, 2)
- Time to study completion: Mean 2.0 (members’ scores: 3, 2, 2, 1, 2)

OVERALL MERIT SCORE: 2.9

Commentary

This design was given a similar rating for value of evidence as the case-control study of mortality. In every other regard, it attracted less favourable ratings in that it was considered more difficult to implement, at higher risk of unsuccessful implementation, and of higher cost. Generally, members felt that it would take around three years to implement, so long as there were not undue delays gaining HREC and administrative approvals.

It was thought that the cohort design would be more credible to researchers who were wary of bias from case-control designs. By including all women (at least those on the Electoral Rolls), the chance of sampling error would be reduced. Where BreastScreen had a record of BreastScreen invitations, the analysis could be by intention to screen, as well as by screening participation, allowing assessment of, and adjustment for screening participation bias. Also to the extent that BreastScreen collected data on family history and other risk factors, adjustment could be made for these factors in sub-analyses of screened women.

It was recognised that data availability would vary by jurisdiction. National analyses would need to be limited to data items routinely available, but supplementary analyses could be undertaken for jurisdictions where a greater range of items existed.

Major limitations of this design also were perceived. Consent could not be sought for checking Medicare records for mammography histories, the NCSCH, National Death Index, cancer registries, and BreastScreen records of deceased people, increasing the potential for difficulties in gaining HREC and administrative approvals.

Consent also would not be realistic for checking records of people not deceased, given the size of the study group. Again, without consent, the potential for difficulties in gaining HREC and administrative approvals would be increased. As for previous designs, members felt that it may be appropriate to engage the AIHW to produce a de-identified file for analysis, or the WA linkage unit, which already has a memorandum of understanding with the Commonwealth to link Medicare data.
A key problem was seen to be the likelihood that few women uninvited to BreastScreen screening would exist on the Electoral Rolls, except for the period of initial rollout, limiting the period available for comparison. However this drawback was seen to apply for all designs when making comparisons by invitation to screen.

It was anticipated that other difficulties may occur classifying women by history of BreastScreen invitation/screening in jurisdictions without centralised record systems. Manual searching of records would not be realistic, given the size of the study group. It may prove necessary to restrict the scope of the study to those jurisdictions with the most accessible records, in order to overcome such difficulties.

Nonetheless the task of linking study group records to Medicare, NCSCH, and National Death Index records would be a huge undertaking, greatly magnified by the size of the study group. Also, risk-factor data may not be uniformly available across BreastScreen services and such data, and information on mammography exposure outside BreastScreen, could not be obtained through self-report, increasing the potential for bias from non-equivalence of comparison groups. Without consent, access to treatment records may not be possible, such that confounding of results from differences in clinical care could not be addressed.

Overall, members felt that implementation of this design would be a major undertaking, with little to be gained in addition to that achievable through a smaller well-designed case-control study.
OPTION 7: PROSPECTIVE COHORT STUDY OF ASSOCIATION OF BREASTSCREEN EXPOSURE WITH RISK OF BREAST-CANCER MORTALITY

Scores

- Value of evidence: Mean 4.8 (members’ scores: 5, 5, 5, 4, 5)
- Ease of implementation: Mean 3.0 (members’ scores: 3, 3, 2, 2, 5)
- Risk of unsuccessful implementation: Mean 2.8 (members’ scores: 3, 3, 1, 2, 5)
- Cost: Mean 1.0 (members’ scores: 1, 1, 1, 1, 1)
- Time to study completion: Mean 1.0 (members’ scores: 1, 1, 1, 1, 1)

OVERALL MERIT SCORE: 2.9

Commentary

This option scored higher on value of evidence than designs 1-6, but unfavourably on cost and time to completion. Advantages of this design include the opportunity to obtain consent for checking records of Medicare for exposure to mammography, the NCSCH, and the National Death Index, which would increase the potential for gaining HREC and administrative approvals.

Disadvantages included the massive scale of the undertaking, the high cost and need for a long period of over 10 years before useful data could be reported. More particularly, it would not be possible to evaluate screening effects from the time of commencement of BreastScreen screening. There would be few uninvited women in the comparison unexposed group available from the Electoral Rolls, once the Program had fully developed, such that comparisons by invitation status may not be feasible. This might be facilitated using Medicare records for sampling, although comparisons could be biased (e.g., women not on the Electoral Rolls may include a disproportionate number of overseas-born women at low risk of breast cancer).

The task of linking study group records to the Medicare, NCSCH, and National Death Index records would be greatly magnified by the size of the study group.
OPTION 8: CASE-COHORT STUDY OF ASSOCIATION OF BREASTSCREEN EXPOSURE WITH REDUCED RISK OF BREAST-CANCER MORTALITY

Scores

- Value of evidence: Mean 4.7 (members’ scores: 5, 4, 5, 5, 4.5)
- Ease of implementation: Mean 3.5 (members’ scores: 3, 4, 2, 4, 4.5)
- Risk of unsuccessful implementation: Mean 2.8 (members’ scores: 3, 4, 1, 2, 4)
- Cost: Mean 1.6 (members’ scores: 2, 2, 1, 1, 2)
- Time to study completion: Mean 1.4 (members’ scores: 2, 2, 1, 1, 1)

OVERALL MERIT SCORE: 3.1

Commentary

The value of evidence from this design was rated at the top of the range, similar to that from the prospective cohort study. While the scale of the undertaking would be lower than for the traditional prospective design, facilitating implementation and reducing cost, the cost and time estimated to completion still received unfavourable scores. Some members felt that approximately 10 years would need to elapse before useful output could be obtained.

As for the traditional prospective design, advantages would include the opportunity to obtain consent for checking records of Medicare for exposure to mammography, the NCSCH, and the National Death Index, which would increase the potential for HREC and administrative approvals. Risk factor data could be collected for women who die from breast cancer, as well as from other women, which would facilitate statistical control.

By appropriate allowance for the sampling, accurate relative risks of breast-cancer death could be calculated by screening invitation/screening exposure without the need to collect data from the great majority of women who did not experience a breast cancer.

A key disadvantage, however, would be that screening effects could not be evaluated for all cancers diagnosed since BreastScreen commencement in the 1990s. As for the traditional prospective design, there would be few uninvited controls available from the Electoral Rolls, increasing the desirability of using Medicare records as an alternative sampling source.
OPTION 9: COMPARISON OF PROGNOSTIC CHARACTERISTICS AND SURVIVAL FOR FEMALE BREAST CANCERS ACCORDING TO WHETHER DETECTED BY SCREENING MAMMOGRAPHY

Scores

• Value of evidence: Mean 2.3 (members’ scores: 2, 3, 1, 2.5, 3)
• Ease of implementation: Mean 3.5 (members’ scores: 3, 3, 4, 3, 4.5)
• Risk of unsuccessful implementation: Mean 3.4 (members’ scores: 3, 4, 4, 4, 2)
• Cost: Mean 4.0 (members’ scores: 3, 4, 4, 5, 4)
• Time to study completion: Mean 2.9 (members’ scores: 2, 3.5, 4, 2, 3)

OVERALL MERIT SCORE: 2.9

Commentary

This design scored poorly on value of evidence, although favourably on cost and (less so) on ease of implementation and risk. The study duration was likely to be about two years or less, so long as there were not undue delays gaining HREC and administrative approvals.

Advantages of this design were seen to be that reasonably early indications of potential effects of BreastScreen exposure on mortality could be obtained, whereas disadvantages were that results would be liable to bias from lead-time effects, length time, and potentially over-diagnosis, and that information on treatment and some TNM staging characteristics may be difficult to retrieve. The value seen in this study design was the potential to use it as an adjunctive design to gain quick interim results in advance of completing a more rigorous investigation.

Information on Medicare-funded and other externally funded mammograms would be needed for statistical adjustment. Consent could not be obtained for access to these data from deceased women, raising the potential for HREC and administrative barriers to gaining access to these data. Again, if this design option were implemented, it would be appropriate to engage the AIHW or WA linkage unit to develop a de-identified linked file for analysis.

When discussing this design option, members also considered the utility of the 1995 national breast-cancer management survey, which had collected data on method of cancer detection, plus prognostic indicators and treatment. While it would be possible in principle to use this dataset to assess the association of mammography detection with prognostic indicators, and with clinical management, interpretation would be complicated by a lack of identification of interval cancers, and by the potential for bias from lead time, length time and over-diagnosis. Follow-up survivals had not been traced in all jurisdictions and likely would be impossible to obtain nationally at this time, since the national file did not include person-identity details.

Nonetheless, where such linking had been undertaken, and where there were parallel studies using other data sources, the results could be used. ETG members were aware of at least two publications that compared prognosis and treatment characteristics of screen-detected and other breast cancers in Australia.
**FINAL SYNTHESIS OF MERITS OF OPTIONAL DESIGNS**

The rank order of design options by overall merit scores was as follows:

- **4.1 – Design Option 2**: Ecological study
- **3.6 – Design Option 5**: Case-control study with mortality endpoint
- **3.6 – Design Option 1**: MISCAN model (ranked below Design Option 5 due to lower value of evidence)
- **3.2 – Design Option 4**: Case-control study with metastatic recurrence as endpoint
- **3.1 – Design Option 8**: Case-cohort study
- **3.0 – Design Option 3**: Case-control study with diagnostic surrogate marker as endpoint
- **2.9 – Design Option 6**: Retrospective cohort study
- **2.9 – Design Option 7**: Conventional prospective cohort study (ranked below Design Option 6 due to unfavourable timeliness)
- **2.9 – Design Option 9**: Comparison of screen-detected with other cancers (ranked below Design Options 6 and 7 due to lower value of evidence)

A summary of the advantages, disadvantages and design criteria scores for each of the considered design options is outlined in Table 4.
Table 4: Summary of design criteria for each considered design option

<table>
<thead>
<tr>
<th>Study design</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Value of evidence</th>
<th>Ease of implementation</th>
<th>Risk</th>
<th>Cost</th>
<th>Time to complete</th>
<th>Overall summary score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. MISCAN – MISCAN micro-simulation of BreastScreen effect on breast-cancer mortality</td>
<td>• Most source data readily available from the ABS, the National Cancer Statistics Clearing House, clinical cancer registries, and BreastScreen (BS) • The MISCAN model has been widely used to indicate screening effects in other settings that have been broadly similar to the results of trials • No need for person-identified data • Inexpensive • Quick</td>
<td>• Theoretical simulations have given widely different estimates of screening effect in other settings, depending on their structure and assumptions, reducing the credibility of results • Local applicability uncertain • Scepticism among some experts about the accuracy of results from simulation models</td>
<td>2.4</td>
<td>4.5</td>
<td>4.5</td>
<td>4.6</td>
<td>4.4</td>
<td>3.6</td>
</tr>
<tr>
<td>2. Ecological Geographic – Ecological study of association of BreastScreen participation with breast-cancer mortality by SLA</td>
<td>• Proof-of-concept already established for this type of design, both in NSW and overseas • Measures outcome at a population/program level and may capture indirect effects as well as direct screening effects • Data likely to be readily available • No need for person-identified data, facilitating HREC approval • Inexpensive • Relatively quick</td>
<td>• Variations in BS participation may be smaller by SLA than ideal for study purposes • Uncertainty about time lags between screening and mortality effects, requiring modelling or “a priori” assumptions • Possible difficulties identifying histories of BS screening in jurisdictions without centralised record systems (eg, NSW) • Data on breast-cancer risk factors not available by SLA requiring assumptions of equivalence • Possible bias from “ecological fallacy”</td>
<td>3.3</td>
<td>4.8</td>
<td>4.4</td>
<td>5.0</td>
<td>4.5</td>
<td>4.1</td>
</tr>
<tr>
<td>3. Case-Control NPI – Case-control study of associations of BreastScreen invitation for screening, and of BreastScreen participation, with a surrogate diagnostic marker of breast-cancer mortality (eg, Nottingham Prognostic Index)</td>
<td>• Earlier surrogate endpoint than mortality, increasing statistical power and the probability of finding screening effects early • Risk-factor data and mammography data could be obtained through self-report, with confirmation of self-reported mammography data through Medicare and BS records • Consent could be sought to check Medicare records, facilitating HREC approval • Where the BS service has a record of BS invitations, the analysis could be by intention to screen, as well as by screening participation, enabling assessment of, and allowance in the analysis for participation bias</td>
<td>• Although a strong correlation of the most severe Nottingham Prognostic Index (NPI) category with risk of breast-cancer death would be expected, not all deaths would be covered by this surrogate measure, nor would all cases identified by the surrogate measure end up dying from breast cancer. • Length and lead-time biases could not be excluded • Could introduce selection and recall bias • External effects may not be addressed • The study would be more time consuming and more expensive than a case-control study that did not involve interviews and prospective accrual</td>
<td>3.1</td>
<td>3.8</td>
<td>3.7</td>
<td>2.4</td>
<td>2.6</td>
<td>3.0</td>
</tr>
</tbody>
</table>
## Table 4: Summary of design criteria for each considered design option (continued)

<table>
<thead>
<tr>
<th>Study design</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Value of evidence</th>
<th>Ease of implementation</th>
<th>Risk</th>
<th>Cost</th>
<th>Time to complete</th>
<th>Overall summary score</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>4. Case-Control Recurrence</strong></td>
<td>- Earlier surrogate endpoint than mortality, increasing statistical power and the probability of finding screening effects early&lt;br&gt;- This endpoint may be less subject to lead-time and length-time bias than a surrogate marker recorded at diagnosis&lt;br&gt;- Risk-factor data and mammography data could be obtained through self-report and self-reported mammography screening invitations and exposure confirmed through BS records and Medicare&lt;br&gt;- Consent could be sought to check Medicare records, facilitating HREC approvals&lt;br&gt;- Where BS has a record of BS invitations, the analysis could be by intention to screen, as well as by screening participation, enabling assessment of participation bias</td>
<td>- There would be a high risk of under-ascertainment of recurrence cases, potentially compromising statistical power and resulting in selection bias and recall bias&lt;br&gt;- Primarily measures outcome at an individual level and may not capture programmatic effects&lt;br&gt;- The study would be more time consuming and more expensive than the aforementioned ecological study&lt;br&gt;- The surrogate measure would be an imperfect measure of breast-cancer death&lt;br&gt;- Could introduce selection and recall bias&lt;br&gt;- External effects may not be addressed</td>
<td>3.6</td>
<td>4.2</td>
<td>3.5</td>
<td>2.4</td>
<td>2.6</td>
<td>3.2</td>
</tr>
<tr>
<td><strong>5. Case-Control Mortality</strong></td>
<td>- The mortality endpoint directly addresses the impact of the program on mortality&lt;br&gt;- The study would be nested within a theoretical retrospective cohort and data would be obtained through linkage (i.e., BS, death registry and cancer registry, and Medicare) so that information would be relatively complete. Possibility of selection and recall bias therefore significantly reduced&lt;br&gt;- Where BS has a record of invitations, the analysis could be by intention to screen, as well as by screening participation, allowing assessment of screening participation bias&lt;br&gt;- Potential for lead and length-time bias reduced as mortality is the endpoint, not an earlier surrogate measure&lt;br&gt;- Statistical power using all available cases should be high&lt;br&gt;- More efficient and less costly than a retrospective cohort study</td>
<td>- Risk-factor data and mammography exposure data could not be obtained from cases through self-report, increasing the potential for bias from non-equivalence of cases and controls (e.g., healthy-screening participation bias)&lt;br&gt;- Primarily measures outcome at an individual level and may not capture programmatic effects&lt;br&gt;- Consent could not be sought for checking Medicare records of deceased people for mammography exposure, increasing the potential for difficulties in gaining HREC and administrative approvals</td>
<td>4.0</td>
<td>3.9</td>
<td>3.8</td>
<td>2.9</td>
<td>3.0</td>
<td>3.6</td>
</tr>
</tbody>
</table>
### Study design

<table>
<thead>
<tr>
<th>Study design</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>6. Retrospective Cohort</strong>&lt;br&gt;Retrospective cohort study to: (1) compare death rates from breast cancer in women who were invited by BreastScreen to screening with those not invited to screening; and (2) compare death rates from breast cancer in women who participated in BreastScreen with those who did not participate in BreastScreen</td>
<td>• By including all screened women (at least those on the Electoral Rolls) the chance of sampling error would be reduced&lt;br&gt;• Where BS had a record of BS invitations, the analysis could be by intention to screen, as well as by screening participation, allowing assessment of and adjustment for screening participation bias&lt;br&gt;• To the extent that BS collected data on family history and other risk factors, some adjustment could be made for these factors in sub-analyses of screened women&lt;br&gt;• Less susceptible to bias than some case-control studies, particularly designs 3 and 4, but results likely to be equivalent to design 5</td>
<td>• Consent could not be sought for checking Medicare, the National Cancer Statistics Clearing House, National Death Index, and BS records of deceased people, increasing the potential for difficulties in gaining HREC and administrative approval&lt;br&gt;• Possible difficulties would occur classifying women by histories of BS invitation/screening in jurisdictions without centralised record systems (eg, NSW)&lt;br&gt;• Linking study group records to the Medicare, the National Cancer Statistics Clearing House, Cancer Registries and National Death Index records would be logistically very difficult&lt;br&gt;• Risk-factor data may not be uniformly available across BS services&lt;br&gt;• Risk-factor data and mammography exposure data could not be obtained through self-report, increasing the potential for bias from non-equivalence of comparison groups (eg, healthy screening participation bias)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value of evidence</th>
<th>Ease of implementation</th>
<th>Risk</th>
<th>Cost</th>
<th>Time to complete</th>
<th>Overall summary score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.9</td>
<td>3.1</td>
<td>2.4</td>
<td>2.0</td>
<td>2.0</td>
<td>2.9</td>
</tr>
</tbody>
</table>

| **7. Cohort Prospective**<br>Prospective cohort study to: (1) compare death rates from breast cancer in women who were invited by BreastScreen to screening with those not invited to this screening; and (2) compare death rates from breast cancer in women who participated in BreastScreen with those who did not participate in BreastScreen | • The prospective cohort design would facilitate the collection of high-quality data and better quality evidence than retrospective designs<br>• Consent could be sought for checking records through Medicare, the National Cancer Statistics Clearing House, and the National Death Index, increasing the potential for gaining HREC and administrative approval<br>• Possibility of bias greatly reduced (i.e., selection and recall bias)<br>• Comprehensive risk-factor data could be collected prospectively | • High cost and substantial logistic difficulties due to prospective enrolment of tens of thousands of women<br>• The study duration would be extensive (i.e., between 5-10 years) to observe sufficient deaths for adequate statistical power. If a surrogate endpoint is used, length and lead time bias could be a problem<br>• The task of linking study group records to the Medicare, National Cancer Statistics Clearing House, and National Death Index records would be greatly magnified by the size of the study group<br>• Loss to follow up would be a potential problem over the time period of the study<br>• Evaluation of program since inception in the 1990s would not be possible |

| 4.8               | 3.0                   | 2.8  | 1.0  | 1.0              | 2.9                  |
## Table 4: Summary of design criteria for each considered design option (continued)

<table>
<thead>
<tr>
<th>Study design</th>
<th>Advantages</th>
<th>Disadvantages</th>
<th>Value of evidence</th>
<th>Ease of implementation</th>
<th>Risk</th>
<th>Cost</th>
<th>Time to complete</th>
<th>Overall summary score</th>
</tr>
</thead>
</table>
| 8. Case Cohort – Case-cohort study of association of exposure of individual women to a BreastScreen invitation for screening, and to BreastScreen screening participation, with risk of breast-cancer mortality | - The case-cohort design would facilitate the collection of high-quality data prospectively  
- Consent could be sought for checking records of BS, Medicare, the National Cancer Statistics Clearing House, and the National Death Index, in prospective recruitment which should increase the potential for gaining HREC and administrative approval  
- Risk-factor data could be collected prior to women dying from breast cancer, as well as for other women  
- Allowing for sampling, accurate relative risks of breast-cancer mortality could be calculated by screening invitation/screening exposure without the need to collect data from the great majority of women who did not later die from breast cancer  
- The task of linking study group records to Medicare, the National Cancer Statistics Clearing House and National Death Index records would be less onerous than for the larger numbers involved in a traditional cohort design | - High cost, although if only a sub-cohort is followed (as proposed), not as high as for a traditional cohort design  
- Potential selection bias from collecting risk-factor data for the majority of women with breast cancer post diagnosis  
- The study duration would need to be extensive (i.e., between 5-10 years) to observe sufficient deaths for adequate statistical power. If a surrogate endpoint were used, length and lead-time biases could be a problem  
- Evaluation of program since its inception in the 1990s would not be possible | 4.7 | 3.5 | 2.8 | 1.6 | 1.4 | 3.1 |
| 9. Assessing association of BreastScreen participation with breast-cancer fatality – A comparison of prognostic characteristics and survival for female breast cancers according to whether detected by screening mammography | - Early inferences of potential effects on mortality could be gained  
- Relatively quick and logistically feasible  
- Information on treatment and some TNM staging features may be difficult to retrieve | - Results prone to bias from lead-time effects, length time and over-diagnosis  
- Information on Medicare-funded and other externally funded mammograms would be needed for statistical adjustment. Consent could not be obtained for access to these data from deceased women, complicating HREC approval and access to these data | 2.3 | 3.5 | 3.4 | 4.0 | 2.9 | 2.9 |
PROPOSALS

PROPOSED STUDY DESIGNS FOR EVALUATING BREASTSCREEN

A scenario consistent with these scores would be to undertake an ecological study to investigate effects of BreastScreen exposure at a population level. In addition, a case-control study could be pursued, based on a mortality endpoint, to investigate effects at an individual level. These studies would be complementary. Broader indirect effects of the screening program (i.e., externality effects) are more likely to be captured through the ecological than case-control design. Meanwhile, a carefully designed case-control study would produce high quality evidence regarding the effects of the Program on mortality from breast cancer at an individual level.

A parallel population survey also is proposed to compare exposures to de facto screening (outside BreastScreen) and risk factors among BreastScreen invitees, participants and other women, in order to assess the potential for bias in the case-control results and allow adjustment of the case-control results for such bias.

PROPOSED ECOLOGICAL STUDY

Title
An ecological study of the association of BreastScreen participation with breast-cancer mortality by SLA.

Aim
To determine associations by residential SLA of BreastScreen participation with: i) time-lagged breast-cancer mortality; or ii) time-lagged annual changes in breast-cancer mortality.

The time lag would either be: set a priori using mean survival periods available from the National Cancer Statistics Clearing House (NCSCH); or obtained through modelling breast-cancer mortality on participation rates to identify the time lag that gave the best fit to the data.

There would be adjustment in the analysis for numbers of mammograms provided outside BreastScreen, relative socio-economic disadvantage of residential area (e.g., SEIFA index), health-service accessibility (e.g., ARIA index), parity estimates using perinatal statistics, other socio-demographic descriptors such as ethnicity mix, State/Territory jurisdiction, and breast-cancer mortality rates for cases diagnosed in the 1980s before the advent of screening mammography.

Although the focus would be on BreastScreen effects of screening 50–69 year olds, it was considered that parallel analyses directed at effects of screening 40–49 year olds, and 70+ year olds, respectively, could be done at little additional cost and would provide comparative data that would assist interpretation of the results for the 50–69 year principal screening target. Furthermore, opportunities could be explored to collect additional data on potential confounders (e.g., risk factors and treatment) in jurisdictions where such data were relatively accessible (i.e., more in-depth local investigations could be nested within the broader national study).
Process outline

- SLA-specific data would be retrieved for the 1990–2005 period for 50–69 year olds, and potentially 40–49 year olds and those aged 70+ years, as relating to:
  - BreastScreen participation (data from BreastScreen)
  - population data and socio-demographic descriptors (interpolations from ABS census data)
  - privately funded mammograms (data from Medicare)
  - mammograms provided to public hospital inpatients (data from hospital separations).

- All SLA areas in Australia would be included in the study to maximise statistical power.

- Annual SLA-specific breast-cancer death data also would be obtained from the NCSCH (or State/Territory cancer registries) by age for diagnoses occurring in 1990–2005 among 50–69 year olds resident in the respective SLA at diagnosis, and potentially 40–49 year olds and those aged 70+ years.

- Breast-cancer mortality rates for diagnoses occurring in 1983–89 among 50–69 year olds resident in the respective SLA at diagnosis, and potentially 40–49 year olds and those aged 70+ years (data from the NCSCH).

- Breast-cancer death rates or annual reductions in breast-cancer mortality would be modelled by SLA according to BreastScreen participation, with adjustment for other characteristics that conditioned BreastScreen risk estimates, such as SEIFA index, ARIA index, State/Territory jurisdiction, other socio-demographic descriptors, mortality rates for breast cancers diagnosed in the 1980s, and mammogram exposure outside BreastScreen. This could involve a Poisson regression analysis.

Timelines

Suggested timelines for the ecological study are as follows:

- **1–6 months**: seek approvals and obtain data from all jurisdictions and sources; finalise study and analysis protocols

- **7–12 months**: data management and analysis

- **13–18 months**: report preparation and final meeting with Steering Committee.

Budget estimate

A ‘ball park’ budget estimate for the ecological study would be between $200,000 and $300,000, depending on the operational complexities encountered and the extent of inclusion of more in-depth nested studies in some jurisdictions where a broader range of data on treatment and/or risk factors was available. This budget estimate would cover salary costs for a part-time project manager, project officer, data manager and statistician, charges for data from the respective agencies, and a 20% on-cost allowance.

Discussion

This ecological proposal was seen to be a comparatively straightforward means of gaining early results, with proof of concept already provided by a previous NSW study (Taylor et al., Cancer Causes Control 2004; 15: 543-50), and by studies in other populations. Identified data would not need to be released.
from databases, which would facilitate HREC and administrative approvals. Results from such a study should be available in 18 months or less, so long as there were not undue delays gaining HREC and administrative approvals.

A number of operational issues would need to be addressed. For example, criteria would need to be developed for distinguishing between Medicare-funded de facto screening mammograms and diagnostic mammograms. The successful methodology developed in other population settings for this purpose would be used as a guide, including the methodology reported by Smith-Bindman et al. in *Medical Care* 2006; 44: 463-70.

An additional issue would be the potential value of adjusting for Medicare and other externally funded mammograms, which took place for de facto screening and in response other risk factors that lead to breast investigations. The analyses could thus account for the effect of screening outside the Program, and potentially other confounders, so that conclusions could be drawn about the effect of BreastScreen in isolation.

Care also would be needed in study implementation to accommodate SLA boundary changes. One approach could be to repeat analyses, first including and then excluding areas where SLA changes had occurred, in order to assess effects on results. Another approach would be to circumvent SLA changes by combining selected ones. Either way, it was anticipated that means would exist to accommodate these issues.

A further requirement of the study would be sufficient variation in BreastScreen participation by SLA to evaluate effects on mortality. During the screening rollout, considerable variability would have existed in all jurisdictions, such that there should not be difficulties satisfying this requirement.

It was considered that this study design could show that mortality from breast cancer was lower in areas where BreastScreen participation was higher, which would be indicative of a BreastScreen effect.
PROPOSED CASE-CONTROL STUDY WITH MORTALITY ENDPOINT

Title

Case-control study of association of exposure to a BreastScreen invitation for screening, and to BreastScreen participation, with risk of breast-cancer mortality.

Aim

To determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with breast-cancer mortality in women diagnosed with breast cancer when 50–69 years of age in 1994–2005.

Adjustment would be made for exposure to other characteristics that might influence BreastScreen invitation/participation, such as exposure to Medicare-funded and other externally funded mammograms, SEIFA index, ARIA index, other socio-demographic descriptors, and State/Territory jurisdiction.

Although the focus would be on associations of BreastScreen exposure with mortality from breast cancers diagnosed in 50–69 year olds, it was considered that parallel analyses for 40–49 year olds, and 70+ year olds, respectively, could be done at little additional cost and would provide comparative data for interpreting results for the 50–69 year principal screening target. Furthermore, opportunities could be explored to collect additional data on potential confounders (e.g., risk factors and treatment) in jurisdictions where such data were relatively accessible (i.e., more in-depth local investigations could be nested within the broader national study).

At the plenary session of the Steering Committee and ETG, this study design was considered further. It was decided that all females with a breast-cancer death should be included, so long as they were recorded on the Electoral rolls and met the inclusion criteria specified below in the Process outline. An equal number of matched controls could be selected rather than three controls per case. This should still achieve close to optimum statistical power to detect small differences in outcome according to BreastScreen exposure and allow testing of variations in outcome by State/Territory jurisdiction. In addition, secular differences in associations of BreastScreen exposure with mortality could be investigated.

Process outline

- State/Territorial cancer registries would identify breast-cancer deaths occurring in 1994–2005 where:
  - the person was diagnosed in 1994 or later
  - age at diagnosis was 40–49, 50–69 or 70+ years respectively
  - the person was recorded on the Electoral Rolls at time of death.

- One control would be selected at random per case from among women with the same date of birth on the Electoral Rolls, who were still alive at the time of death of the case, and where there was no evidence from State/Territorial registries of a breast-cancer diagnosis before the date of diagnosis of the case.
• State/Territory BreastScreen services would search their records and classify cases according to whether they had been: i) invited for screening (y/n); and ii) screened (y/n), in the 3 years prior to diagnosis, or earlier. Controls would be similarly classified for the same time ‘window’ as their respective cases.

• Exposure to mammograms outside BreastScreen would be investigated using Medicare and if feasible, hospital separation data.

• A standard case-control analysis would be performed to determine the association of: i) BreastScreen screening invitation; and ii) BreastScreen screening participation, with being a case (i.e., breast-cancer death) as opposed to a control (i.e., two separate analyses). The relative odds of being a case would be determined for these predictors, adjusting for other characteristics that conditioned risk estimates for BreastScreen invitation/participation, such as exposure to mammograms outside BreastScreen, SEIFA and ARIA indices (drawn from residential postcodes), and State/Territory jurisdiction. Interaction terms could be used to investigate differences in outcome by State/Territory jurisdiction and by calendar year groupings. Conditional logistic regression analysis could be used for this purpose.

Statistical Power

A sample of up to 5,420 cases (breast-cancer deaths) and 5,420 matched controls would be anticipated where diagnoses would have occurred since 1994 and at 50–69 years of age. This would provide a statistical power of 96% to detect an odds ratio of 0.85 (versus the alternative equal odds using a Chi-Square test) at a 0.05 significance level. This assumes a correlation between the matched groups of 0.2 and a history of ever having BreastScreen exposure of 60% for the control group.

If the inclusion ages were broadened to included 40–74 year olds, about 16,980 cases and 16,980 matched controls would be expected, which would provide a statistical power of 100% to detect an odds ratio of 0.85.

Timelines

Suggested timelines for the case-control study are as follows:

• **1–6 months:** obtain appropriate clearances and approvals for data from all jurisdictions; submit ethics applications to Commonwealth/State/Territory jurisdictions; finalise study protocol; submit ethics application for linking to Medicare database

• **7–12 months:** establish database; data management and cleaning of BreastScreen data; linking to data from Medicare database (may extend beyond 12-month timeline)

• **13–24 months:** data analysis; report preparation and final meeting with Steering Committee; may extend to 36 months depending on operational complexities.
Budget estimate

A ‘ball park’ budget estimate for the case-control study would be between $550,000 and $650,000, depending on the operational complexities encountered and the extent of inclusion of nested in-depth studies in some jurisdictions. This budget estimate would cover salary costs for a part-time project manager, project officer, data manager and statistician, charges for data from the respective agencies, and a 20% on-cost allowance.

Discussion

Members preferred a case-control study with a mortality endpoint to an alternative design relying on a surrogate measure, since mortality was the outcome of interest. It was recognised that surrogate measures had a number of advantages, including their availability in advance of mortality, allowing earlier evaluation. Also, women who were alive and to whom a surrogate measure applied would be able to report on risk-factor exposures, and histories of de facto screening through Medicare, and could provide consent for researchers to access to records, which would facilitate HREC and administrative approvals. On the other hand, surrogate measures were imperfect measures of mortality that would be less credible. Depending on the study design, there would also be questions about effects of lead-time, length-time and selection bias when surrogate measures were used.

Members recognised that the linking of cases and controls to BreastScreen and Medicare records would be a major task. For this reason, the decision was made to reduce the numbers of controls per case from three to one. This would reduce the burden of linking, but would have little effect on statistical power, given the large numbers of cases that would be available.

Members considered that the proposed case-control study with a mortality endpoint would provide more valid results. It would be important, however, to link to Medicare records, despite absence of consent. Otherwise the study would only indicate the effectiveness of BreastScreen exposure against a comparator that included de facto screening.

The risk of not gaining approval for linking to Medicare records seemed small, although the time to approval may vary depending on ethical and bureaucratic processes. The Commonwealth already permits linkage of Medicare data with Western Australian health-service data. Through a prescribed process, researchers can receive the fully de-identified files they require for analysis. It seems unlikely that a similar approval would not be given with the AIHW. ETG members were aware of negotiations for this type of linkage between the AIHW, Department of Health and Ageing, and Medicare where the present case-control proposal could be a pathfinder project. While there may be the fallback option of having the linkage undertaken through the WA linkage unit, it seemed unlikely to members that the present negotiations would be unsuccessful and that this recourse would be necessary.

While Electoral Rolls have been proposed as the sampling frame for controls, it was considered a disadvantage that up to 10% of women would not be recorded on the Rolls. While this should not bias comparisons if cases also were restricted to those on the Rolls, the generalisation of results to other women not on the Rolls would be uncertain. Medicare records were also considered as a sampling frame, but access to the necessary Medicare information was regarded as potentially more problematical. In view of the high coverage of the population by the Electoral Rolls, and the likelihood of ready access to them, they were regarded as the preferred sampling frame.
A disadvantage of using a mortality outcome is the inability to obtain from deceased persons self-reported data on exposure to de facto screening or on risk factors related to family history of breast cancer, genetic predisposition, reproductive history, use of hormone replacement therapy, and history of breast surgery for non-malignant conditions. The option of seeking this information through proxy informants was not regarded by members to be realistic.

Members felt that a parallel population survey should be implemented to indicate differences between BreastScreen participants, invitees and other women by exposure to de facto screening and risk factors. The results could be used to indicate the potential for bias in the case-control results and permit adjustments to be made for this bias. Evidence from the published literature on effects on mortality of different risk factors would be used to guide this process.

It was considered that results of a case-control study could be that participation in BreastScreen was associated with a decreased risk of dying from breast cancer and that this effect was not due to more frequent screening outside BreastScreen. Results would also indicate the strength of this association.

While light could be shed on differences in protective association by socio-demographic and jurisdictional characteristics through the proposed ecological and case-control studies, such questions would be better addressed through separate modeling projects. Also, should a protective association be found for BreastScreen, further research would be needed to explore the extent to which this was due to the screening per se, as compared with possible improvements in clinical care associated with BreastScreen implementation.
PROPOSED RISK-FACTOR SURVEY

Aim

To establish a profile of Australian women by history of exposure to BreastScreen, to *de facto* screening funded externally to BreastScreen, and to breast-cancer risk factors.

Although the focus would be on 50–69 year olds, it was considered that parallel data collection for 40–49 year olds, and 70+ year olds, respectively, would provide comparative information, which along with the results of the case-control data, would facilitate interpretation of results for the 50–69 year principal screening target.

Process outline

- A national telephone survey would be undertaken using either the electronic white pages directory as a random sampling frame or potentially random digit dialling or an alternative methodology. A CATI (computer-aided telephone interviewing) process would be employed.
- Up to 10 separate call backs could be used, if needed, to recruit households.
- A non-replacement sample could be used, with a view to gaining 3,000 female respondents aged 40 years or more. Only one interview would be conducted per household. Where more than one woman was eligible, the one who was last to have her birthday could be selected.
- Interview questions would be pilot tested and implementation of the survey would be subject to quality assurance. Questions would cover:
  - socio-demographic descriptors such as age, ethnicity, country of birth, metropolitan, regional urban or rural place of residence, socio-economic status (household income/SEIFA index) and service access (ARIA index)
  - past exposure to BreastScreen (screening participation and invitations to be screened)
  - history of *de facto* screening funded externally to BreastScreen
  - risk-factor exposures, including those related to personal history of breast cancer, family history of breast cancer, past breast problems involving a biopsy or other surgical procedure, hormone replacement therapy, reproductive history, genetic susceptibility, and excess body weight.
- Univariate and multivariable analyses would be used to assess risk profiles of women by BreastScreen exposure and exposure to *de facto* screening. These analyses would be weighted by the inverse of the woman’s probability of selection and re-weighted to correspond with the estimated residential population of Australia by age, sex and place of residence.
Statistical power

Adequate statistical precision would be achieved by including 3,000 respondents. For example, a calculated proportion of 50% would have a 95% confidence interval of around 48% to 52% for the whole sample. For a sub-set of 1,000 respondents, the interval would be from approximately 47% to 53%. If the calculated proportion were lower at 10%, the 95% confidence interval would be around 9% to 11% for 3,000 respondents and 8% to 12% for a sub-set of 1,000 respondents.

Timelines

It is suggested that the survey take place concurrently with the proposed case-control study. The following timelines could apply:

- **1–6 months**: finalise study protocol and submit ethics application
- **7–12 months**: undertake survey
- **13–18 months**: data management and analysis
- **18–24 months**: Report preparation and final meeting with Steering Committee.

Budget estimate

This is estimated to be between $80,000 and $110,000 depending on the extent of questioning and the sampling routine chosen.

Discussion

Survey data would provide a risk-profile context for interpreting case-control results. The approach is ecological and methodologically inferior to obtaining risk-factor data directly from case-control study participants. It is a ‘fall back’ approach, given the impossibility of gaining this information directly from deceased women.

The value of the survey data is that, taken together with existing data on the strength of the associations between risk factors and breast-cancer mortality, they would allow a sensitivity analysis of the possible extent of confounding by performing external adjustment of the odds ratios.

Because exposure to some risk factors, such as exposure to hormone replacement therapy for example, may have varied over time, the risk-factor survey should, as far as possible, address exposures over the past 15 years in addition to current exposures.

The survey data could be biased to the extent that not all Australians would be accessible by telephone. This bias is not anticipated to be large, however, and it is expected that the risk-profile data obtained will greatly assist interpretation of the case-control results.
ISSUES NOT ADDRESSED THROUGH RECOMMENDED DESIGNS

This proposal addresses the evaluation of BreastScreen effects on breast-cancer mortality. Outcomes outside the scope of the project are outlined below.

- **Effects on morbidity:**
  - favourable effects would include: i) diagnosis of breast cancers at an earlier stage when they would be more amenable to breast-conserving surgery; ii) avoiding the morbidity associated with breast-cancer progression to a fatal outcome; iii) reassuring BreastScreen participants following screening of an absence of detected cancers; and iv) reassuring participants when detected tumours have a favourable prognosis.
  - unfavourable effects would include: i) unnecessary anxiety among women without breast cancer that screening or assessment outcomes may be a cancer diagnosis; ii) unnecessary investigation and follow-up of women without a cancer diagnosis; and iii) potentially the diagnosis and treatment of DCIS and possibly of small invasive cancers that would not have progressed, if left untreated.

- **Differential effects on mortality:**
  - effects of BreastScreen on mortality may vary by socio-demographic sub-groups classified by ethnicity, socio-economic status, service access and other characteristics; although some study design options considered in this project had the potential to reveal differences in effect, this was not a primary objective of the project.
  - effects of BreastScreen on mortality may vary by family history and other risk factors; again, although some of study design options had the potential to reveal different effects in this context, this was not a primary objective of this project.

- **Effects of new technologies:**
  - breast imaging is experiencing significant changes in technology, with the potential for different mortality outcomes; means of evaluating such differences were not addressed in this project.

- **Mechanisms that may underlie a protective association of BreastScreen exposure with mortality:**
  - apart from effects of screening per se on breast-cancer mortality, there is the potential for better outcomes from improvements along the screening pathway, including improvements in treatment over the time period of the study.
  - investigations of mechanisms underlying any protective association of BreastScreen exposure with mortality, and the apportioning of this association to screening per se as opposed to other factors, is outside the scope of this project; different customised design options would be needed for this purpose than the ones presented in this report, which are directed at assessing whether an association exists in the first place, and if so, on what scale.
RECOMMENDATIONS

It is recommended that:

1. The effects of the BreastScreen Australia Program on mortality should be assessed in two separate studies, one that assesses the effect of the Program on sub-populations within Australia, and one that assesses the effect at the level of individual women who participated in, or were invited to BreastScreen.

2. The study of sub-populations should assess the association of BreastScreen participation with breast-cancer mortality by Statistical Local Area (SLA).
   - The study of effects on individual women would compare rates of exposure to BreastScreen in women who died of breast cancer with rates in women who had not died of this disease. Two analyses would be undertaken in this case-control study, the first defining BreastScreen exposure to include receiving an invitation to BreastScreen screening, and the second restricting this definition to BreastScreen screening participation.
   - Concurrently with the case-control study, a population survey should assess how the exposure of women to BreastScreen has been related to (“diagnostic”) screening mammograms outside the Program and how differences in BreastScreen exposure are related to breast-cancer risk factors, medical care and other social and demographic factors. Because exposure to risk factors may vary over time, the aim would be to assess past as well as current exposures.
REFERENCES


INCLUDED PUBLICATIONS

Ref ID: 1156

Ref ID: 432

Ref ID: 2418

Ref ID: 1292

Ref ID: 3772

Ref ID: 1307

Ref ID: 1154
Ref ID: 3656

Ref ID: 1781

Ref ID: 2031

Ref ID: 493

Ref ID: 3532

Ref ID: 2593

Das B, Feuer EJ, Mariotto A. Geographic association between mammography use and mortality reduction in the US. *Cancer Causes Control* 2005;16(6):691–9.
Ref ID: 220

Ref ID: 69

Ref ID: 70

Ref ID: 1015

Ref ID: 2501

Ref ID: 180
Ref ID: 661

Ref ID: 2006

Ref ID: 748

Ref ID: 3715

Ref ID: 2177

Ref ID: 1954

Ref ID: 1663

Ref ID: 1399

Ref ID: 3687

Ref ID: 764

Ref ID: 823

Ref ID: 1630

Ref ID: 1912

Ref ID: 1349


Ref ID: 3804


Ref ID: 3926


Ref ID: 3857


Ref ID: 368


Ref ID: 686


Ref ID: 3750


Ref ID: 1014


Ref ID: 3500


Ref ID: 3888


Ref ID: 23

Ref ID: 2240


Ref ID: 3829


Ref ID: 749


Ref ID: 2696


Ref ID: 1838


Ref ID: 1320


Ref ID: 824


Ref ID: 687


Ref ID: 487


Ref ID: 2290


Ref ID: 3946


Ref ID: 3947
Ref ID: 1973

Ref ID: 2353

Ref ID: 2088

Ref ID: 2885

Ref ID: 2891
APPENDICES

APPENDIX A-1: TERMS OF REFERENCE OF THE STEERING COMMITTEE

AIM OF PROJECT
To review and assess the design options for evaluating the impact of the BreastScreen Australia program on breast-cancer mortality in the population with due regard for feasibility and quality of evidence.

TERMS OF REFERENCE
The Steering Committee has an overseeing role focused on the strategic directions of the project, including:

• definition of the objectives of the literature review
• critical review and approval of project methodology
• critical review and comment on draft project reports
• approval of the final project report prior to submission to the Department of Health and Ageing.

PROJECT DURATION AND MEETING FREQUENCY
This project will start in July 2006 and is expected to take approximately 6 months to complete. During this time, it is expected that Steering Committee members will be invited to attend up to three meetings, which may be held face-to-face or via teleconference.

APPENDIX A-2: TERMS OF REFERENCE OF THE EXPERT TECHNICAL GROUP

AIM OF PROJECT
To review and assess the design options for evaluating the impact of the BreastScreen Australia program on breast-cancer mortality in the population with due regard for feasibility and quality of evidence.

TERMS OF REFERENCE
The Expert Technical Group will be instrumental in identifying key objectives for the project and synthesising results, including providing:

• expert input to facilitate scoping and planning of the literature review and data extraction process
• appraisal of the accuracy and completeness of data extraction
• critical review and ranking of design options described in the literature
• contributing to the final project report.
PROJECT DURATION AND MEETING FREQUENCY

This project will start in July 2006 and is expected to take approximately 6 months to complete. During this time, it is expected that Expert Technical Group members will be invited to attend up to four meetings, which may be held face-to-face or via teleconference.

APPENDIX A-3: STATED POTENTIAL CONFLICTS OF INTEREST OF STEERING COMMITTEE AND EXPERT TECHNICAL GROUP MEMBERS

Where a ‘Conflict of Interest’ means not only a financial interest, but any matter which influences, or may appear to influence, proper consideration or decision-making on a matter.

STEERING COMMITTEE

Professor Dallas English

Professor English advised members that he was a member of the Board of BreastScreen Victoria. Members were unanimous that this was not a conflict of interest. Professor English and Associate Professor Dorota Gertig also advised members that they were co-investigators on a current NHMRC submission examining ways to evaluate BreastScreen Victoria. Members agreed this was a complementary study and were unanimous that it did not present a conflict of interest.

EXPERT TECHNICAL GROUP

Professor David Roder

Professor Roder advised members that he was an investigator on a study examining ways to evaluate BreastScreen South Australia. Members agreed this was a complementary study and were unanimous that it did not present a conflict of interest.

Dr Chris Stevenson

Dr Stevenson advised members that he may have a potential conflict of interest as he may make a submission to undertake the record linkage for the proposed case-control study. Members of both the Steering Committee and the ETG agreed that each of them may be in a position to make a submission to undertake one or more aspects of the recommended study designs given the relatively small number of experts in the area. Members were unanimous that this was not a conflict of interest at this point.

Associate Professor Dorota Gertig

See above.
APPENDIX A-4: INTERNATIONAL EXPERTS CONTACTED FOR THE REVIEW

**Dr Roger Blanks**  
Epidemiologist, Institute of Cancer Research, Sutton, Surrey, United Kingdom

**Dr Heather Bryant**  
Research Scientist, Division of Population Health and Information, Tom Baker Cancer Centre, Calgary, Alberta, Canada

**Professor Nicholas Day**  
Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom

**Professor Stephen Duffy**  
Cancer Research UK Centre for Epidemiology, Mathematics and Statistics, Wolfson Institute of Preventive Medicine, London, United Kingdom

**Professor Joann Elmore**  
Professor of Medicine & Adjunct Professor of Epidemiology, Harborview Medical Center Seattle, Washington, USA

**Professor Suzanne Fletcher**  
Professor of Ambulatory Care and Prevention, Harvard School of Medicine, Professor of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA

**Professor Matti Hakama**  
Research Professor, Tampere School of Public Health, University of Tampere, Finland

**Professor Hakan Jonsson**  
Department of Oncology, Umea University Hospital, Umea, Sweden

**Associate Professor Harry de Koning**  
Department of Public Health, Erasmus Medical Centre, Rotterdam, The Netherlands

**Dr Sue Moss**  
Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton, Surrey, United Kingdom

**Dr Eugenio Paci**  
Unit of Clinical and Descriptive Epidemiology, Centre for Study and Prevention of Cancer, Florence, Italy

**Dr Robert Smith**  
Director of Cancer Screening, American Cancer Society, Atlanta, Georgia, USA

**Professor László Tabár**  
Professor of Radiology, Department of Mammography, Falun Central Hospital, Falun, Sweden

**Professor Noel Weiss**  
Professor of Epidemiology, School of Public Health and Community Medicine, University of Washington, Seattle, Washington, USA
## APPENDIX A-5: DATA EXTRACTION FORM

### 1. Article Details

<table>
<thead>
<tr>
<th>Study No.:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td></td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td></td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.):</td>
<td></td>
</tr>
<tr>
<td>Impact Factor (2003):</td>
<td></td>
</tr>
</tbody>
</table>

### 2. Description of the study (State page number when possible)

<table>
<thead>
<tr>
<th>Objectives:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Location of study:</td>
<td></td>
</tr>
</tbody>
</table>

### 3. Study Design and Population (When possible, report as stated in article and specify page numbers)

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants:</td>
<td></td>
</tr>
<tr>
<td>Target screening interval:</td>
<td></td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
<td></td>
</tr>
<tr>
<td>Participation rate:</td>
<td></td>
</tr>
<tr>
<td>Study design:</td>
<td></td>
</tr>
<tr>
<td>Entry criteria:</td>
<td></td>
</tr>
<tr>
<td>Recruitment strategy:</td>
<td></td>
</tr>
</tbody>
</table>

**Nature of Comparison Groups:**

<table>
<thead>
<tr>
<th>Group 1 (Reference group): (N= xxx)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 (Intervention group): (N= xxx)</td>
<td></td>
</tr>
</tbody>
</table>

| Outcome/s of interest reported: |  |
4. Bias reduction and statistical methodology (When possible, report as stated in article and specify page numbers)

Data required in order to conduct primary analyses:

At what time points?:

Statistical analyses performed:

Adjustments performed for inter-group differences:

Investigating effect of extent of screening exposure:

Other methods of bias reduction (incl loss to follow-up):

6. Study strengths and weaknesses

Study assumptions:

Strengths:

Weaknesses:

7. Other comments
APPENDIX A-6: PROMPTERS OF SELECTED STUDY DESIGNS, OUTCOME OPTIONS AND TYPES OF BIAS

STUDY DESIGNS

1. Theoretical Models
These are simulations of breast-cancer progression from pre-clinical states to death, with estimated transition probabilities across states, and with estimated effects on these transitions of exposures to screening and therapies. Models generally are calibrated to local incidence, mortality and other disease characteristics, such as stage. While calibration may increase model credibility, results remain sensitive to the assumptions used in the model. There are many model types. They may be macro or micro simulations, and biologically or epidemiologically based. Examples include MISCAN and MICROLIFE models, and other Markov simulations. Apart from calibration to local data, these are theoretical models with results that are subject to multiple assumptions. Use of models of these types to evaluate Australian screening effects would rely on assumptions inferred from the international literature where local applicability would be uncertain.

2. Ecological Designs
In these designs, measures apply to population groups or sub-groups, not individuals. Examples include analyses of breast-cancer mortality rates in population sub-groups over time, geographic differences, or time trends by geographic area, according to screening exposure. Such studies have the advantage of addressing ecological effects of screening. They normally can be completed with available de-identified data, avoiding HREC barriers and increasing feasibility. However, results would be prone to confounding from the so-called ‘ecological fallacy’

3. Case Series
In these studies, outcomes (eg, survivals) may be compared in patient series before and after the introduction of screening to infer screening effects. These studies are vulnerable to lead-time, length-time, over diagnosis and related biases and to confounding from other influences.

4. Case-Control Designs
Generally, people dying of breast cancer are compared with controls of the same age, who are still alive at the time of the death case. Comparisons may be of past exposures to screening or to screening invitations. These studies are efficient, but considerable care is needed in case/control selection and other methodological features to avoid bias. Nonetheless, aggregated results from case-control studies appear to be consistent with trials data. A disadvantage is that measures of past exposures may need to rely on self-reporting (not possible from dead cases) or on data already collected in administrative data systems, which may be of questionable accuracy and completeness.
5. Cohort Designs

Generally, people exposed to screening or screening invitations are followed to compare their breast-cancer outcomes with those not so exposed. If these are prospective as opposed to retrospective designs, then measures of exposures and confounders can be attuned to study needs. If to avoid unacceptably long study durations, the data are collected retrospectively, then reliance is needed on data already available in administrative data systems, with the drawbacks of uncertain accuracy and completeness.

6. Hybrid Cohort/Case-Control Designs

These designs include cohort studies with nested case-control components. Where the cohort study is prospective, associated high quality data can be supplemented with additional data collected in a nested case-control investigation. If the cases are dead (or cannot self-report for other reasons), the additional case-control data may be of more limited value. Other case-cohort hybrids include the collection of additional data on cases at diagnosis (i.e., when access to self-reporting and clinical data are still possible), plus corresponding data on a sample cohort. Such designs recognise the rareness of breast cancer, and the desirability of collecting a broad range of data on possible confounders, but they avoid the need to collect these broader data on the entire cohort.

OUTCOME OPTIONS

1. Breast-cancer mortality – subject to bias from misattribution
2. All causes mortality – subject to type-2 error from inadequate study power
3. Excess mortality – useful to infer breast cancer mortality among breast-cancer patients, but not conducive to population-based mortality assessment
4. Refined (selective) breast-cancer mortality – useful to avoid attenuating effects from including cancers diagnosed prior to commencement of screening
5. Surrogate mortality endpoints (e.g., large tumours or large/node positive/high grade tumours) – useful to increase power when numbers of deaths saved would be small due to short follow-up period. Useful for collecting self-reported data on possible confounders. Allows earlier analyses before deaths have occurred. Need to have a credible link from the surrogate endpoint to mortality. Potential for bias from lead-time and related effects
6. Screening performance indicators – also useful to increase power. These indicators would need to be evidence-based and have a credible link with mortality
TYPES OF BIASES

1. **Lead-time** – The aim of screening is to find cancers earlier (in lead-time periods). Lead-time bias occurs when case-survival comparisons point to artificial gains for cancers detected through mammography (i.e., if the diagnoses are earlier but without changes in the dates of death).

2. **Length bias** – Tumours found in mammography screening tend to be those that progress more slowly. Length-time bias occurs if case-survival comparisons point to artificial gains for cancers detected through mammography (i.e., the difference are due to the selective pick-up through screening of slowly progressing tumours).

3. **Over-diagnosis** – Screening would find a proportion of cancers that would not have progressed if left undetected. Over-diagnosis bias would occur if case-survival comparisons pointed to artificial gains for cancers detected through mammography (i.e., the differences are due to dilution of the detected cancers by non-progressive lesions).

4. **Self-selection** – High-risk people may self-select for screening. Alternatively, those with a breast-cancer history or with symptoms may be ineligible for screening. Mortality comparisons of screened and unscreened people can therefore be affected by selection bias.

5. **Healthy-screening effect** – People who don’t get breast cancer are more likely to continue with screening, artificially causing a negative association of screening duration with breast-cancer mortality.

6. **Misclassification of screening and diagnostic mammograms** – Mammography is a common feature of diagnostic work-ups for breast cancer. If diagnostic mammograms were misclassified as screening mammograms, an artificial positive association of screening with breast-cancer death would result.

7. **Attenuation** – Attenuation bias occurs if the exposed are defined as women invited for screening, when some of these don’t get screened. Similarly, there is attenuation bias if the non-exposed include some not invited to screening but who nonetheless were screened.

8. **Inadequate follow-up** – Results would bias towards the null if assessment occurred too early, before enough deaths had been prevented.
APPENDIX A-7: POSSIBLE STUDY DESIGNS FOR EVALUATING THE EFFECT OF BREASTSCREEN ON BREAST CANCER MORTALITY

Example only

1. ECOLOGICAL STUDY OF ASSOCIATION OF BS PARTICIPATION WITH BREAST CANCER MORTALITY BY SLA

Aim:
To determine the association by residential SLA of BS screening participation with time-lagged BCA mortality, adjusting for numbers of mammograms provided outside BS, SEIFA index of relative socio-economic disadvantage of residential area, ARIA score, other socio-demographic descriptors, state/territory jurisdiction, and BCA mortality rates prior to the advent of mammography screening (i.e., rates in the 1980s)

Process:
- Obtain annual SLA specific data for the 1990-2005 period for 50–69 year olds by age, as relating to:
  - BS participation (data from BreastScreen)
  - Population data and socio-demographic descriptors (interpolated data from ABS)
  - Breast cancer mortality rates in the 1980s (data from ABS)
  - Privately funded mammograms (data from Medicare)
  - Mammograms provided to public hospital inpatients (data from hospital separations)
- Obtain from the National Cancer Statistics Clearing House (or from state/territory cancer registries if necessary) annual SLA specific breast-cancer death data by age for diagnoses occurring in 1990–2005 among 50–69 year olds resident of the respective SLA at diagnosis.
- Model breast cancer death rates by SLA according to BS participation, with adjustment for age, SEIFA index, ARIA index, state/territory jurisdiction, other socio-demographic descriptors, mortality rates for breast cancer in the 1980s, and mammogram exposure outside BS. This would involve weighted regression analysis (possibly Poisson regression). An alternative outcome would be the annual reduction in breast cancer mortality (as used in an earlier NSW study)

Advantages:
- Proof-of-concept already has been established for this type of design, both in NSW and overseas.
- Outcome may capture indirect effects (i.e., ‘externalities’) as well as direct screening effects
- Data likely to be available
- No need for person-identified data, facilitating HREC approval
- Inexpensive
- Quick
Disadvantages/risks:

- Variations in BS participation may be smaller by SLA than ideal for study purposes (apparently not a problem in earlier NSW study)
- Uncertainty about time lags between screening and mortality effects, requiring modelling
- Possible difficulties identifying histories of BS screening in jurisdictions without centralised record systems (eg, NSW)
- Data on many breast-cancer risk factors not available by SLA (eg, family history), requiring assumptions of equivalence
- Possibly ‘ecological fallacies’
- Negative attitudes of some experts to ecological study designs

Comments:

This seems to be a low-risk approach for fast tracking initial findings. This approach could be performed in advance of gaining results from a more time-consuming alternative approach

APPENDIX A-8: OVERVIEW OF NATIONAL DATA SOURCES

BREASTSCREEN REGISTRY DATA

Each state and territory BreastScreen service maintains a register of women who attend for screening. These registers contain demographic data for each woman screened in a BreastScreen service as well as data on the mammogram and its outcomes and any follow-up data for each time the woman has attended a BreastScreen service for screening. Each state and territory screening program annually extracts an agreed subset of the register data and sends it to the AIHW for compilation into a national monitoring dataset. The registers have identifying data for each woman but the national dataset is de-identified – that is, it contains a record for each woman screened containing screening outcome data but no identifying information.

The BreastScreen program in each state and territory annually links its register to the jurisdiction’s cancer registry (see section 3 below) to identify cancers diagnosed in women who have had a clear screening mammogram (known as interval cancers). There is some under-identification of these interval cancers because this linkage may miss those women whose cancers were diagnosed in a different jurisdiction.

The strength of the BreastScreen register data is in its complete coverage of all women screened at a BreastScreen service. However, there is evidence that a substantial proportion of mammography examinations outside the Program are performed for screening purposes. Some of this is reimbursed under Medicare and recorded as diagnostic mammography and some is done for screening purposes and paid for by the women. There is also some overlap between the registers in that a woman who has a screen in two different jurisdictions may be recorded on the registers of each jurisdiction. Although the national breast-screening program dates from January 1991, the individual state and territory registers commenced at different times. Western Australia had the earliest register, starting in 1989. The New South Wales, Queensland, and South Australian registers started in 1991, the Victorian, Tasmanian and ACT registers in 1993, and the Northern Territory register at the end of 1994.
CANCER REGISTRY DATA

Cancer is a notifiable disease in each Australian State and Territory, and each jurisdiction has a cancer registry which records each new case of cancer in that jurisdiction. The registries contain demographic data on each person with a cancer and detailed data on each tumour diagnosed in each person. The registries annually pass an agreed monitoring dataset to the National Cancer Statistics Clearing House (NCSCH) at the AIHW. Both the registry data and the NCSCH data contain identifying information. This enables the NCSCH to match data from each jurisdiction to remove duplicate records (i.e., records for a person who has been diagnosed and/or treated for the same cancer in two or more jurisdictions).

The strength of the cancer registry data is its almost complete coverage of cancer in Australia. However, the registries contain little or no treatment or outcomes data other than recording a person’s death. This recording of death may be incomplete as a death occurring in a different jurisdiction to the registry may not be recorded. From time to time, the AIHW matches NCSCH data to national deaths data to identify cancer patients who have died in any part of Australia. This will still miss those deaths occurring outside Australia (e.g., migrants who return ‘home’ to die). There will commonly be a lag of 12 to 18 months for a cancer to be registered on a jurisdictional registry and a further 12 to 18 months for that cancer to be recorded on the NCSCH.

National cancer incidence data are available from the NCSCH from 1982.

DEATHS DATA

The AIHW maintains a National Death Index. This contains data for all deaths registered in Australia from 1980 and includes the name and date of birth of the person who has died, the place of death and the cause of death. The Index’s strength is in its completeness and its inclusion of the underlying cause of death as recorded on the death certificate (for all deaths from 1980), and all other associated causes as recorded on the death certificate (for all deaths from 1997).

MEDICARE AND PHARMACEUTICAL BENEFITS SCHEME DATA

These data sets record all medical services reimbursed under the Medicare scheme and all prescribed medication attracting a benefit under the Pharmaceutical Benefits Scheme (PBS). The strength of these databases is that they provide the most complete coverage of out-of-hospital medical services and prescribed medication in Australia. For example, the Medicare dataset potentially allows the identification of de-facto screening mammography outside the BreastScreen services, though the difficulty here is that such mammography will be recorded as diagnostic and is not readily differentiated from genuinely diagnostic mammography. These databases contain identifying information, but there are legal and logistical difficulties in accessing this information for linkage purposes. This is discussed further in section 8 below.

A second strength of the Medicare database is that it contains a record for all people entitled to Medicare rebates. Hence it provides an effectively complete list of all Australian women since 1983–84 (when the Medicare scheme commenced). If this list could be linked to the BreastScreen registers, the cancer registries and the death index, then it would allow examination of cancer and mortality outcomes for all Australian women according to whether or not they have had mammography.
HOSPITAL DATA

The AIHW maintains a national database of hospital services covering both public and private hospital in-patient services since 1993–94. The strength of the database is its wide coverage of services. Its main limitations from the point of view of the mortality study are:

1. That it only covers admitted patients – and many breast screening and cancer-related services may be provided to non-admitted patients in same-day clinics;
2. It is a database of hospital stays rather than hospital patients, with no patient identifying information, so that a single person having many hospital stays cannot be distinguished from many people having a single stay; and
3. The AIHW holds data from 1993–94, though the first year should be regarded as incomplete. Hence there are no data covering the period prior to the introduction of the breast-screening program in 1991 and either no data, or at best, only incomplete data for its first four years of operation.

BEACH GP SURVEY DATA

The BEACH program is a national GP survey program that began in 1998 and involves a random sample of around 1,000 GP’s each year. It collects details of 100 consecutive patient consultations from each GP in the sample. The information collected covers the reason given for the person attending the GP, the GP’s assessment of the problem underlying this reason and any medications prescribed or referrals made during the consultation. This data set collects basic patient demographic information but no patient identifying information. In principle, these data could be used to investigate mammography referral patterns as part of identifying those diagnostic mammography which are in fact de-facto screening mammograms. However, the sample size is small, which means that there will not be enough patients referred for mammography, and in particular de facto screening mammography, to facilitate this analysis.

POTENTIAL FOR DATA LINKAGE

The hospitals data and the BEACH GP survey data contain no personal identifying information, so they are unsuitable for data linkage. However, the value of the other national data sets could be greatly enhanced by linkage between them. For example, the cancer registry, BreastScreen register and death index datasets could potentially be linked to create a record for each woman with an incident breast cancer, identifying whether or not the cancer was detected by the BreastScreen service and whether or not she has died from that cancer or from some other reason. Such a linked dataset could potentially be used to explore differences in outcomes between BreastScreen detected and other cancers. This section explores the potential for such linkages to be done as part of a mortality study.

Medicare data are currently available to members of the Australian Government Department of Health and Ageing (DoHA) and may be linked to PBS data for approved medical research purposes – with the approval of the Department Secretary. Confidential linked de-identified Medicare and PBS data may be passed on to an evaluator under an existing legal instrument. Researchers would be required to agree to certain conditions under guidelines issued by the Privacy Commissioner. However, the greatest use of the Medicare and PBS data in a mortality study would be in linkage to one or more of the other datasets.
Currently data linkage between the other datasets at a state/territory level with Medicare and PBS is only possible in Western Australia, because of an existing Memorandum of Understanding (MOU) with DoHA. Other States and Territories, (particularly New South Wales and the ACT) are progressing data linkage within their jurisdictions. Should these jurisdictions wish to link to Medicare and/or PBS data in the future, this would be done under a similar MOU to that which operates between WA and DoHA. However, this process requires a substantial amount of time and is not expected to be in place within the timeframe of the BreastScreen Australia Evaluation.

The NCSCH database and the death index have been linked regularly for survival analyses and are currently being linked for all cancer incident cases in all states and territories to the end of 2003. The AIHW is exploring linkage of the NCSCH, the death index, the BreastScreen register data and Medicare data as part of the BreastScreen evaluation. Ethics committee approvals for such a linkage will depend on a detailed protocol covering the linkage methods and the subsequent use of the linked data. Hence, the linkage cannot proceed further until details of the evaluation, including the mortality study methodology, have been finalised. Further, some practical difficulties involved in the linkage of such large datasets remain to be resolved.

In addition to national level data linkage, significant linkage is being done at the state and territory level. As noted above, Western Australia has an advanced program of data linkage covering most major health data sets, including the Medicare and PBS data. The New South Wales Cancer Institute has linked the BreastScreen register, cancer registry and deaths data for that state, and an application has been made to extend this linkage to the Medicare data. The New South Wales data are particularly interesting as this is the only jurisdiction whose cancer registry records stage at diagnosis for breast cancer cases. This is important, as it would allow for the examination of any shift to earlier diagnosis in the screen-detected cases.
### APPENDIX A-9: STUDY DESIGN RANKING FORM

1. **Proposed Study Number:**

   Study Design and Population:

   Statistical Methodology:

   Methods to Achieve Valid Results:

   Rationale and Assumptions:

2. **Ranking of Likely Value of Evidence:**
   (1=little or no value; 2=doubtful value; 3=possibly good; 4=probably good; 5=good)
   Score (1-5):
   Reasons:

3. **Ranking of Likely Ease of Implementation:**
   (1=not doable; 2=probably not doable; 3=possibly doable; 4=probably doable; 5=readily doable)
   Score (1-5):
   Reasons:

4. **Ranking of Risk of Unsuccessful Implementation:**
   (1=too risky; 2=probably too risky; 3=risks possibly manageable; 4=risks probably manageable; 5=not risky)
   Score (1-5):
   Reasons:

5. **Ranking of Anticipated Cost:**
   (1=costly end; 2=probably at costly end; 3=possibly mid-range in cost; 4=probably at low cost end; 5=at low cost end)
   Score (1-5):
   Reasons:

6. **Ranking of Likely Time to Complete Study:**
   (1=over 3yrs; 2=about 3yrs.; 3=about 2yrs.; 4=about 18 months; 5=about 1yr or less)
   Score (1-5):
   Reasons:
## APPENDIX A-10: DATA EXTRACTION FORMS FOR INCLUDED PUBLICATIONS

### ANTILLA et al, 2002

#### 1. Article Details

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>1156</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Programme sensitivity and effectiveness of mammography service screening</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>Anttila et al, 2002</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.):</td>
<td>J Med Screen 9: 153–58</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>2.483</td>
</tr>
</tbody>
</table>

#### 2. Description of the study

| Objectives: | To evaluate effectiveness of mammography service screening |
| Location of study: | Helsinki, Finland |

#### 3. Study Design and Population

| Nature of screening programme & mammography: | Helsinki breast cancer screening service. NR, appears to be mammography alone. Women are invited based on birth year. |
| Target age for screening participants: | 50–59 years |
| Target screening interval: | 2 years |
| Start date and maturity of programme under evaluation: | 1986, 11 years |
| Participation rate: | Average participation rate was 82% (range 79%–84%) |
| Study design: | Retrospective before and after study (historical cohort). Mammography screening database, and linked records between cancer incidence and mortality from Finnish Cancer registry. |
| Entry criteria: | N/A |
| Recruitment strategy: | N/A |

**Nature of Comparison Groups:**

- **Group 1 (Reference group):** (N= 155,400 women-years) Reference non-screened cohort: Immediately preceding 5 year birth cohort (1930–34)
- **Group 2 (Intervention group):** (N= 161,400 women-years) Screened cohort: The first screening cohort (1935–39) (p 155)

| Outcome/s of interest reported: | Unadjusted and adjusted breast cancer mortality at age ≥50 years |

#### 4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | Incidence and mortality for women within specific birth year cohorts |
| At what time points?: | Data for two periods, before and after introduction of screening. Mean of 10 years follow-up from first invitation. NB. Also need data on incidence and mortality rates for pre-screen ages for purpose of adjustments, and the prior age cohort for purpose of comparison. |
| Statistical analyses performed: | Mortality data from non-screened group censored to equal the duration of follow-up of the screened group (p155). Excludes diagnoses and deaths prior to the screening age (i.e., <50 years). Significance tested assuming the number of observed cases followed a Poisson distribution (p156). No further detail provided. |
| Adjustments performed for inter-group differences: | Adjusted for potential treatment or other background effects by adjusting the difference between cohorts by their death rate from incident cancer in the pre-screening ages, 40–49, and in age at death (p 156). |
| Investigating effect of extent of screening exposure: | NR |
| Other methods of bias reduction (incl. loss to follow-up): | Use of reference group immediately prior to screened group. Truncation of follow-up of earlier cohort. Adjustments as described above. |
6. Study strengths and weaknesses

| Study assumptions: | Not possible to link women from actual women from mammography to incidence and mortality. Assumes all those within screened cohort were screening (actually ~80%). |
| Study strengths: | Individual patient level data not available for screening (p157). |
| Weaknesses: | Even though adjustment for risk of disease over time, can not rule out possibility that small numbers might have partially affected the results (p157). |
| | Not clear to what extent non-screened women within the screened birth cohort dilute the screening effect. |
| | Relatively small cohort. Not concurrent cohorts. |

7. Other comments

Result is non-significant

BAKER et al. 2004

1. Article Details

| Study No.: | 432 |
| Title: | Comparing breast cancer mortality rates before-and-after a change in availability of screening in different regions: extension of the paired availability design. |
| Author & year: | Baker et al, 2004 |
| Impact Factor (2005): | No impact factor available |
| Also include as a methodological publication |

2. Description of the study

Objectives: Compare breast cancer mortality before and after a change in the availability of screening in different regions

Location of study: Sweden

3. Study Design and Population

| Nature of screening programme & mammography: | Local screening mammography programmes in six counties (originally seven, but one excluded as large proportion enrolled in RCT). No other detail provided. |
| Target age for screening participants: | Varied between counties |
| Target screening interval: | NR |
| Start date and maturity of programme under evaluation: | NR |
| Participation rate: | Varied between counties (range 64%–92%) |
| Study design: | Paired availability design (uses data from all historical controls, both screened and unscreened, to reduce selection bias) |
| Entry criteria: | N/A |
| Recruitment strategy: | N/A |

Nature of Comparison Groups:

| Group 1 (Reference group): | (N= 292,369 eligible women) Reference: Prior to mass screening (although 14% in one county) |
| Group 2 (Intervention group): | (N= 309,879 eligible women) Screening: After introduction of mass screening |
| Outcome/s of interest reported: | Incident breast cancer deaths |
### 4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | Uses data from multiple regions that all have before and after data available. Breast cancer deaths only where the cancer was diagnosed within the period of interest. Actual receipt of screening accounted for (not availability of screening) |
| At what time points?: | Data for two periods, before and after introduction of screening. Time periods must be same length, and sufficiently long to maximise impact on mortality (p13) |
| Statistical analyses performed: | Complex analyses – refer to original paper |
| Adjustments performed for inter-group differences: | Removes selection bias because a comparison is made between i) mortality in all subjects in the current time period when screening is widely available with ii) mortality in all subjects in the previous time period when screening was less available (p13) (i.e., doesn’t just report the outcomes in those that obtained screening) \(\text{Adjustments for lead time related to prior screening (time from screen detection to clinical detection in the absence of screening)}\) \(\text{Adjustment for lead time related to age-range at diagnosis}\) |

#### 6. Study strengths and weaknesses

| Study assumptions: | Assumes stable population (characteristics of population are the same over time i.e., inward and outward migration must be patients of the same nature) \(\text{Assumes stable treatment (assumes screening modality and treatment after diagnosis do not change over time)}\) \(\text{Assumes measure of mortality does not change over time.}\) \(\text{Assumes stable preferences (factors affecting decision to receive screening).}\) \(\text{Assumes stable screening effects (that overall impact of screening on deaths is constant over time)}\) |
| Strengths: | Actual receipt of screening accounted for (not availability of screening). This may include small amount of screening occurring in the earlier period. \(\text{Time period duration can vary between regions/states (but not between before and after periods for any one state).}\) \(\text{Adjustments for lead time related to prior screening (time from screen detection to clinical detection in the absence of screening)}\) \(\text{Adjustment for lead time related to age-range at diagnosis}\) |
| Weaknesses: | Caution that Assumption 2 (stable treatment) may not hold due to changed in systemic therapy during the periods of interest, therefore may overstate the benefit of screening. Treatment changes over time are not accounted for. \(\text{Assumption 4 (stable preferences) also likely to be violated if screening campaign used.}\) \(\text{Are periods truly long enough to satisfy ‘Requirement 1’ (screening to have impact on mortality, p13)}\) |

#### 7. Other comments

Could be used to account for treatment changes over time if had a region where there was no screening in either time period.
### 1. Article Details

<table>
<thead>
<tr>
<th>Study No.</th>
<th>2418</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Population-based breast cancer survival: Mammographic screening activities in Central Italy.</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>Barchielli et al, 1994</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.):</td>
<td>Cancer 74(12):3126–34.</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>4.800</td>
</tr>
</tbody>
</table>

NB. More recent analyses in Barchielli et al, 2001 (Study No. 1292)

### 2. Description of the study

**Objectives:** To evaluate the effect of screening programme on 5 year survival of patients with invasive cancer

**Location of study:** Central Italy

### 3. Study Design and Population

**Nature of screening programme & mammography:** Screening was performed in several rural municipalities (~15% of women in province).

**Target age for screening participants:** 40–70 years

**Target screening interval:** NR

**Start date and maturity of programme under evaluation:** 1970, 26 years

**Participation rate:** Average participation in ‘screened’ rural municipalities ~52% Some asymptomatic mammography in urban Florence (~10%)

**Study design:** Retrospective concurrent cohort study

**Entry criteria:** N/A

**Recruitment strategy:** N/A

**Nature of Comparison Groups:**

**Group 1 (Reference group):** (N = ~577,000) Reference: All women in other provinces, excluding those below

**Group 2 (Intervention group):** (N = ~33,000) Screened: Women in several rural municipalities

**Outcome/s of interest reported:** Paper reports breast cancer incidence, mortality and mortality: incidence ratio in different municipalities (including the screening area), but performs no analyses with these data.

### 4. Bias reduction and statistical methodology

**Data required in order to conduct primary analyses:** Concurrent data from two regions: breast cancer mortality, age (NB. focus of paper was survival)

**At what time points?:** Annual over 6 years

**Statistical analyses performed:** Cox model (for survival only). No analyses with mortality data.

**Adjustments performed for inter-group differences:** Age-standardised to European population.

**Investigating effect of extent of screening exposure:** None

**Other methods of bias reduction (incl. loss to follow-up):**
6. Study strengths and weaknesses

Study assumptions:

| Strengths: | Differences in mortality may be due to differences in treatment. Screened group were all rural women. May introduce bias due to baseline differences with respect to remainder of group (e.g., socioeconomic, access to follow-up and treatment). Does not exclude possibility of lead-time and length bias (for survival analysis only?). |

7. Other comments

Not designed to measure impact of mammography per se. Focus is on survival only.

BARCHIELLI & PACI, 2001

1. Article Details

| Study No.: | 1292 |
| Title: | Trends in breast cancer mortality, incidence, and survival, and mammographic screening in Tuscany, Italy. |
| Author & year: | Barchielli & Paci, 2001 |
| Citation (Jnl, Vol(Num):Pgs.): | Cancer Causes Control 12(3):249–55. |
| Impact Factor (2005): | 3.195 |

More recent analyses of Central Italy mammography screening programme is provided in Paci 2002 (Study No. 1014) and Gorini et al, 2004 (Study No. 3715)

2. Description of the study

Objectives: Study described breast cancer mortality time trends (1970–97), taking into account diffusion of screening

Location of study: Central Italy. Compares ‘Florence area’ with ‘rest of Tuscany’

3. Study Design and Population

| Nature of screening programme & mammography: | Screening by personal invitation in ‘Florence area’ since end 1990. NB. some rural municipalities had screening since 1970s Other ‘Florence area’ municipalities had screening introduced some time after 1992 No population screening programmes in rest of Tuscany |
| Target age for screening participants: | 50–69 years |
| Target screening interval: | NR |
| Start date and maturity of programme under evaluation: | Highly variable, but most screening started –1990 (incl. Florence city). 7 years |
| Participation rate: | Average compliance rate was 60% |
| Study design: | Descriptive only. Does not directly compare mammography screening with no screening |
| Entry criteria: | N/A |
| Recruitment strategy: | N/A |

Nature of Comparison Groups:

| Group 1 (Reference group): | N/A, descriptive only |
| Group 2 (Intervention group): | N/A, descriptive only |
| Outcome/s of interest reported: | Breast cancer mortality |
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer deaths. Age.

At what time points?: Over 27 year period

Statistical analyses performed: Age-adjusted time trends in mortality were analysed by Poisson regression models, assuming year of death as a continuous variable to estimate the average change by calendar year.

Adjustments performed for inter-group differences: Breast cancer mortality is age-standardised.

Investigating effect of extent of screening exposure: NR

Other methods of bias reduction (incl. loss to follow-up):

6. Study strengths and weaknesses

Study assumptions: Descriptive only, but basic premise is linking temporal pattern of breast cancer mortality to availability of mammography. Does not directly compare mammography screening with no screening

Strengths:

Weaknesses: Very weak design to address impact of mammography upon mortality.

7. Other comments

BLANKS et al, 2000

1. Article Details

Study No.: 3772

Title: Effect of NHS breast screening programme on mortality from breast cancer in England and Wales, 1990-8: Comparison of observed with predicted mortality.

Author & year: Blanks et al, 2000


Impact Factor (2005): 9.052

2. Description of the study

Objectives: To assess the impact of the NHS breast screening programme on mortality from breast cancer in women aged 55–69 years, over the period 1990–98

Location of study: England and Wales
<table>
<thead>
<tr>
<th><strong>3. Study Design and Population</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Nature of screening programme &amp; mammography:</td>
</tr>
<tr>
<td>Target age for screening participants:</td>
</tr>
<tr>
<td>Target screening interval:</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
</tr>
<tr>
<td>Participation rate:</td>
</tr>
<tr>
<td>Study design:</td>
</tr>
<tr>
<td>Entry criteria:</td>
</tr>
<tr>
<td>Recruitment strategy:</td>
</tr>
<tr>
<td><strong>Nature of Comparison Groups:</strong></td>
</tr>
<tr>
<td>Group 1 (Reference group):</td>
</tr>
<tr>
<td>Group 2 (Intervention group):</td>
</tr>
<tr>
<td>Outcome/s of interest reported:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>4. Bias reduction and statistical methodology</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Data required in order to conduct primary analyses:</td>
</tr>
<tr>
<td>At what time points?:</td>
</tr>
<tr>
<td>Statistical analyses performed:</td>
</tr>
<tr>
<td>Adjustments performed for inter-group differences:</td>
</tr>
<tr>
<td>Investigating effect of extent of screening exposure:</td>
</tr>
<tr>
<td>Other methods of bias reduction (incl. loss to follow-up):</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>6. Study strengths and weaknesses</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Study assumptions:</td>
</tr>
<tr>
<td>Strengths:</td>
</tr>
<tr>
<td>Weaknesses:</td>
</tr>
</tbody>
</table>

| **7. Other comments** |
1. Article Details

**Study No.:** 1154  
**Title:** Repeated mammographic screening reduces breast cancer mortality along the continuum of age.  
**Author & year:** Broeders et al, 2002  
**Impact Factor (2005):** 2.483

Supersedes Van Dijck et al, 1996 (Study No. 2088) and Van Dijck et al, 1994 (Study No. 2353) and Van Dijck et al, 1985 (Study No. 2885). Duplicates the Nijmegen data reported in Broeders et al, 2001 (Study No. 1307), but there it is compared to another city without screening.

2. Description of the study

**Objectives:** To describe the effect of mammographic screening on breast cancer mortality along the continuum of age, based on a 20 year follow-up period (abstract)

**Location of study:** Nijmegen city, The Netherlands

3. Study Design and Population

**Nature of screening programme & mammography:** Local population based mammography screening programme by invitation. Two view mammography at initial screen, with only mediolateral view subsequently (for ~80%).

**Target age for screening participants:** 35–65 years initially, then over 35 from 1977, then gradually adjusted to 50–69 between 1989–1997, 50–74 thereafter.

**Target screening interval:** 2 years

**Start date and maturity of programme under evaluation:** 1975, with 20 year follow-up

**Participation rate:** ~55% after first round (p164)

**Study design:** Case-referent

**Entry criteria:** Eligible population: all women who had received at least one invitation to participate in screening between 1 January 1975 and 1 January 1997

**Recruitment strategy:** see below for cases and referents

**Nature of Comparison Groups:**

**Group 1 (Controls):**  
(N= 785)  
For each case, a set of eligible referents was selected from the population who i) were alive and residing in Nijmegen at the time of death of the case, ii) had been invited to participate in the round of the case’s index screening (see below), and iii) were free of breast cancer at their index invitation.  
Five referents were randomly selected from each set. Referents were selected from those at risk at the time of death of each case, known as incidence-density sampling.

**Group 2 (Cases):**  
(N= 157)  
Defined as women from the eligible population who died from primary and histologically-confirmed breast cancer between 1 January 1987 and 1 January 1997. Cause of death based on medical records and death certificate, and classified by a medical panel unaware of screening history.

**Outcome/s of interest reported:** Screen-detected cases, index screening was the screening examination at which the cancer was detected. For interval cancer cases, and cases who had never attended screening, the most recent invitation preceding the diagnosis was defined as the index screening.
### 4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | Ability to link mortality, medical records and screening databases on an individual patient basis (for cases and referents). Must know date of diagnosis of cases. NB. Must ensure referents and cases are mutually exclusive |
| At what time points?: | Screening data over 20 year period, mortality data over 10 year period. |
| Statistical analyses performed: | The odds ratio is the estimator of the ratio of breast cancer mortality rates in women who had accepted the invitation to the index screening, relative to those who had not. |
| Adjustments performed for inter-group differences: | NR |
| Investigating effect of extent of screening exposure: | Indicated indirectly by comparing the mortality odds ratios results between the index screen alone and index + previous screen? |
| Other methods of bias reduction (incl. loss to follow-up): | NR |

### 6. Study strengths and weaknesses

| Study assumptions: | Results only relate to participation in the ‘index’ screening (not ever screened), although additional analyses were also performed with attendance at both index + the previous screening). For case-referent methodology to be sound using incidence-density sampling, the proportion screened amongst the population invited must be constant over the study period. This was poorly reported (reasonably stable around ~55%). |
| Strengths: | Results presented separately by age at index screening |
| Weaknesses: | Subject to self-selection bias. Higher risk women may have attended screened, or more health-conscious women may have attended screening. |

### 7. Other comments

Results not significant in any age group. Overall results are not shown.

---

**BROEDERS et al, 2001**

#### 1. Article Details

| Study No.: | 1307 |
| Author & year: | Broeders et al, 2001 |
| Impact Factor (2005): | 4.700 |

Supersedes Van Dijck et al, 1996 (Study No. 2088) and Van Dijck et al, 1994 (Study No. 2353) and Van Dijck et al, 1985 (Study No. 2885). Duplicates the Nijmegen data reported in Broeders et al, 2002 (Study No. 1154), but here it is compare to another city without screening (Arnhem).

#### 2. Description of the study

| Objectives: | Compares breast cancer mortality rates over two decades in city with screening (Nijmegen) with a control city (Arnhem) and to the Netherlands as a whole. NB. Arnhem and rest of Netherlands did have screening from 1989 |
| Location of study: | The Netherlands |
### 3. Study Design and Population

**Nature of screening programme & mammography:**
- **Nijmegen:** Local population based mammography screening programme by invitation. Two view mammography at initial screen, with only mediolateral view subsequently (for ~80%).
- **Arnhem:** Women were offered national screening from 1989.

**Target age for screening participants:**
- **Nijmegen:** 35–65 years initially, then over 35 from 1977, then gradually adjusted to 50–69 between 1989–1997, 50–74 thereafter.
- **Arnhem:** 50–69 years to 1997.

**Target screening interval:** 2 years

**Start date and maturity of programme under evaluation:**
- **Nijmegen:** 1975, with 20-year follow-up
- **Arnhem:** 1989 onward

**Participation rate:**
- **Nijmegen:** 88% in first round, then stabilises as ~66% in subsequent rounds, p304 (NB. this contradicts ~55% after first round reported on p164 of Broeders 2002)
- **Arnhem:** ‘averaging around 77%’ p304

**Study design:**
- Retrospective concurrent cohort

**Entry criteria:**
- N/A

**Recruitment strategy:**
- N/A

**Nature of Comparison Groups:**
- **Group 1 (Reference):** (N= ~19,000 women aged 50-79, exact numbers each year in paper)
- **Group 2 (Intervention):** (N= ~19,000 women aged 50-79, exact numbers each year in paper)

**Outcome/s of interest reported:**
- Breast cancer mortality (age-standardised)

### 4. Bias reduction and statistical methodology

**Data required in order to conduct primary analyses:** Breast cancer deaths, population by age for each geographic area

**At what time points?**

**Statistical analyses performed:**
- Age-standardised mortality ratio for women 50-79 years (ratios of Nijmegen vs. Arnhem and vs. Netherlands as a whole).
  - For each calendar year, the expected number of breast cancer deaths in Nijmegen was derived by applying the age-specific breast cancer mortality rates in Arnhem or the Netherlands to the Nijmegen population (p303). The ratio of the observed to expected deaths is the SMR.
  - Then the trend over time is investigated using a Poisson regression, with extra-Poisson variation added to the resulting model (detail in Appendix of paper). Forces mortality to be the same at beginning of model.
  - Analysis is performed allowing for 2 different assumptions for delay in impact of screening upon mortality (5 yrs and 10 yrs).

**Adjustments performed for inter-group differences:**
- Age standardised to world standard population.

**Investigating effect of extent of screening exposure:**
- Not clear how the introduction of screening in Arnhem and rest of Netherlands after 1989 was accounted for.

**Other methods of bias reduction (incl. loss to follow-up):**
6. Study strengths and weaknesses

Study assumptions:
- Appears the analysis assumes no impact of post 1989 screening in Arnhem and rest of Netherlands?
- Assumes introduction of adjuvant chemotherapy and hormonal therapies was consistent across geographical areas.

Strengths:
- Deaths detected prior to the introduction of screening have not been excluded from death statistics.

Weaknesses:
- No corrections for attendance at screening on an individual level. Also doesn’t account for opportunistic screening in each cohort. Likely to be increased awareness of mammography in Arnhem with the programme underway in Nijmegen.
- Small populations of Nijmegen and Arnhem increase the ‘noise’ in mortality in these areas.
- No information of patient characteristics (e.g., risk factors) in the different cohorts of women.

7. Other comments

CHAMBERLAIN et al., 1988

1. Article Details

Study No.: 3656
Title: First results on mortality reduction in the UK trial of early detection of breast cancer
Author & year: Chamberlain et al., 1988 (on behalf of UK trial of early detection of breast cancer group)
Citation (Jnl, Vol(Num):Pgs.): Lancet 2: 411–416
Impact Factor (2005): 23.878

Superseded by Ellman et al., 1993 (Study No. 2501) and Moss et al., 1999 (Study No. 3926) therefore data not extracted

CHEN et al., 1998

1. Article Details

Study No.: 1781
Title: Evaluation by Markov chain models of a non-randomised breast cancer screening programme in women aged under 50 years in Sweden.
Author & year: Chen et al., 1998
Citation (Jnl, Vol(Num):Pgs.): J Epidemiol Community Health; 52(5):329–35
Impact Factor (2005): 3.003

Also included as a methodological publication

2. Description of the study

Objectives: To apply Markov chain models to data from an uncontrolled service screening (models that have previously used data from RCTs)
Location of study: Uppsala, Sweden
### 3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th>Local service screening in Uppsala county by invitation. Two view at first screen, then single view thereafter (exception if first screen indicated dense breasts).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants:</td>
<td>39-49 years</td>
</tr>
<tr>
<td>Target screening interval:</td>
<td>Every 20 months</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
<td>1988</td>
</tr>
<tr>
<td>Participation rate:</td>
<td>NR</td>
</tr>
<tr>
<td>Study design:</td>
<td>Retrospective analysis of single cohort. Results compared to those observed using RCT data. Markov model based on health states representing phases of tumour development</td>
</tr>
<tr>
<td>Entry criteria:</td>
<td>N/A</td>
</tr>
<tr>
<td>Recruitment strategy:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

#### Nature of Comparison Groups:

| Group 1 (Reference): | N/A |
| Group 2 (Intervention): | (N= 25,660) Women aged 39–49 in Uppsala invited to screening |

#### Outcome/s of interest reported:

Model ultimately predicts breast cancer mortality for screened population relative to a (theoretical) unscreened control group

### 4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | For first, second and third screening round: Number screened, negative cases, prevalent cases with and without nodal involvement. Interval cancers between each round of screening with and without nodal involvement. Time since last screen for interval cancers |
| At what time points?: | For each screening round, and between rounds |
| Statistical analyses performed: | A Markov chain model with transitions between health states. Transitions are independent of the transitions earlier in the model. Spontaneous regression of disease is not possible, and no progression after clinical diagnosis (due to excision). Model used transition probabilities observed in Uppsala data. Tumour status ultimately translated into mortality data using data from the RCT. |

#### Adjustments performed for inter-group differences:

| Investigating effect of extent of screening exposure: | Used data from first 3 screening rounds and intervals immediately following them only, as tumour data from subsequent screens and intervals were very sparse (p330) |

#### Other methods of bias reduction (incl. loss to follow-up):

### 6. Study strengths and weaknesses

#### Study assumptions:

Model assumes inputs for sensitivity and attendance.

#### Strengths:

Results highly modelled, rather than representing actual observed results

#### Weaknesses:

Base model does not adjust for sensitivity and appears to assume an attendance, rather than applying an observed attendance. An additional analysis attempt to adjust for various sensitivities. Use of existing model for tumour characteristic to survival and mortality is specific to the treatments used at the time of model development.

### 7. Other comments
**CHU et al, 1996**

1. **Article Details**

   - **Study No.:** 2031
   - **Title:** Recent trends in U.S. breast cancer incidence, survival, and mortality rates.
   - **Author & year:** Chu et al, 1996
   - **Citation (Jnl, Vol(Num)/Pgs.):** J Natl Cancer Inst 88(21):1571–9
   - **Impact Factor (2005):** 15.171

2. **Description of the study**

   **Objectives:** To determine trends in breast cancer mortality rates, relative to various medical interventions, including mammography screening.

   **Location of study:** USA

3. **Study Design and Population**

   **Nature of screening programme & mammography:** Doesn’t review a particular mammography programme, but recommendations to use screening mammography from age 40 in 1980s.

   **Target age for screening participants:** >40 years

   **Target screening interval:** Start date and maturity of programme under evaluation: 1980s onward

   **Participation rate:** 10-20% in early 1980s, ~20% in 1987, ~40% in 1990, ~50% in 1992. NB. These data for mammography for any reason. Screening mammography thought to be ~70%.

   **Study design:** Descriptive only. Temporal trends only.

   **Entry criteria:** N/A

   **Recruitment strategy:** N/A

   **Nature of Comparison Groups:**

   - **Group 1 (Reference):** N/A
   - **Group 2 (Intervention):** (N= NR, US female population)
   - **Outcome(s) of interest reported:** Breast cancer mortality

4. **Bias reduction and statistical methodology**

   **Data required in order to conduct primary analyses:** Breast cancer mortality data for white women only. Screening participation but only for indirect association with mortality.

   **At what time points?:** Annual data 1969 to 1993

   **Statistical analyses performed:** Linear regression of log-transformed mortality rates to quantify direction and magnitude of temporal trends. Piecewise regression analyses for sudden changes in slope. Trends examined by age group.

   **Adjustments performed for inter-group differences:** Age adjustment to 1970 US standard population.

   **Investigating effect of extent of screening exposure:** No, indirect temporal association only.

   **Other methods of bias reduction (incl. loss to follow-up):**

5. **Study strengths and weaknesses**

   **Strengths:**

   **Weaknesses:** Weak design. Descriptive only. Not possible to differentiate impact of mammography from changes in treatment etc. Data is for entire US female population, irrespective of age.

6. **Other comments**
COBURN et al, 2004

1. Article Details

Study No.: 493
Title: Decreased breast cancer tumor size, stage, and mortality in Rhode Island: An example of a well-screened population.
Author & year: Coburn et al, 2004
Citation (Jnl, Vol(Num):Pgs.): Cancer Control 11(4):222-30
Impact Factor (2005): No impact factor

2. Description of the study

Objectives: To determine the effect of high mammography rates on breast cancer presentation and outcomes (including mortality)
NB. Rhode Island has one of highest screening rates in USA
Location of study: Rhode Island, USA

3. Study Design and Population

Nature of screening programme & mammography: Local mammography screening programme by invitation
Target age for screening participants: Mammography screening promoted to all women
Target screening interval: 2 years for 40–49 years, 1 year for >50 years
Start date and maturity of programme under evaluation: Promoted since the 1980s
Participation rate: Approximately 70% in 1990. Greater than 80% biannual screening since 1997 (p227). Up to 88% n 50–64 years in 1999-2001. See Fig 1 for detail. (NB. results based on survey of sub-group of women)
Study design: Descriptive only. Temporal trends only.
Entry criteria: N/A
Recruitment strategy: N/A

Nature of Comparison Groups:
Group 1 (Reference): N/A
Group 2 (Intervention): (N= NR)
Outcome/s of interest reported: Breast cancer mortality

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality. Age. Data not linked to individual screening attendance.
At what time points?: 1987–2001
Statistical analyses performed: The change in mortality was analysed by comparing the rate ratio of 1987-1989 to 1998–2000 (method not stated, presumably chi-square test as per other outcomes)
Adjustments performed for inter-group differences: Age adjusted to 2000 US standard population
Investigating effect of extent of screening exposure: Does not directly investigate impact of mammography. Indirect temporal association only.
Other methods of bias reduction (incl. loss to follow-up):
6. Study strengths and weaknesses

Study assumptions:

Strengths: Weak design. Descriptive only. Not possible to differentiate impact of mammography from changes in treatment etc.

7. Other comments

COLLETT E et al, 1984

1. Article Details

Study No.: 3532
Title: Evaluation of screening for breast cancer in a non-randomised study (the DOM project) by means of a case-control study.
Author & year: Collette et al, 1984
Citation (Jnl, Vol(Num):Pgs.): Lancet 1(8388):1224–6
Impact Factor (2005): 23.878

2. Description of the study

Objectives: Investigate whether there is any association between risk of breast cancer mortality and participation in a population-based screening programme
Location of study: Utrecht, The Netherlands

3. Study Design and Population

Nature of screening programme & mammography: Local screening study (not programme per se) involving clinical examination and mammography. 75% of eligible women in Utrecht were screened, but not clear if all were invited.
NB. Although not reported here, Collette et al 1992 states all eligible women in Utrecht were encouraged to attend and that if women didn’t participate in one round, they were not invited to the next round.
Target age for screening participants: 50-64 years
Target screening interval: 0, 12, 18 and 24 months
Start date and maturity of programme under evaluation: 1974
Participation rate: 72%, if all were invited.
Study design: Matched case control.
Entry criteria: Eligible population is women eligible for screening within the study period.
Recruitment strategy: see below for cases and referents

Nature of Comparison Groups:

Group 1 (Controls): (N= 138) Controls had to have lived in Utrecht at the time of death of the case, and have the same year of birth as the case. Three controls randomly selected 'with help of local authorities'.

Group 2 (Cases): (N= 46) Defined as a breast cancer death in a woman born between 1911 and 1925, diagnosis and death occurring after the screening project started.

Outcome/s of interest reported: Screening history taken for the time up to and including the date of diagnosis of the case.
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer deaths, with date of diagnosis of cases. Age and screening history for cases and controls.

At what time points?: 1973–1981

Statistical analyses performed: Relative risk of breast cancer death compared between ever-screened and never-screened, using chi-squared statistic. Also stratified by age or birth cohort.

Adjustments performed for inter-group differences: Age matching

Investigating effect of extent of screening exposure: Additional analysis conducted by number of screening examination (with matching ignored)

Other methods of bias reduction (incl. loss to follow-up):

6. Study strengths and weaknesses

Study assumptions: Assumes cases and controls are matched for risk factors other than age i.e., from same population without differences other than screening history.

Strengths: Excludes deaths with a diagnosis before the screening programme.

Screening history extracted from records so not subject to recall bias.

Weaknesses: At high risk of self-selection bias if non-attendees had been invited by chose not to attend. Period of investigation may not be long enough to see impact of mammography upon mortality. Small sample so likely to be consider noise in mortality rates over time.

7. Other comments

Investigates impact of actual screening (i.e., attending screening) rather than availability of a screening programme (i.e., invitation).

COLLETTE et al, 1992

1. Article Details

Study No.: 2593

Title: Further evidence of benefits of a (non-randomised) breast cancer screening programme: The DOM project.

Author & year: Collette et al, 1992

Citation (Jnl, Vol(Num):Pgs.): J Epidemiol Community Health 46(4):382–6

Impact Factor (2005): 3.003

Update of case-control study reported by Collette et al, 1984 (Study No. 3532), but also uses difference evaluation methods.

2. Description of the study

Objectives: To demonstrate the benefits of breast cancer screening on mortality

Location of study: Utrecht, The Netherlands
3. Study Design and Population

Nature of screening programme & mammography: Local screening study (not programme per se) with clinical examination and mammography. Women who did not participate in a particular screening round were not invited to attend the next round. Appears that screening was intended for all women of eligible age in Utrecht (p383).

Target age for screening participants: 50–64 years

Target screening interval: 0, 12, 18, 24 and 48 months.

Start date and maturity of programme under evaluation: 1974, 5 rounds in 10 years.

Participation rate: 72% in first round

Study design: Paper reports 5 different retrospective study designs/evaluation methods.
1. Case-control (focus of this data extraction form, see comments below re other designs)
2. Compare deaths in screened and unscreened (i.e., comparable to the case-control approach)
3. Comparing breast cancer mortality before and after start of programme
4. Comparing breast cancer mortality in different large cities

Entry criteria: N/A

Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Controls): (N= 348) Controls had to have lived in Utrecht at the time of death of the case, and have the same year of birth as the case. Three controls randomly selected ‘with help of local authorities’.

Group 2 (Cases): (N= 116) Defined as a breast cancer death in a woman born between 1911 and 1925, diagnosis and death occurring after the screening project started.

Outcome/s of interest reported: Screening history taken for the time up to and including the date of diagnosis of the case.

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer deaths, with date of diagnosis of cases. Age and screening history for cases and controls.

At what time points?: Data were available up to 1987, six years longer than original analyses (there was also one additional screening round)

Statistical analyses performed: Odds ratio for breast cancer death between screened and not screened. Also stratified by age or birth cohort.

Adjustments performed for inter-group differences: Age matching

Investigating effect of extent of screening exposure: Additional analysis conducted by number of screening examination (with matching ignored). See notes re Method 5 below.

Other methods of bias reduction (incl. loss to follow-up):

6. Study strengths and weaknesses

Study assumptions: Assumes cases and controls are matched for risk factors other than age i.e., from same population without differences other than screening history.

Strengths: Authors state ‘no evidence of confounding’ (abstract). Screening history extracted from medical records so not subject to recall bias.

Weaknesses: However, authors also state (p384) that a ‘healthy screened effect’ has been reported for this study elsewhere, but then discount this as a bias against screening because these women are at risk of dying for a longer time. This appears counterintuitive when analysis period is of equal duration for all women, and a healthy screened effect could contribute to reduced risk factors for breast cancer. Despite authors contention of ‘no confounding’ the study is at high risk of self-selection bias if non-attendees had been invited by chose not to attend. Small sample so likely to be consider noise in mortality rates over time.
7. Other comments

Investigates impact of actual screening (i.e., attending screening), rather than availability of a screening programme (i.e., invitation).
Method 2 is essential comparable to the case-control results.
Method 3 compares mortality before and after commencement of the programme, and therefore can not isolate effect of mammography versus other factors. Only indirectly investigates impact of mammography from temporal changes.
Method 4 compares mortality in Utrecht with mortality in similar cities. Descriptive comparison of temporal trends only. NB. Analyses appears to be of mortality for the same cohort over time, as authors state their increasing age is the reason for the increase over time in the other cities.
Method 5 uses the fact that women in Utrecht only got screens for certain time period to investigate the extent of screening activity upon mortality. Descriptive comparison of temporal trends only.

DAS et al., 2005

1. Article Details

Study No.: 220
Title: Geographic association between mammography use and mortality reduction in the US.
Author & year: Das et al, 2005
Citation (Jnl, Vol(Num),Pgs.): Cancer Causes Control 16(6):691–9
Impact Factor (2005): 3.195
NCI analysis (Updates Chu et al, Study No. 2031)

2. Description of the study

Objectives: To examine population data on breast cancer screening and mortality to see if there is any geographic association between these factors, after adjusting for therapy use.
Location of study: USA

3. Study Design and Population

Nature of screening programme & mammography: Doesn’t review a particular mammography programme, but recommendations to use screening mammography from age 40 in 1980s
Target age for screening participants: >40 years
Target screening interval: NR
Start date and maturity of programme under evaluation: NR
Participation rate: NR
Study design: Descriptive only
Entry criteria: N/A
Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Reference): N/A
Group 2 (Intervention): N/A
Outcome/s of interest reported: Breast cancer mortality
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality, mammography use (or surrogate for it), data on adjuvant therapy use complete with age and stage of disease for adjustments. Individual level data not required.

At what time points?: Annual data, 1970–2000, exact years used in analyses is not reported.

Statistical analyses performed: Regression analysis of estimated annual percent reduction in breast cancer mortality was performed on mammography use at the state level, with mammography usage from BRFSS survey information. A secondary regression analyses was performed to adjust for use of adjuvant therapy, at health service level.

Annual percent change in incidence of early cancer was used as surrogate for mammography use at the HSA level, as survey data (BRFSS) was insufficient (p694).

Adjustments performed for inter-group differences: Mortality rate age-adjusted to 2000 US standard population. Adjuvant therapy usage also adjusted for age and stage of disease.

Investigating effect of extent of screening exposure: None

Other methods of bias reduction (incl. loss to follow-up):

6. Study strengths and weaknesses

Study assumptions: Annual percent change in incidence of early cancer was used as surrogate for mammography use

Strengths:

Weaknesses: Measures of therapy and mammography are somewhat coarse and subject to inaccuracies. Data analysed at state and HSA level, not at individual patient level, therefore ecological bias may exist. Second analyses by HSA level relies upon stage at diagnosis as a proxy for extent of mammography. Other factors (increased awareness, self examination etc) could contribute to stage shift.

7. Other comments

DUFFY et al, 2006A

1. Article Details

Study No.: 70
Title: Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data.
Author & year: Duffy et al, 2006a for Swedish Organised Service Screening Evaluation Group
Citation (Jnl, Vol(Num):Pgs.): Cancer Epidemiol Biomarkers Prev 15(1):45-51
Impact Factor (2005): 4.460
Companion paper describing alternative method is Duffy et al, 2006b (Study No. 69)

2. Description of the study

Objectives: Report the effect of the introduction of mammography screening upon breast cancer mortality
Location of study: 13 geographic areas in 9 counties Sweden (includes the ‘7 counties’ reported elsewhere)
A review of methodological options for evaluating the impact of BreastScreen Australia on breast cancer mortality

December 2006

3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th>Service mammography screening</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants:</td>
<td>40–69 in eight areas, 50–69 in five areas</td>
</tr>
<tr>
<td>Target screening interval:</td>
<td>~2 years (p49)</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
<td>Date of commencement varied by county</td>
</tr>
<tr>
<td>Participation rate:</td>
<td>range 70% to 90% (Table 2 and p48)</td>
</tr>
<tr>
<td>Study design:</td>
<td>Retrospective before and after study (historical cohort)</td>
</tr>
<tr>
<td>Entry criteria:</td>
<td>N/A</td>
</tr>
<tr>
<td>Recruitment strategy:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Nature of Comparison Groups:

| Group 1 (Reference):          | (N= 542,187) |
| Total (average) population in the pre-screening epoch |
| Group 2 (Intervention):       | (N= 566,423) |
| Total (average) population in the screening epoch |

Outcome/s of interest reported: Breast cancer mortality (incidence-based mortality i.e., deaths only from tumours diagnosed in the screening epoch). NB. companion paper Duffy et al 2006b uses all breast cancer mortality (not just incidence based)

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality by geographical area. Actual screening attendance data, not invitation to screening, i.e., screening history for individual deaths.

At what time points?: For equal period of time before and after introduction of screening, within an area. Varied between areas.

Statistical analyses performed: Used Poisson regression to compare deaths in the pre-screening epoch from tumours diagnosed in that epoch, with the corresponding deaths in the screening epoch (p47). Also separated screening deaths and person-years by screening exposure, and estimated the change in mortality compared with the pre-screening epoch in the screening-exposed and unexposed groups separately. Results from all areas were combined using the inverse variance weighted averages of the relative risks in the logarithmic scale.

Adjustments performed for inter-group differences: Corrected for self-selection bias (as described in Duffy et al, 2002) but with the additional refinements of i) adjusting after estimating the effect of being screened vs. being invited to screening and ii) using each area’s own relative risk for death among the unexposed group in the screening epoch to adjust for selection bias, instead of the estimated relative risk from the RCTs. (see p47-48)

Investigating effect of extent of screening exposure: Compares screening exposure and non-exposure amongst screening epoch, but not volume of screening.

Other methods of bias reduction (incl. loss to follow-up): Analysis restricted to women under 70. Required equal follow-up in each area between pre-screening epoch and screening epoch.

6. Study strengths and weaknesses

Study assumptions:

Strengths: Distinguished between those invited to screening and those attending

Weaknesses: Adjustment for self-selection may be inaccurate, as is based on relative risk of unexposed in the screening epoch with pre-screening epoch (assumes this is due to selection bias). But this is confounded by any other changes occurring over time. Effect may be due to changes in therapy over time, particularly use of adjuvant hormonal and chemotherapy. This is not adjusted for in the study.

7. Other comments
DUFFY et al, 2006b

1. Article Details

| Study No.: | 69 |
| Title: | Reduction in breast cancer mortality from the organised service screening with mammography: 2. Validation with alternative analytic methods. |
| Author & year: | Duffy et al, 2006b for Swedish Organised Service Screening Evaluation Group |
| Citation (Jnl, Vol(Num):Pgs.): | Cancer Epidemiol Biomarkers Prev 15(1):52-6 |
| Impact Factor (2005): | 4.460 |

Companion paper describing alternative method is Duffy et al, 2006a (Study No. 70)

2. Description of the study

Objectives: Report the effect of the introduction of mammography screening upon breast cancer mortality

Location of study: 13 geographic areas in 9 counties Sweden (includes the ‘7 counties’ reported elsewhere)

3. Study Design and Population

Nature of screening programme & mammography: Service mammography screening

Target age for screening participants: 40–69 in eight areas, 50–69 in five areas

Target screening interval: ~2 years

Start date and maturity of programme under evaluation: Date of commencement varied by county

Participation rate: range 70% to 90% (from Duffy et al, 2006a)

Study design: Retrospective before and after study (historical cohort)

Entry criteria: N/A

Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Reference): (N= 555,677) Total (average) population in the entire period of study, not reported just for pre-screening epoch

Group 2 (Intervention): (N= 555,677) Total (average) population in the entire period of study, not reported just for screening epoch

Outcome/s of interest reported: All breast cancer mortality (NB. companion paper Duffy et al 2006a uses incidence-based mortality i.e., deaths only from tumours diagnosed in the screening epoch)

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality by geographical area. Actual screening attendance data, not invitation to screening, i.e., screening history for individual deaths.

At what time points?: Varied between areas. For this method, the time does not need to be equal for pre-screening and screening epochs (in contrast with companion paper)

Statistical analyses performed: Poisson regression used, and for each year in the period of observation, all breast cancer deaths are included, irrespective of year of diagnosis. Death was linked to time of and exposure status at diagnosis. The log-linear model includes year of diagnosis, year of death, screening exposure and a correction for lead-time (p53). There is also a term representing the change in fatality independent of screening (such as changes in therapy), and a term reflecting the trend of change in incidence over time (p53).

Adjustments performed for inter-group differences: Corrected for lead-time (see paper for details p53). Corrected for self-selection bias (as described in more detail in paper). This paper (but not companion paper) accounts for changes in incidence and fatality of breast cancers taking place during the period of study independent of screening.

Investigating effect of extent of screening exposure: None (just exposed, unexposed)

Other methods of bias reduction (incl. loss to follow-up): Analysis restricted to women under 70.
6. Study strengths and weaknesses

Study assumptions:

Strengths: Distinguished between those invited to screening and those attending. Includes a correction for lead-time (and tests different methods of this as a sensitivity analysis). Adjusts for contemporaneous changes in incidence and fatality (companion paper did not).

Weaknesses: Adjustment for self-selection may be open to debate

7. Other comments

DUFFY et al, 2002

1. Article Details

Study No.: 1015
Title: The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish Counties: A collaborative evaluation.
Author & year: Duffy et al, 2002
Citation (Jnl, Vol(Num):Pgs.): Cancer 95(3):458–69
Impact Factor (2005): 4.800

'7 counties’ study. Fewer areas than included in Duffy et al, 2006a and 2006b. NB. Also less sophisticated methodology.

2. Description of the study

Objectives: To investigate the impact of organised mammography service screening on breast cancer mortality
Location of study: 7 counties, Sweden

3. Study Design and Population

Nature of screening programme & mammography: Local service mammography screening (NB. in one county a RCT was ongoing for portion of time)
Target age for screening participants: Varied by county (see Table 1, p461)
Target screening interval: varied by county: 1.5–2 years
Start date and maturity of programme under evaluation: Varied by county.
Participation rate: Varied by county (refer Table 1, p461)
Study design: Retrospective before and after study (historical cohort), but also reports comparison of concurrent data for exposed and unexposed women within the screening epoch (retrospective concurrent cohort).
Entry criteria: N/A
Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Reference): \(N=\ 3.7\text{ million person-years}\) Pre-screening epoch
Group 2 (Intervention): \(N=\ 3.8\text{ million person-years}\) Screening epoch
Outcome/s of interest reported: Breast cancer mortality. Excludes deaths diagnosed before the start of the screening epoch
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses:
Breast cancer mortality by geographical area. Actual screening attendance data, not invitation to screening, i.e., screening history for individual deaths.

At what time points?:
Equal duration for pre-screen and screen epoch

Statistical analyses performed:
Poisson regression to compare mortality in the screening epoch with the pre-screening epoch. Also, separately, compared death rates between the exposed and unexposed women within the screening cohort. There was correction for lead-time in the two counties with substantial screening occurring within the pre-screening period.

Adjustments performed for inter-group differences:
Corrections were made for lead-time bias in two counties only
Correction for self-selection bias (based on RCT evidence).

Investigating effect of extent of screening exposure:
None (just exposed, unexposed)

Other methods of bias reduction (incl. loss to follow-up):
Excludes deaths where the diagnosis occurred before the start of the screening epoch.
Analysis restricted to women under 70.

6. Study strengths and weaknesses

Strengths:

Weaknesses:
In one county there was 14% screening in the pre-screening period, and in another a RCT was ongoing with 33% participation.
Main analysis doesn’t directly correct for changes in treatment over time, and increased awareness, self-examination etc. Separate analysis attempts to determine extent of these by comparing exposed and unexposed women within the screened group.

7. Other comments

ELLMAN et al, 1993

1. Article Details

Study No.: 2501
Title: Breast cancer mortality after 10 years in the UK trial of early detection of breast cancer.
Author & year: Ellman et al, 1993, on behalf of the UK Trial of Early Detection of Breast Cancer group
Citation (Jnl, Vol(Num):Pgs.): Breast 2(1):13-20
Impact Factor (2005): 1.705

Superseded by Moss et al, 1999, Study No. 3926 (same approach but longer follow-up), therefore data not extracted.

ELMORE et al, 2005

1. Article Details

Study No.: 180
Title: Efficacy of breast cancer screening in the community according to risk level.
Author & year: Elmore et al, 2005
Citation (Jnl, Vol(Num):Pgs.): J Natl Cancer Inst 97(14):1035-43
Impact Factor (2005): 15.171
2. Description of the study

Objectives: To assess the efficacy of community-based breast cancer screening among women in two different age groups, and at two different levels of breast cancer risk.

Location of study: Six health plans in USA

3. Study Design and Population

Nature of screening programme & mammography: Community-based screening by mammography and/or clinical examination within 6 health care plans.

Target age for screening participants: NR

Target screening interval: NR

Start date and maturity of programme under evaluation: NR

Participation rate: See Table 2 and Fig 1 for rate of screening for cases and controls. Rate of mammography screening is low (~30% in average risk groups and ~50% in high risk groups). Clinical examination alone screening is significant.

Study design: Retrospective matched case-control study.

Entry criteria: N/A

Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Controls): (N= 2501)
Controls had to be free of breast cancer at their index date, within one year of age of the case, and within the same age group (40–49 yr, 50–65 yr) when the case was diagnoses, alive on the date the case had died, and continuously enrolled in a health plan during the index period, and who were active health plan members at the time of the cases' breast cancer diagnosis. Index date= same index date as their ‘case’. Index period= 3 year period leading up to index date.

Group 2 (Cases): (N= 1351)
Case defined as women who had died of breast cancer between Jan 1983 and Dec 1998, who had an initial diagnosis between Jan 1983 and Dec 1993, who was between 40 and 65 years at diagnosis, and who was continuously enrolled in a health plan during the index period, and who were active health plan members at the time of breast cancer diagnosis. Index date= date of diagnosis. Index period= 3 year period leading up to index date.

Outcome/s of interest reported: Occurrence of cancer screening during the index period.

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer deaths, with date of diagnosis for cases. Age, risk factors and screening history for cases and controls. Mammograms were identified as screening or diagnostic from patient records.

At what time points?: Period reviewed for deaths was 15 years. Screening information obtained over 3 years.

Statistical analyses performed: Primary analysis involved calculation of odds ratio using conditional logistic regression retaining matching by health plan, calendar time, age and breast cancer risk level.

Adjustments performed for inter-group differences: Model also included other confounder such as race, co-morbidity, and age at first birth.

Investigating effect of extent of screening exposure: None.

Other methods of bias reduction (incl. loss to follow-up): Only screening mammograms were included.
6. Study strengths and weaknesses

Study assumptions:

| Strengths: | Large study. Includes data on risk factors (e.g., race co-morbidity, age at first birth, family history), so able to match for this between cases and controls and also to conduct analyses by risk sub-groups. Screening history extracted from records so not subject to recall bias. |
| Weaknesses: | Low proportion with screening. Potential confounding by high level of clinical screening. Index period was only the three years prior to date of diagnosis. Data extraction from medical records subject to error, especially with regard to identification of risk factors. |

7. Other comments

Investigates impact of actual screening (i.e., attending screening), rather than availability of a screening programme (i.e., invitation).

FIELDER et al, 2004

1. Article Details

| Study No.: | 661 |
| Title: | A case-control study to estimate the impact on breast cancer death of the breast screening programme in Wales. |
| Author & year: | Fielder et al, 2004 |
| Citation (Jnl, Vol(Num):Pgs.): | J Med Screen 11(4):194-8 |
| Impact Factor (2005): | 2.483 |

2. Description of the study

| Objectives: | To estimate the effect of service screening on breast cancer mortality in Wales |
| Location of study: | Wales, UK |

3. Study Design and Population

| Nature of screening programme & mammography: | National mammography screening programme. Mammograms are double read. |
| Target age for screening participants: | 50–64 years |
| Target screening interval: | 3 years |
| Start date and maturity of programme under evaluation: | 1989 ("fully underway by 1991"), 10 years + |
| Participation rate: | NR |
| Study design: | Retrospective match case-control |
| Entry criteria: | N/A |
| Recruitment strategy: | N/A |
### Nature of Comparison Groups:

<table>
<thead>
<tr>
<th>Group 1 (Controls)</th>
<th>(N= 717)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Two controls matched to case by year of birth, one from same GP practice and one from different GP practice within the same district.</td>
</tr>
<tr>
<td></td>
<td>N8. Controls could have breast cancer, as long as they were alive at the time the case was diagnosed (p195)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 2 (Cases)</th>
<th>(N= 419)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Deaths with breast cancer as either the primary or secondary cause of death, in women aged 50–75 years at diagnosis, who were diagnosed after the instigation of screening in 1991. Cases were selected from the Breast Test Wales database by working through the list of deaths, most recent first, until the required number had been identified.</td>
</tr>
</tbody>
</table>

| Outcome/s of interest reported: | Occurrence of cancer screening during the index period. |

### 4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | Breast cancer mortality (as primary or secondary cause of death), with date of diagnosis. Screening history for both cases and controls. |
| At what time points?: | See note above re. time period for death of cases. |
| Statistical analyses performed: | Conditional logistic regression. |
| Adjustments performed for inter-group differences: | Self-selection bias corrected for using method of Duffy et al, 2002. In addition another analysis was run adjusting for socioeconomic status index. |
| Investigating effect of extent of screening exposure: | Analyses performed for ever vs never screened and also investigating number of screens. |
| Other methods of bias reduction (incl. loss to follow-up): | Also excludes any cancers diagnosed early in the programme (see p195 ‘screen-detected cases’ for detailed explanation) |

### 6. Study strengths and weaknesses

| Study assumptions: | Assumes cases and controls are matched for risk factors other than age i.e., from same population without differences other than screening history. |
| Strengths: | Excludes cases diagnosed prior to the screening programme. Excludes any screening occurring in controls after date of diagnosis of their case. |
| Screening history extracted from records so not subject to recall bias. | |
| Weaknesses: | Does the selection of most recent cases introduce bias? |

### 7. Other comments

|  | Investigates impact of actual screening (i.e., attending screening) , rather than availability of a screening programme (i.e., invitation). |
GARNE et al, 1997

1. Article Details

| Study No.: | 2006 |
| Author & year: | Garne et al, 1997 |
| Citation (Jnl, Vol(Num):Pgs.): | Cancer 79(1):69–74 |
| Impact Factor (2005): | 4.800 |

2. Description of the study

| Objectives: | To investigate age-adjusted breast cancer mortality 1964–1992 (only indirectly related to introduction of mammography) |
| Location of study: | Malmo and rest of Sweden |

3. Study Design and Population

| Nature of screening programme & mammography: | NR, not a major focus of the paper |
| Target age for screening participants: | 45–69 years then 40–74 years. Since 1991 offered to all women 50–69 years |
| Target screening interval: | NR |
| Start date and maturity of programme under evaluation: | Mammography occurred in Malmo as a result of the RCT 1976-1986, and then as a result of Swedish health authority recommendations. |
| Participation rate: | NR |
| Study design: | Descriptive only. Temporal trends only. |
| Entry criteria: | N/A |
| Recruitment strategy: | N/A |

Nature of Comparison Groups:

| Group 1 (Reference): | (N= ~120,000) Typical number of women in Malmo over this time period |
| Group 2 (Intervention): | NR |
| Outcome/s of interest reported: | Breast cancer mortality |

4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | Breast cancer deaths. Age. |
| At what time points?: | 1964-1992 |
| Statistical analyses performed: | Poisson regression to investigate changes in mortality over time. Impact of mammography not investigated statistically. |
| Adjustments performed for inter-group differences: | Age-standardised to 1970 Swedish population |
| Investigating effect of extent of screening exposure: | No, indirect temporal association only. |
| Other methods of bias reduction (incl. loss to follow-up): | 3 year moving average applied to descriptive mortality (but not Poisson analysis) |

6. Study strengths and weaknesses

| Study assumptions: |
| Strengths: |
| Weaknesses: | Weak design. Descriptive only. Not possible to differentiate impact of mammography from changes in treatment etc. |

7. Other comments
1. Article Details

| Study No.: | 748 |
| Title: | Evaluation of the organised mammographic screening programme in Australia. |
| Author & year: | Giles & Amos, 2003 |
| Citation (Jnl, Vol(Num),Pgs.): | Ann Oncol. 14(8):1209-11 |
| Impact Factor (2005): | 4.319 |

2. Description of the study

Objectives: NR, but to provide some evidence of potential benefit (p1210). Not primary intention of paper to measure changes in mortality at this stage.

Location of study: Australia

3. Study Design and Population

Nature of screening programme & mammography: National screening mammography programme. Two view mammography. Free for women 50–69 years (by invitation), but accessible to all women ≥40 years

Target age for screening participants: Free for women 50–69 years, but accessible to all women ≥40 years

Target screening interval: 2 years (for 50–69 years)

Start date and maturity of programme under evaluation: 1992, 10 years

Participation rate: Aim to achieve 60% participation after 5 years. In 1997–98 was 67.4%

Study design: Descriptive only. Temporal trends only.

Entry criteria: N/A

Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Reference): N/A

Group 2 (Intervention): NR

Outcome/s of interest reported: Breast cancer mortality

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality (age-specific)

At what time points?: 1983–1999

Statistical analyses performed: None, descriptive only

Adjustments performed for inter-group differences: None

Investigating effect of extent of screening exposure: None

Other methods of bias reduction (incl. loss to follow-up): None

5. Study strengths and weaknesses

Study assumptions:

Strengths:

Weaknesses: Programme too young to observe changes in mortality. Weak design. Descriptive only. Not possible to differentiate impact of mammography from changes in treatment etc.

6. Other comments

A review of methodological options for evaluating the impact of BreastScreen Australia on breast cancer mortality

December 2006
GORINI et al, 2004

1. Article Details

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>3715</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Breast cancer mortality trends in two areas of the province of Florence, Italy, where screening programmes started in the 1970s and 1990s.</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>Gorini et al, 2004</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.):</td>
<td>Br J Cancer 90(9):1780-3</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>4.115</td>
</tr>
</tbody>
</table>

Considerable overlap with data in Paci and Palli papers but different methodology.

2. Description of the study

Objectives: To compare breast cancer mortality in areas with early introduction of screening (1970s) and late introduction of screening (1990s)

Location of study: Province of Florence, Italy

3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th>Local mammography screening programmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES: Mammography screening programme in 23 rural municipalities in province of Florence that commenced screening early</td>
<td></td>
</tr>
<tr>
<td>LS: Mammography screening programme in rest of province (including city of Florence) that commenced screening late</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target age for screening participants:</th>
<th>ES: 40–69 years up to 1989, then 50-69 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LS: 50–69 years</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target screening interval:</th>
<th>NR</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Start date and maturity of programme under evaluation:</th>
<th>ES: 1970s, ~20 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS: 1990s, ~10 years</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participation rate:</th>
<th>ES: ~ 8000-9000 mammograms per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>LS: ~ NR just for LS area.</td>
<td></td>
</tr>
</tbody>
</table>

NB. “relatively low compliance in the first years of the programme in both areas (abstract)”

<table>
<thead>
<tr>
<th>Study design:</th>
<th>Retrospective cohort study, comparing two areas that started mammography at two different time points: early (ES) and late (LS)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Entry criteria:</th>
<th>N/A</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Recruitment strategy:</th>
<th>N/A</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Nature of Comparison Groups:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1 (Reference): (N= NR)</td>
</tr>
<tr>
<td>Late screening</td>
</tr>
<tr>
<td>Group 2 (Intervention): (N= 70,000 women 25 years and over)</td>
</tr>
<tr>
<td>Early screening</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outcomes of interest reported:</th>
<th>Breast cancer mortality</th>
</tr>
</thead>
</table>
4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | Breast cancer mortality. Age. Individual mammography screening history not required. |
| At what time points?: | Annual data over 10-20 years period. |
| Statistical analyses performed: | To estimate percent change in rates for period for trends in mortality were examined using a log-linear regression model with the year of death as a continuous variable. Analyses conducted by age group. NB. Change statistics are within each group (not between them) |
| Adjustments performed for inter-group differences: | Age-standardised to European Standard population. |
| Investigating effect of extent of screening exposure: | None |
| Other methods of bias reduction (incl. loss to follow-up): | None |

6. Study strengths and weaknesses

Study assumptions:

**Strengths:**
- Descriptive only, estimated percent change statistics are within each group (not between them).
- Likely to be contamination by low compliance, and no accounting for mammography in LS area prior to commencement of formal programme.
- Not possible to differentiate impact of mammography from changes in treatment etc.
- Likely to be considerable differences in other risk factors and access to treatment between areas (ES primarily rural and LS primarily urban). Although bias may be in opposite direction, magnitude of impact may not be generalisable to a more urban population.

**Weaknesses:**
- Likely to be contamination by low compliance, and no accounting for mammography in LS area prior to commencement of formal programme.
- Not possible to differentiate impact of mammography from changes in treatment etc.
- Likely to be considerable differences in other risk factors and access to treatment between areas (ES primarily rural and LS primarily urban). Although bias may be in opposite direction, magnitude of impact may not be generalisable to a more urban population.

7. Other comments

**HAKAMA et al, 1995**

1. Article Details

| Study No.: | 2177 |
| Title: | Effectiveness of screening for breast cancer in women under 50 years at entry: The Kotka pilot project in Finland. |
| Author & year: | Hakama et al, 1995 |
| Citation (Jnl, Vol(Num):Pgs.): | Int J Cancer: 63(1):55–7 |
| Impact Factor (2005): | 4.700 |

Superseded by Hakama et al, 1999 (Study No. 1663), which uses same methodology.

**HAKAMA et al, 1997**

1. Article Details

| Study No.: | 1954 |
| Title: | Effectiveness of the public health policy for breast cancer screening in Finland: Population based cohort study |
| Author & year: | Hakama et al, 1997 |
| Impact Factor (2005): | Superseded by Hakama et al, 1999 (Study No. 1663), which uses same methodology. |
### 1. Article Details

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>1663</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Implementation of screening as a public health policy: Issues in design and evaluation.</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>Hakama et al, 1999</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num)/Pgs.):</td>
<td>J Med Screen; 6(4):209-16</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>Supersedes Hakama 1995 (Study No. 2177) and Hakama 1997 (Study No. 1954). Also included as a methodological publication</td>
</tr>
</tbody>
</table>

### 2. Description of the study

**Objectives:** To propose principles of design and measures of effect for cancer screening programme as a public health policy (abstract)

**Location of study:** Finland

### 3. Study Design and Population

**Nature of screening programme & mammography:** National mammography programme with women identified and invited by birth cohort. Two view mammography, interpretation is by two radiologists.

NB. Earlier paper (Hakama et al, 1995) states that the programme included clinical breast examination and information re breast self-examination.

**Target age for screening participants:** 50–59, but women can continue to be screened to 64 (p209)

**Target screening interval:** 2 years

**Start date and maturity of programme under evaluation:** Gradually implemented since 1987 (in three age cohorts) but then expanded to all even year cohorts. Maturity 8 years for first cohort. (see Fig 1, 209 for screening by age cohort)

**Participation rate:** Participation is ~90% for those invited by municipal councils (p210). NB. But not all municipal councils within Finland had programmes.

**Study design:** NB. Several methods. Retrospective before and after study (historical cohort), also with comparisons between control age cohorts (who started at least 2 years later) (but with no accounting for municipality or individual participation in screening). Retrospective additional individual level analyses by intention to screen at an individual level with individual controls who were cluster randomised and matched for age.

**Entry criteria:** N/A for ecological analyses

**Recruitment strategy:** N/A

**Nature of Comparison Groups:**

- **Group 1 (Reference):** (N= 68,862) largely odd birth year cohorts
- **Group 2 (Intervention):** (N= 89,893 invited, 76,389 screened) largely even year cohorts

**Outcome/s of interest reported:** Breast cancer mortality (reported as SMRs)
### 4. Bias reduction and statistical methodology

**Data required in order to conduct primary analyses:**
Breast cancer mortality, including only those diagnosed after introduction of screening. For individual analyses, invitations and screening records at an individual level, and the ability to identify individual ‘controls’, matched for municipality and approximate age (adjacent birth years). N8. Requires programme to have been introduced in a gradual fashion determined by birth year cohorts.

**At what time points?:**
Annual over 9 years

**Statistical analyses performed:**
No statistical analyses performed on before and after and adjacent age cohort data. Descriptive only.

Rationale for individual level method:
It is possible to approximate the individually randomised trial in the gradual implementation phase of mammography programme.

Paper describes three possible analyses at the individual level (p211). States best is ‘internal control – refined mortality’. This involves making comparison within each municipality between cohorts that were classified by screening invitation status and year of birth (even vs. odd birth years), with subsequent consolidation for all the municipalities taking part in programme (p211). In this analysis, mortality is confined to deaths diagnosed after the onset of screening.

**Adjustments performed for inter-group differences:**
None

**Investigating effect of extent of screening exposure:**
None

**Other methods of bias reduction (incl. loss to follow-up):**

### 6. Study strengths and weaknesses

**Study assumptions:**
Assumes the one year difference between odd birth year cohorts and even birth year controls does not introduce bias (p212), and that these cohorts are fundamentally comparable.

Assumes screening amongst controls in 1990–92 does not effect mortality for cancers diagnosed in 1990-92 (p212)

**Strengths:**
Author’s state ‘internal control - refined mortality’ analyses removes any bias due to self-selection of either municipalities or individual to attend the programme.

Authors state that using ‘refined mortality’ removes bias and dilution.

**Weaknesses:**
Comparison between early intervention cohorts and adjacent cohorts assumes difference in mortality would be visible in 1990-95 (i.e., − 3–8 years after screening commences in early cohorts (p210). Follow-up time without contamination of introduction of screening may be too short.

‘Internal control – refined mortality’ method assumes comparability in risk between adjacent age cohorts.

### 7. Other comments

**JONSSON et al, 2000**

**1. Article Details**

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>1399</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Service screening with mammography in Sweden: Evaluation of effects of screening on breast cancer mortality in age group 40–49 years.</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>Jonsson et al, 2000</td>
</tr>
<tr>
<td>Citation (Jnl. Vol(Num):Pgs.):</td>
<td>Acta Oncol 39(5):617–23</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>2.362</td>
</tr>
</tbody>
</table>

N8. Same methodology as Jonsson et al, 2001 (Study No. 3687), Jonsson et al, 2003 (Study No. 764) and Jonsson et al, 2003 (Study No. 823)

Data extracted for Jonsson et al, 2001 (Study No. 3687), where age of interest was 50–69 years.
<table>
<thead>
<tr>
<th>Study No.</th>
<th>3687</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Service screening with mammography of women aged 50-69 years on Sweden: effects on mortality from breast cancer.</td>
</tr>
<tr>
<td>Author &amp; year</td>
<td>Jonsson et al, 2001</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.)</td>
<td>J Med Screen: 8(3):152–60</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>2.483</td>
</tr>
</tbody>
</table>

Methodology replicates that used in Jonsson et al, 2000 (Study No. 1399), Jonsson et al, 2003 (Study No. 764) and Jonsson et al, 2003 (Study No. 823)

2. Description of the study

Objectives: To estimate the effect of the population based service screening programme in Sweden on mortality from breast cancer among women aged 50–69.

Location of study: Sweden (majority of counties)

3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th>Nationwide mammography screening programme. 7 geographical areas had an early start to mammography screening (study areas), whilst 5 had a late start to mammography screening (control areas)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants:</td>
<td>50–69 years</td>
</tr>
<tr>
<td>Target screening interval:</td>
<td>Variable, but mean screening interval 23 months.</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
<td>Study areas: 1986-7 Control areas: 1993 or later Mean follow-up 10.6 years</td>
</tr>
<tr>
<td>Participation rate:</td>
<td>NR</td>
</tr>
<tr>
<td>Study design:</td>
<td>Retrospective concurrent cohort, with a historic reference period</td>
</tr>
<tr>
<td>Entry criteria:</td>
<td>N/A</td>
</tr>
<tr>
<td>Recruitment strategy:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Nature of Comparison Groups:

<table>
<thead>
<tr>
<th>Group 1 (Reference):</th>
<th>(N= 98,608 women) Control area with late starting programmes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 (Intervention):</td>
<td>(N= 161,986 women) Study area with early starting programmes</td>
</tr>
</tbody>
</table>

Outcome/s of interest reported: Breast cancer mortality (‘refined’ mortality = deaths from breast cancers diagnosed ‘after a certain time point and in a certain age group’). Cumulative refined mortality was computed with the mean number of person-years as denominator (person-years divided by years of follow-up)
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses:
- Annual age-specific breast cancer mortality. Mortality was smoothed with the Lowess method.
- Number of women and duration of follow-up. Additional data was obtained from breast cancer deaths (date, age and residence at diagnosis, date and cause of diagnosis).

At what time points:
- Annual data for study period and reference period.

Statistical analyses performed:
- Cumulative relative risks were estimated to determine the impact of screening availability.
- The refined breast cancer mortality was also analysed with a multiplicative Poisson model with the number of breast cancer deaths as dependent variable, and year of follow-up, age during follow-up, geographical area, and period all as categorical co-variates (p154). Screening effect was measured as a dummy variable (1 for study area cohorts in the study period, and 0 for elsewhere). The logarithm of the number of person-years in each cell in the cohorts were taken as offset.

Adjustments performed for inter-group differences:
- To adjust for possible geographical differences in mortality between the study areas and control areas, the RR for the study period was divided by the RR for the reference period. Authors state that this adjustment also corrects for the slight difference in duration of follow-up between the study group and the control group (p154).

Investigating effect of extent of screening exposure:
- None

Other methods of bias reduction (incl loss to follow-up):
- 'Inclusion bias' and 'lead time bias' are both possible in this study. Allowances for these are discussed in detail in the publication (p155).

6. Study strengths and weaknesses

Study assumptions:
- The ratio of relative risks assumes multiplicative effects between the groups and the time periods.

Strengths:
- Uses reference period to correct for geographical difference not due to screening.

Weaknesses:
- Overlap between reference and study periods.
- Age was defined as age at diagnosis in this study. There may have been an interaction between lead-time and age, with a women in the unscreened group assigned an age that was older than a comparable woman in the screened group (due to diagnosis at an earlier stage of disease). Impact of this is discussed on p155.
- May have been opportunistic screening in control group

7. Other comments

JONSSON et al, 2003

1. Article Details

Study No.: 764
Title: Service screening with mammography of women aged 70-74 years in Sweden: Effects on breast cancer mortality.
Author & year: Jonsson et al, 2003
Citation (Jnl, Vol(Num):Pgs.): Cancer Detect Prev 27(5):360–9
Impact Factor (2005): 1.599

NB. Same methodology as Jonsson et al, 2001 (Study No. 3687), Jonsson et al, 2000 (Study No. 1399) and Jonsson et al, 2003 (Study No. 823)
Data extracted for Jonsson et al, 2001 (Study No. 3687), where age of interest was 50–69 years.
1. Article Details

**JONSSON et al, 2003**

- **Study No.**: 823
- **Title**: Service screening with mammography. Long-term effects on breast cancer mortality in the county of Gavleborg, Sweden.
- **Author & year**: Jonsson et al, 2003
- **Citation (Jnl, Vol(Num):Pgs.):** Breast: 12(3):183–93
- **Impact Factor (2005):** 1.705

Note: Same methodology as Jonsson et al, 2000 (Study No. 1399), Jonsson et al, 2001 (Study No. 3687) and Jonsson et al, 2003 (Study No. 764). Data extracted for Jonsson et al, 2001 (Study No. 3687).

**KRICKER et al, 1999**

- **Study No.**: 1630
- **Title**: Breast cancer in New South Wales in 1972-1995: Tumor size and the impact of mammographic screening.
- **Author & year**: Kricker et al, 1999
- **Citation (Jnl, Vol(Num):Pgs.):** Int J Cancer 81(6):877–80
- **Impact Factor (2005):** 4.700

2. Description of the study

**Objectives:** To see if mammographic screening has been associated with reductions in mortality from breast cancer (abstract)

**Location of study:** NSW, Australia

3. Study Design and Population

**Nature of screening programme & mammography:** BreastScreen NSW, state-based mammography screening service.

**Target age for screening participants:** 50–69 (p880, Discussion)

**Target screening interval:** NR

**Start date and maturity of programme under evaluation:** Pilot programmes in 1989, main programme in 1991. ‘BreastScreen NSW was not yet fully implemented in 1995’

**Participation rate:** ‘By 1995, an estimated 72% of women in their 50s, 67% of women in their 60s and 52% of women in their 40s had had an initial screen’.

**Study design:** Descriptive investigation of temporal trends only.

**Entry criteria:** N/A

**Recruitment strategy:** N/A

**Nature of Comparison Groups:**

- **Group 1 (Reference):** N/A
- **Group 2 (Intervention):** (N= NR)
- **Outcome/s of interest reported:** Breast cancer mortality
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality. No accounting for individual screening history.

At what time points?: Annual

Statistical analyses performed: None, descriptive only.
Paper states ‘temporal trends investigated with Poisson regression’ (p878), although not clear if this is applies to mortality results as these are not reported.

Adjustments performed for inter-group differences: Age-standardisation of mortality, using 1991 Australian population

Investigating effect of extent of screening exposure: None

Other methods of bias reduction (incl. loss to follow-up): None

6. Study strengths and weaknesses

Study assumptions:

Strengths:

Weaknesses:

High level of participation in non-targeted age groups contaminates any comparison in trends between age cohorts.
Interpretation of temporal trends stated by authors are debatable. Selective reporting.
Programme too young to observe changes in mortality.
Weak design. Descriptive only. Not possible to differentiate impact of mammography from changes in treatment etc.

7. Other comments

LENNER & JONSSON, 1997

1. Article Details

Study No.: 1912
Title: Excess mortality from breast cancer in relation to mammography screening in Northern Sweden
Author & year: Lenner & Jonsson, 1997
Citation (Jnl, Vol(Num):Pgs.): J Med Screen. 4:6–9
Impact Factor (2005): 2.483

NB. Same methodology as Jonsson et al, 2000 (Study No. 1399), Jonsson et al, 2001 (Study No. 3687), Jonsson et al, 2003 (Study No. 764) and Jonsson et al, 2003 (Study No. 823). Data extracted for Jonsson et al, 2001 (Study No. 3687).
NB. This study compares ‘excess mortality’ between screened and unscreened cohorts, whereas other report ‘refined’ breast cancer mortality.
### 1. Article Details

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>1349</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Predicted long-term mortality reduction associated with the second round of breast screening in East Anglia.</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>McCann et al. 2001</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.):</td>
<td>Br J Cancer 84(3):423–8</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>4.115</td>
</tr>
</tbody>
</table>

### 2. Description of the study

**Objectives:** To estimate the reduction in mortality expected from the second round of screening of East Anglia breast screening programme (abstract)

**Location of study:** East Anglia, UK

### 3. Study Design and Population

**Nature of screening programme & mammography:** Regional mammography screening programme

**Target age for screening participants:**

**Target screening interval:**

Start date and maturity of programme under evaluation: ‘Staggered introduction across the region’ (p423). First centre started screening in 1989, completing first round in 1993 and second round in 1996. Last centre started screening in 1991, completing first round in 1995 and second round in 1998. Women were also invited by year of birth in 5 year age bands over a 3 year cycle starting with the oldest. At time of study, ‘all but one screening centre had completed two rounds of screening’ (p424)

**Participation rate:** NR

**Study design:** Retrospective concurrent cohort. One birth cohort of women eligible for at least two screening rounds therefore provided us with two comparable groups of cancers: one diagnosed in women already invited for screening and the other diagnosed in those awaiting invitation (p424). (NB modelled, not measured, mortality outcome based on prognostic factors)

**Entry criteria:** N/A

**Recruitment strategy:** N/A

### Nature of Comparison Groups:

<table>
<thead>
<tr>
<th>Group 1 (Reference):</th>
<th>(N= 451 cancers) Cancers diagnosed in women awaiting invitation (Uninvited group)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 (Intervention):</td>
<td>(N= 950 cancers) Cancers diagnosed in women already invited (Invited group) NB. Cancers detected at the first screen were excluded.</td>
</tr>
</tbody>
</table>

**Outcome/s of interest reported:** Relative predicted mortality only (not actually measured) NB. Absolute population mortality not actually reported, analyses of relative mortality relies on deaths per cancers detected.
4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | Cancers diagnosed, the time of their diagnosis, the screen or screening interval they related to for that woman (first, second or interval), and their prognostic tumour characteristics. N8. Cancers detected at the first screen were excluded. |
| At what time points?: | Any cancer diagnosed between 1989 and 1996 |
| Statistical analyses performed: | Prediction of mortality was done using two sets of data - data from this entire East Anglia cohort (n=1401) and data from the Swedish Two-County trial. N8. Not clear if current cohort had sufficient follow-up to obtain suitable survival data, otherwise why not use mortality directly? Estimated survival to 88 months. Relative mortality was obtained by dividing the predicted death rate (=deaths/cancers) in the invited group by that in predicted in the in uninvited group. Confidence intervals were estimated assuming a multinomial distribution for the 112 possible prognostic classes (p424). |
| Adjustments performed for inter-group differences: | None, assumes the invited and uninvited are comparable populations. Exception is age at diagnosis which was different between groups, but adjusted for in the prognostic equation. |
| Investigating effect of extent of screening exposure: | Relates to second screen only |
| Other methods of bias reduction (incl. loss to follow-up): | Correction for lead time bias based on data from the Two Counties trial by estimating survival at 124 months (not 88 months) for screen detected cancers |

6. Study strengths and weaknesses

| Study assumptions: | Assumes actual mortality will be same as predicted mortality. Assumes the invited and uninvited groups are comparable populations. Assumes the correction for lead time bias based on Two Counties data also applies here. |
| Strengths: | Excluding the first round of screening excludes the slow growing tumours with good prognosis that might have introduced length bias. |
| Weaknesses: | Women in the invited group were three years older than those in the uninvited group, but age at diagnosis was adjusted for. Relatively short follow-up period. |

7. Other comments

MILTENBERG et al., 1998

1. Article Details

| Study No.: | 3804 |
| Title: | Seventeen-year evaluation of breast cancer screening: The DOM project, The Netherlands. |
| Author & year: | Miltenberg et al, 1998 |
| Citation (Jnl, Vol(Num):Pgs.): | Br J Cancer. 78(7):962–5 |
| Impact Factor (2005): | 4.115 |

Update of case-control study reported by Collette et al, 1984 (Study No. 3532) and Collette et al, 1992 (Study No. 2593)

2. Description of the study

| Objectives: | To demonstrate the benefits of breast cancer screening on mortality |
| Location of study: | Utrecht, The Netherlands |
3. Study Design and Population

Nature of screening programme & mammography: Local screening study (not programme per se) with clinical examination and mammography. Women who did not participate in a particular screening round were not invited to attend the next round, and therefore were eligible as controls.

Target age for screening participants: 50–64 years

Target screening interval: 0, 12, 18, 24 and 48 months.

Start date and maturity of programme under evaluation: 1974, 5 rounds in 10 years. 17 year follow-up reported here.

Participation rate: 72% attended first round

Study design: Matched case-control

Entry criteria: Eligible population is women eligible for screening within the study period.

Recruitment strategy: See below for cases and referents

Nature of Comparison Groups:

Group 1 (Reference): (N= 531)
Controls had to have lived in Utrecht at the time of death of the case, and have the same year of birth as the case. Three controls randomly selected from the screening invitation file

Group 2 (Intervention): (N= 177)
Defined as a breast cancer death in a woman born between 1911 and 1925, whose diagnosis and death occurred between 1975 and 1992.

Outcome/s of interest reported: Screening history taken for the time up to and including the date of diagnosis of the case.

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer deaths, with date of diagnosis of cases. Age and screening history for cases and controls.

At what time points?: 1975–1992

Statistical analyses performed: Relative risk of breast cancer death compared between ever-screened and never-screened, using chi-squared statistic. Also stratified by age or birth cohort.

Adjustments performed for inter-group differences: Age matching

Investigating effect of extent of screening exposure: Additional analysis conducted by number of screening examination (with matching ignored)

Other methods of bias reduction (incl. loss to follow-up): Impact of lead time bias tested by excluding breast cancer deaths with a short follow-up period after diagnosis (i.e., deaths of patients who were less likely to have been screened), because their inclusion would give the impression of a disproportionately large number of deaths from breast cancer in unscreened women. (p962)

6. Study strengths and weaknesses

Study assumptions: Assumes cases and controls are matched for risk factors other than age i.e., from same population without differences other than screening history.

Strengths: Excludes deaths with a diagnosis before the screening programme. Screening history extracted from records so not subject to recall bias.

Weaknesses: At high risk of self-selection bias if non-attendees had been invited but chose not to attend. Small sample so likely to be considerable noise in mortality rates over time.

7. Other comments

Investigates impact of actual screening (i.e., attending screening), rather than availability of a screening programme (i.e., invitation).
MOSS et al, 1999

1. Article Details

Study No.: 3926
Title: 16-year mortality from breast cancer in the UK Trial of Early Detection of Breast Cancer (TEDBC)
Author & year: Moss et al, 1999 (for UK trial of Early Detection of Breast cancer group)
Citation (Jnl, Vol(Num):Pgs.): Lancet: 353:1909–14
Impact Factor (2005): 23.878

Supersedes Ellman et al, 1993 (Study No. 2501)

2. Description of the study

Objectives: To investigate the effect of screening (including mammography) and education on breast cancer mortality.

Location of study: Scotland and England

3. Study Design and Population

Nature of screening programme & mammography: TEDBC was a prospective non-randomised trial (not a service programme per se) set up to investigate impact of screening (by mammography and clinical examination) and education re breast self-examination. Involving eight geographic districts (two screening, two education, and four controls). NB. National programme introduced soon after

Target age for screening participants: 45–64 years
Target screening interval: 2 years mammography (but clinical examination screening every year)
Start date and maturity of programme under evaluation: 1979. Follow-up of 16 years.
Participation rate: 60–70% at first screening
Study design: Prospective concurrent cohort study
Entry criteria: Age 45–64 years within the 7 years of the trial
Recruitment strategy: Involves eight geographic districts (two screening, two education, and four controls)

Nature of Comparison Groups:

Group 1 (Reference): \(N = 127,123 + 63,373\)
Four control centres without screening and two with education regarding breast self-examination

Group 2 (Intervention): \(N = 45,607\)
Two screening centres

Outcome/s of interest reported: Breast cancer mortality

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality for women who were resident in the study areas, and whose cancer was diagnosed after beginning of trial. Deaths for cases diagnosed 7–10 years after date of entry were tested in sensitivity analyses (as potential for impact of national screening programme). Age.

At what time points?: Annual, for 16 years (Data also required for earlier pre-screening period for purpose of adjustment)

Statistical analyses performed: Analysis compares observed number of deaths with expected number of deaths if there were no difference between screened and unscreened areas, after adjustment for age in 5-year age-groups and individual year in trial. The expected numbers were estimated using Poisson regression. Expected numbers were adjusted for pre-trial mortality

Adjustments performed for inter-group differences: Pre-trial mortality used to adjust for differences between groups. Also adjusted for age and period in trial.

Investigating effect of extent of screening exposure: Analyses conducted separately for initial cohort and later-entry cohorts, but no specific consideration of extent of screening.

Other methods of bias reduction (incl. loss to follow-up):
6. Study strengths and weaknesses

Study assumptions:

Strengths: Restricting deaths to those diagnosed within 7 years removed any impact of the national screening programme.

Weaknesses: Restricting deaths to those diagnosed within 7 years increases the possible impact of lead-time bias. Result will be contaminated by the clinical screening occurring in addition to the mammographic screening. Use of pre-trial mortality to adjust for differences between groups may not be adequate, and differences such as socioeconomic status may remain. Doesn’t appear to account for incidental screening in control groups or non-attendance in screened group.

7. Other comments

MOSS et al, 1992

1. Article Details

Study No.: 3857
Title: A case-control evaluation of the effect of breast cancer screening in the United Kingdom trial of early detection of breast cancer.
Author & year: Moss et al, 1992
Citation (Jnl, Vol(Num):Pgs.): J Epidemiol Community Health 46(4):362–4
Impact Factor (2005): 3.003

2. Description of the study

Objectives: Case-control evaluation of the effect of breast cancer screening
Location of study: Guildford and Stoke-on-Trent, UK

3. Study Design and Population

Nature of screening programme & mammography: TEDBC was a prospective non-randomised trial (not a service programme per se) set up to investigate impact of screening (by mammography and clinical examination) and education re breast self-examination. Data for current study comes from Guildford (a screening area) and Stoke-on-Trent (a control area)
Target age for screening participants: 45–64 years
Target screening interval: 2 years mammography (but clinical examination screening every year)
Start date and maturity of programme under evaluation: 1979. Seven years follow-up
Participation rate: In Guildford, 72% accepted initial invitation
Study design: Uses retrospective case-control methodology to evaluate data from a prospective concurrent cohort study (TEDBC) NB. Two different methods reported in paper, using controls from within the screening area only (‘Study B’; data not extracted here) and controls from both areas (Study A, data extracted below). The former is heavily affected by selection bias.
Entry criteria: Age 45–64 years within the 7 years of the trial
Recruitment strategy: Involve two geographic districts (one screening, one control)
### Nature of Comparison Groups:

<table>
<thead>
<tr>
<th>Group 1 (Controls):</th>
<th>(N= 990)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Five controls drawn from pooled population of both areas, matched on age at entry to trial. Controls were free from breast cancer at time of diagnosis of case, and alive at time of death of case.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 2 (Cases):</th>
<th>(N= 198)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All breast cancer deaths in both areas in cases diagnosed after entry to the trial that occurred up to the end 1986.</td>
<td></td>
</tr>
</tbody>
</table>

### Outcome/s of interest reported:
Screening history (never screened, ever screened)

### 4. Bias reduction and statistical methodology

<table>
<thead>
<tr>
<th>Data required in order to conduct primary analyses:</th>
<th>Breast cancer mortality. Date of diagnosis. Age. Screening history.</th>
</tr>
</thead>
<tbody>
<tr>
<td>At what time points?:</td>
<td>Annual</td>
</tr>
<tr>
<td>Statistical analyses performed:</td>
<td>A matched analyses using the Mantel-Hansel method. A logistic regression analysis with conditional likelihood functions was then used to include the effect of screening attendance.</td>
</tr>
<tr>
<td>Adjustments performed for inter-group differences:</td>
<td>Cases and controls obtained from both districts. Mortality was age-standardised.</td>
</tr>
<tr>
<td>Investigating effect of extent of screening exposure:</td>
<td>N/A</td>
</tr>
<tr>
<td>Other methods of bias reduction (incl. loss to follow-up):</td>
<td>Postal survey was used to determine risk factors, but response was poor so this adjustment was not possible. Time in study was measured from actual date of entry to the trial of each case or control, rather than calendar time, with added restriction that controls must have a length of follow-up at least as great as the time from entry to death of the case.</td>
</tr>
</tbody>
</table>

### 6. Study strengths and weaknesses

<table>
<thead>
<tr>
<th>Study assumptions:</th>
<th>Assumes cases and controls come from same underlying population, comparable for all factors other than screening history.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengths:</td>
<td>Screening history extracted from records so not subject to recall bias.</td>
</tr>
<tr>
<td>Weaknesses:</td>
<td>Unable to control for any underlying differences between cases and controls. Cases and controls, and screened and unscreened could be different with respect to risk of breast cancer. Breast awareness likely to be higher in the Guildford area (area with screening programme). Like all case control studies investigates the impact of attending screening, rather than availability of a screening programme. May not adequately control for self-selection bias of screenees.</td>
</tr>
</tbody>
</table>

### 7. Other comments
Investigates impact of actual screening (i.e. attending screening), rather than availability of a screening programme (i.e., invitation).
**OLSEN et al, 2005**

### 1. Article Details

<table>
<thead>
<tr>
<th>Study No.</th>
<th>368</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Breast cancer mortality in Copenhagen after introduction of mammography screening: Cohort study.</td>
</tr>
<tr>
<td>Author &amp; year</td>
<td>Olsen et al, 2005</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num)/Pgs.)</td>
<td>Br Med J: 330(7485):220–2</td>
</tr>
<tr>
<td>Impact Factor (2005)</td>
<td>9.052</td>
</tr>
</tbody>
</table>

### 2. Description of the study

**Objectives:** To evaluate the effect on breast cancer mortality during the first 10 years of mammography service screening programme.

**Location of study:** Copenhagen, Denmark

### 3. Study Design and Population

**Nature of screening programme & mammography:** Local service mammography screening.

**Target age for screening participants:** 50–69 years

**Target screening interval:** 2 years

**Start date and maturity of programme under evaluation:** 1991, 10 years

**Participation rate:** 71% in first round

**Study design:** Retrospective concurrent cohort study (also with historical control).

**Entry criteria:** N/A

**Recruitment strategy:** N/A

**Nature of Comparison Groups:**

<table>
<thead>
<tr>
<th>Group 1 (Reference)</th>
<th>(N= NR, but person-years by age-cohort reported on p222)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Historical control</td>
<td></td>
</tr>
<tr>
<td>2. National control group</td>
<td></td>
</tr>
<tr>
<td>3. Historical national control group</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group 2 (Intervention)</th>
<th>(N= ~40,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All women invited to screening in Copenhagen during first 5 rounds (1991-2001)</td>
<td></td>
</tr>
</tbody>
</table>

**Outcome/s of interest reported:** Breast cancer mortality

### 4. Bias reduction and statistical methodology

**Data required in order to conduct primary analyses:** Annual breast cancer mortality, individual screening data, emigration, age at diagnosis. Personal id numbers used to link patients and to construct the various control groups and their pseudo-invitation rounds.

**At what time points?** Annual, 1991–2001

**Statistical analyses performed:**

- Used Poisson regression model with a study group, historical control group, national control group and a historical national control group to study effect of both invitation to screen and participation in screening.
- To analyse the effect of invitation to screening, we compared breast cancer deaths, adjusting for age (age during follow-up), time period, and region, using a Poisson regression model with the variables five year age group, exposure, period and region.

**Adjustments performed for inter-group differences:** Regression variables therefore controlled for time trends and regional differences, but not the impact that an interaction between the two may have had from the impact of screening.

The authors attempted to correct the result for selection bias (see p221) based on the mortality rate of non-attendees vs. general population.

**Investigating effect of extent of screening exposure:** NR

**Other methods of bias reduction (incl. loss to follow-up):**
6. Study strengths and weaknesses

Study assumptions: Opportunistic screening is minimal.

Strengths: Able to exclude breast cancers diagnosed before the programme started. Studied effect of both invitation to screen and participation in screening.

Weaknesses: Breast cancer mortality was different in Copenhagen and rest of Denmark in the pre-screening period.

7. Other comments

OTTO et al., 2003

1. Article Details

Study No.: 686
Title: Initiation of population-based mammography screening in Dutch municipalities and effect on breast-cancer mortality: a systematic review.
Author & year: Otto et al, 2003
Citation (Jnl, Vol(Num):Pgs.): Lancet: 361(9367):1411–7
Impact Factor (2005): 23.878

2. Description of the study

Objectives: To assess the effect of mammography screening programme on breast cancer mortality rates.
Location of study: The Netherlands

3. Study Design and Population

Nature of screening programme & mammography: Nationwide service screening programme (after pilot programmes in Utrecht and Nijmegen)
Target age for screening participants: 50–69 years
Participation rate: NR
Study design: Descriptive study. Temporal trends only.
Enter criteria: N/A
Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Reference): N/A
Group 2 (Intervention): (N= 8,227,529 person-years)
Person years within the screening period.

Outcome/s of interest reported:
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses:
Mortality and population data at municipality level (these were grouped into clusters). Age.

At what time points?:
1980-1999

Statistical analyses performed:
Mortality rates calculated using breast cancer deaths and mid-year female population by age group.
Assessed time trends in breast cancer mortality before and after introduction of the screening programme by Poisson log-linear regression analysis for three age groups, based on annual percentage change. Relative rates were calculated for each time point since introduction of screening, taking the year of introduction as reference (year 0).

Adjustments performed for inter-group differences:
Mortality rates adjusted to the European standard population.

Investigating effect of extent of screening exposure:
None

Other methods of bias reduction (incl. loss to follow-up):
Analyses large city areas with long first screening round separately. These may have diluted a possible effect of screening. Authors indirectly investigate the impact of changes in treatment and argue that the timing of the reduced mortality due to treatment and due to screening would not have coincided (p1416)

6. Study strengths and weaknesses

Study assumptions:
Assumes temporal changes are due to mammography rather than any other factor (eg. increased awareness, self-examination, treatment changes)

Strengths:
Large study size. Excluded areas were pilot projects were undertaken. At least it standardises the time of screening commencement across municipalities using time 0.

Weaknesses:
Descriptive study. Weak design difficult to differentiate screening effects from treatment changes over time. Impact of treatment could only be indirectly assessed. Unable to exclude deaths from cancers diagnosed before the initiation of screening.

7. Other comments

PACI et al, 2002

1. Article Details

Study No.:
1014

Title:
Quantification of the effect of mammographic screening on fatal breast cancers: The Florence Programme 1990-96

Author & year:
Paci et al, 2002

Citation (Jnl, Vol(Num):Pgs.):
Br J Cancer: 87(1):65–9

Impact Factor (2005):
4.115

2. Description of the study

Objectives:
To quantify the effect of mammographic screening on fatal breast cancers

Location of study:
Florence city, Italy
3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th>Local population based mammography screening. Two view.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants:</td>
<td>50–69 years</td>
</tr>
<tr>
<td>Target screening interval:</td>
<td>2 years</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
<td>1990, ~8 years follow-up</td>
</tr>
<tr>
<td>Participation rate:</td>
<td>NR</td>
</tr>
<tr>
<td>Study design:</td>
<td>Predominantly retrospective before and after study (historical cohort). However also able to compare screened and unscreened; and invited vs. not-invited within the same time period (concurrent cohorts).</td>
</tr>
<tr>
<td>Entry criteria:</td>
<td>N/A</td>
</tr>
<tr>
<td>Recruitment strategy:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Nature of Comparison Groups:**

- **Group 1 (Reference):** (N= 61,786 women)
  Before screening programme (1985-86)
- **Group 2 (Intervention):** (N= 59,947 women)
  After introduction of screening programme (1990-96)

**Outcome/s of interest reported:** Breast cancer mortality (incidence cases only)

4. Bias reduction and statistical methodology

**Data required in order to conduct primary analyses:**

Breast cancer deaths. All were linked to screening database and method and timing of diagnosis. Therefore able to determine whether deaths was in an invitee vs. non-invitee, and attendee vs. non-attendee

**At what time points?:**

Deaths diagnosed 1990-96, with deaths followed up to 1999

**Statistical analyses performed:**

Cases are able to be expressed over correct population denominators to calculate incidence breast cancer mortality i.e., population at diagnosis and age at diagnosis (rather than death) (p66). Mortality is compared to a reference period (pre-screening). By comparing change from reference in the invited and not invited, the authors estimate the impact of ‘other’ factors, such as treatment. Also compares incidence-based mortality in screened and unscreened; and invited vs. not-invited within the same time period (1990-96) as rate ratios. Adjustments performed for inter-group differences:

**Investigating effect of extent of screening exposure:**

Cases were able to be differentiated between round of screening that they had been detected at (first vs. repeat vs. interval)

**Other methods of bias reduction (incl. loss to follow-up):**

6. Study strengths and weaknesses

**Study assumptions:**

**Strengths:**
Breast cancer deaths had to have been diagnosed after the commencement of screening. Able to accurately calculate the population and person-years of follow-up for use as denominator of population mortality. Uses person-years to take into account duration of follow-up.

**Weaknesses:**
Follow-up may be too short to fully detect impact upon mortality. Does not account for changes in treatment over time (other than descriptively). Descriptive changes over time only. Comparison of screened vs. not-screened is subject to self-selection bias.

7. Other comments
### PACI et al, 2002

1. **Article Details**

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>3750</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Assessment of the early impact of the population-based breast cancer screening programme in Florence (Italy) using mortality and surrogate measures.</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>PACI et al, 2002</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.):</td>
<td>Eur J Cancer: 38(4):568–73</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>3.706</td>
</tr>
</tbody>
</table>

Replicates data and methodology presented in PACI et al, 2002 (Study No.1014), therefore data not extracted.

### PALLI et al, 1986

1. **Article Details**

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>3500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>A case-control study of the efficacy of a non-randomized breast cancer screening program in Florence (Italy).</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>PALLI et al, 1986</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.):</td>
<td>Int J Cancer: 38(4):501–4</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>4.700</td>
</tr>
</tbody>
</table>

2. **Description of the study**

- **Objectives:** A case-control study to evaluate the efficacy of a non-randomised breast cancer screening programme in Florence.

3. **Study Design and Population**

- **Location of study:** Rural municipalities in Florence province.
- **Nature of screening programme & mammography:** Mammography screening programme in 19 rural municipalities by invitation. Two view. Physical examination and other tests ‘performed in selected cases’.
- **Target age for screening participants:** 40–70 years.
- **Target screening interval:** ‘Average interval between screening rounds 2.5 years.’
- **Start date and maturity of programme under evaluation:** 1970, 14 years follow-up.
- **Participation rate:** 60% on first invitation.
- **Study design:** Matched case-control.
- **Entry criteria:** N/A.
- **Recruitment strategy:** N/A.

4. **Nature of Comparison Groups:**

- **Group 1 (Controls):** (N= 285) Five living women matched for year of birth and residence, randomly selected from municipality lists at the beginning of 1985.
- **Group 2 (Cases):** (N= 57) All female residents in screening area who died from breast cancer 1977-1984, who were diagnosed after their first screening invitation.

**Outcome/s of interest reported:** Screening history.
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer deaths after initiation of programme. Age. Screening records enabled screening history (including timing) to be determined for both cases and controls.

At what time points?: Annually, 1970–84

Statistical analyses performed: Mantel-Haenszel method for matched data. All variables were simultaneously analysed by multivariate analysis, conditional upon matched data.

Adjustments performed for inter-group differences: Information was included in the model relating to current occupation, level of education, marital status and place of birth.

Investigating effect of extent of screening exposure: None, analysis just ever screened vs. never screened (more detail is considered in Palli et al., 1989)

Other methods of bias reduction (incl. loss to follow-up):

6. Study strengths and weaknesses

Study assumptions: Assumes cases and controls come from same underlying population, comparable for all factors other than screening history.

Strengths: Breast cancer deaths limited to those after commencement of programme. Screening records enabled screening history (including timing) to be determined for both cases and controls.

No recall bias, as all data obtained from records.

Weaknesses: Despite attempts to correct for demographic factors, likely that selection bias still present. Like all case control studies investigates the impact of attending screening, rather than availability of a screening programme. Small number of deaths.

7. Other comments

Significant impact of several demographic factors upon mortality, emphasises the need to collect and include this information in such analyses.

Investigates impact of actual screening (i.e., attending screening), rather than availability of a screening programme (i.e., invitation).

PALLI et al., 1989

1. Article Details

Study No.: 3888

Title: Time interval since last test in a breast cancer screening programme: A case-control study in Italy.

Author & year: Palli et al., 1989

Citation (Jnl, Vol(Num):Pgs.): J Epidemiol Community Health. 43(3):241–8

Impact Factor (2005): 3.003

Update of Palli et al., 1986 (Study No. 3500), with more detail regarding extent of screening exposure and interval since last screen

2. Description of the study

Objectives: To evaluate a population based screening programme for breast cancer, with particular attention to interval since last screen.

Location of study: 23 small towns, Florence Province, Italy
3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th>Mammography screening programme in 23 rural municipalities by invitation. Two view. Physical examination and other tests performed in selected cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants:</td>
<td>40–70 years</td>
</tr>
<tr>
<td>Target screening interval:</td>
<td>Average interval between screening rounds 30 months</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
<td>1970, 17 years follow-up</td>
</tr>
<tr>
<td>Participation rate:</td>
<td>60% on first invitation</td>
</tr>
<tr>
<td>Study design:</td>
<td>Matched case-control</td>
</tr>
<tr>
<td>Entry criteria:</td>
<td>N/A</td>
</tr>
<tr>
<td>Recruitment strategy:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Nature of Comparison Groups:**

| Group 1 (Controls): | (N= 515) Five living women matched for year of birth and town of residence. |
| Group 2 (Cases): | (N= 103) All female residents in screening area who died from breast cancer 1977-1987, who were diagnosed after their first screening invitation. Diagnosis also had to be within three years of the last invitation. |

Outcome/s of interest reported: Screening history (and recency)

4. Bias reduction and statistical methodology

<table>
<thead>
<tr>
<th>Data required in order to conduct primary analyses:</th>
<th>Breast cancer deaths after initiation of programme. Age. Screening records enabled screening history (including timing and recency) to be determined for both cases and controls.</th>
</tr>
</thead>
<tbody>
<tr>
<td>At what time points?:</td>
<td></td>
</tr>
<tr>
<td>Statistical analyses performed:</td>
<td>The analyses considered separately the number of screens and the time since last screen, and also considered these together. The analysis was a conditional logistic model using PHGLM, with crude and adjusted odds ratios reported.</td>
</tr>
<tr>
<td>Adjustments performed for inter-group differences:</td>
<td>Demographic and risk factors, extent of screening.</td>
</tr>
<tr>
<td>Investigating effect of extent of screening exposure:</td>
<td>Adjusted for volume and recency of screening, and a combination of the two.</td>
</tr>
<tr>
<td>Other methods of bias reduction (incl. loss to follow-up):</td>
<td>In addition to information available from records, postal questionnaires were used to determine comparability of possible confounding factors between cases and controls (relative to complete questionnaire on behalf of cases).</td>
</tr>
</tbody>
</table>

6. Study strengths and weaknesses

<table>
<thead>
<tr>
<th>Study assumptions:</th>
<th>Assumes cases and controls come from same underlying population, comparable for all factors other than screening history.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengths:</td>
<td>Breast cancer deaths limited to those after commencement of programme. Screening records enabled screening history (including timing and recency) to be determined for both cases and controls. Able to investigate impact of volume and recency of screening. No recall bias as all data obtained from records.</td>
</tr>
<tr>
<td>Weaknesses:</td>
<td>Despite attempts to correct for demographic factors, can not discount possibility of selection bias. Like all case control studies investigates the impact of attending screening, rather than availability of a screening programme (i.e., invitation).</td>
</tr>
</tbody>
</table>

Relatively small number of deaths.

7. Other comments

Impact of number of screens emphasises the need to collect and include this information in such an analyses where possible. Investigates impact of actual screening (i.e., attending screening), rather than availability of a screening programme (i.e., invitation).
1. Article Details

<table>
<thead>
<tr>
<th>Study No.</th>
<th>23</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Service screening mammography reduces breast cancer mortality among elderly women in Turku.</td>
</tr>
<tr>
<td>Author &amp; year</td>
<td>Parvinen et al, 2006</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num),Pgs.)</td>
<td>J Med Screen: 13(1):34–40</td>
</tr>
<tr>
<td>Impact Factor (2005)</td>
<td>2.483</td>
</tr>
</tbody>
</table>

2. Description of the study

Objectives: To assess the effects of service screening mammography of breast cancer refined mortality

Location of study: Turku city, Finland (compared with Helsinki and Tampere)

3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography</th>
<th>Population based service screening.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants</td>
<td>55–69 years in Turku and 55–59 years in Tampere</td>
</tr>
<tr>
<td>Target screening interval</td>
<td>2 years</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation</td>
<td>1987 in Turku and Tampere, screening period 1987–97, plus 4 years further follow-up</td>
</tr>
<tr>
<td>Participation rate</td>
<td>NR</td>
</tr>
<tr>
<td>Study design</td>
<td>Retrospective concurrent cohort study (also with historical reference period)</td>
</tr>
<tr>
<td>Entry criteria</td>
<td>N/A</td>
</tr>
<tr>
<td>Recruitment strategy</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Nature of Comparison Groups:

<table>
<thead>
<tr>
<th>Group 1 (Reference)</th>
<th>(N= 309,302 person-years, 1987–97) Helsinki, no screening of the birth cohorts under investigation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 (Intervention)</td>
<td>(N= 147,185 person-years, 1987-97) Tampere, screening</td>
</tr>
<tr>
<td>Group 3 (Intervention)</td>
<td>(N= 143,848 person-years, 1987-97) Turku, screening</td>
</tr>
</tbody>
</table>

Outcome/s of interest reported: Breast cancer mortality (refined mortality = deaths from ‘new’ breast cancers)

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer deaths for screening period and pre-screening period in each city. Time of diagnosis of cancer to determine ‘refined’ mortality. Age.

At what time points?: Annual for 15 years.

Statistical analyses performed: Relative risks were computed using Poisson regression. To evaluate developments in background risk, routine trends in the general population before screening were used. A log-linear assumption for a temporal trend was assumed and the trend was calculated per 10 years of time and adjusted for five-year age groups.

Adjustments performed for inter-group differences: Ratios of relative risk takes into account differences between pre-screening and screening that might be common to all cities (eg. changes in treatment over time).

Investigating effect of extent of screening exposure: None, no individual screening information or screening attendance captured.

Other methods of bias reduction (incl. loss to follow-up):
6. Study strengths and weaknesses

<table>
<thead>
<tr>
<th>Study assumptions:</th>
<th>Assumes changes in mortality between pre-screening and post-screening, due to factors other than screening, are constant between cities.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengths:</td>
<td>Comparable duration of follow-up between pre-screening and screening periods, and between cities. Uses reference period to correct for other differences that may be present between cities.</td>
</tr>
<tr>
<td>Weaknesses:</td>
<td>Individual patient level data not available for screening. Not clear to what extent non-attending women within the cities with screening, or opportunistically screened women within the pre-screening period or unscreened city, may dilute the effect.</td>
</tr>
</tbody>
</table>

7. Other comments

PEER et al, 1995

1. Article Details

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>2240</th>
<th>Study No.:</th>
<th>3829</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Effect on breast cancer mortality of biennial mammographic screening of women under age 50.</td>
<td>Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening.</td>
<td></td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>Peer et al, 1995</td>
<td>Quinn et al, 1995</td>
<td></td>
</tr>
</tbody>
</table>

Same methodology as Broeders et al, 2001 (Study No. 1307), which investigated women over 50 years, therefore data not extracted here.

QUINN et al, 1995

1. Article Details

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>3829</th>
<th>Study No.:</th>
<th>749</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Changes in incidence of and mortality from breast cancer in England and Wales since introduction of screening.</td>
<td>Evaluation of the UK breast screening programmes.</td>
<td></td>
</tr>
</tbody>
</table>

Is an alternative analyses of the data presented in Blanks 2000. Detail of the analysis is extracted here.

2. Description of the study

Objectives: To fit an alternative model to that used by Blanks et al, 2000

Location of study: England and Wales, UK
### 3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th>Nationwide mammography screening programme by invitation. Two view at first screen, one view thereafter.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants:</td>
<td>50–64 years</td>
</tr>
<tr>
<td>Target screening interval:</td>
<td>3 years</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
<td>1990 (NB. Blanks reports 1988), ~10 years follow-up.</td>
</tr>
<tr>
<td>Participation rate:</td>
<td>Coverage ~75% nationally. 73% for first round.</td>
</tr>
<tr>
<td>Study design:</td>
<td>Single cohort. Descriptive.</td>
</tr>
<tr>
<td>Entry criteria:</td>
<td>N/A</td>
</tr>
<tr>
<td>Recruitment strategy:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Nature of Comparison Groups:**

| Group 1 (Reference): | N/A |
| Group 2 (Intervention): | (N= 299,000 first screen + 884,000 subsequent screens) |
| Outcome/s of interest reported: | Relative mortality (not reported whether breast cancer or all mortality, presumably breast cancer) |

### 4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | [Breast cancer] mortality by single years of age. |
| At what time points?: | Annual |
| Statistical analyses performed: | The authors fit an alternative model to Blanks et al, 2000, specifically: 
\[ \text{rate} = \exp(f(\text{age}) + g(\text{year of birth}) + h(\text{current year}) + k(\text{years since first covered by screening programme})), \] 
where f, g, h, and k are all smooth functions to be estimated from the data using Poisson regression. (p1207) |
| Adjustments performed for inter-group differences: | Investigating effect of extent of screening exposure: |

### 6. Study strengths and weaknesses

| Study assumptions: | Value dependent upon validity of model assumptions. |
| Weaknesses: | Unable to isolate impact of mammography from other factors such as contemporaneous changes in treatment. As was the case with Blanks et al, 2000, deaths were not restricted to those with cancer diagnosed after the introduction of programme. |

### 7. Other comments

Minimal detail provided.
1. Article Details

Study No.: 2696
Author & year: Sigurdsson et al., 1991
Citation (Jnl, Vol(Num):Pgs.): Int J Cancer 48: 523–28
Impact Factor (2005): 4.700

2. Description of the study

Objectives: Evaluate possible effects of screening upon time trends in breast cancer mortality
Location of study: Iceland

3. Study Design and Population

Nature of screening programme & mammography: Nationwide mammography screening programme.
Target age for screening participants: 40–69 years
Target screening interval: 2 years
Start date and maturity of programme under evaluation: 1987 (although breast examination and referred mammography were used as screening since 1973)
Participation rate: ~65% in 1988 and 1989 (p525)
Study design: Descriptive only, temporal trends (only indirectly related to mammography)
Entry criteria: N/A
Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Reference): (N= NR) 35 year old women (who have breast examination screening)
Group 2 (Intervention): (N= 34,311 women in 1989) Women in Iceland aged 40–69 years

Outcome/s of interest reported: Breast cancer mortality

4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality. Extent of participation in mammography (although only indirectly related to mortality)
At what time points?: Annual, 1955-1989
Statistical analyses performed: In order to evaluate time trends in mortality rate, a curvilinear regression line was fitted to the rates for the individual years.
Adjustments performed for inter-group differences: Age-standardised to 'Segi’s world population'
Investigating effect of extent of screening exposure: None
Other methods of bias reduction (incl. loss to follow-up): None
6. Study strengths and weaknesses

Study assumptions:

Strengths:

Weaknesses:
- Inadequate follow-up (3 years) from commencement of mammography screening and end of study.
- Design cannot account for temporal changes not due to mammography screening (e.g., treatment changes)
- Does not account for attendance at screening.
- Screening with breast examination prior to introduction of mammography will confound the investigation of the impact of mammography

7. Other comments

SMITH et al, 1998

1. Article Details

Study No.: 1838
Author & year: Smith et al, 1998
Citation (Jnl, Vol(Num):Pgs.): Med J Aust.: 168(1):11–4
Impact Factor (2005): 2.127

2. Description of the study

Objectives: To analyse breast cancer mortality trends in Australia and to see if mammographic screening has led to a reduction in mortality (abstract)

Location of study: Australia

3. Study Design and Population

Target age for screening participants: 50–69 years
Target screening interval: NR
Start date and maturity of programme under evaluation: From 1988 onward.
Participation rate: reaches ~50–70% (depending on age group within target range) by 1994 (all <35% in 1990)
Study design: Descriptive study, temporal trends only. Only indirectly related to mammography.

Entry criteria: N/A
Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Reference): N/A
Group 2 (Intervention): (N= NR)
Outcome/s of interest reported: Breast cancer mortality
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer mortality, by age group

At what time points?: 1921–1994

Statistical analyses performed: None, descriptive only.

Adjustments performed for inter-group differences: Age-standardised to world population.

Investigating effect of extent of screening exposure: None. NB. Number screened based on precarious assumptions.

Other methods of bias reduction (incl. loss to follow-up):

6. Study strengths and weaknesses

Study assumptions:

Strengths: High level of participation in non-targeted age groups contaminates any comparison in trends between age cohorts.

Weaknesses: Interpretation of temporal trends stated by authors is subjective. Programme too young to observe changes in mortality. Weak design. Descriptive only. Not possible to differentiate impact of mammography from changes in treatment etc.

7. Other comments

TABAR et al., 2001

1. Article Details

Study No.: 1320

Title: Beyond randomized controlled trials: Organized mammographic screening substantially reduces breast carcinoma mortality.

Author & year: Tabar et al, 2001

Citation (Jnl, Vol(Num):Pgs.): Cancer: 91(9):1724–31

Impact Factor (2005): 4.800

Superseded by Tabar et al, 2003 therefore data not extracted here.

TABAR et al, 2003A

1. Article Details

Study No.: 687

Title: Mammography service screening and mortality in breast cancer patients: 20-year follow-up before and after introduction of screening.

Author & year: Tabar et al, 2003

Citation (Jnl, Vol(Num):Pgs.): Lancet: 361(9367):1405–10

Impact Factor (2005): 23.878

Supersedes Tabar et al, 2001 (Study No. 1320)
### 2. Description of the study

**Objectives:** To assess the long-term effect of mammographic service screening on death from breast cancer, taking into account potential biases from self-selection, changes in breast cancer incidence, and classification of cause of death.

**Location of study:** Östergötland and Dalarna counties, Sweden

### 3. Study Design and Population

**Nature of screening programme & mammography:** Regional mammography programmes. NB. Between 1978 and 1985, ~50% of women were screened as part of an RCT (and other 50% were controls). After 1985, all were invited.

**Target age for screening participants:** 40–69 years

**Target screening interval:** ‘18 months in women < 55 years and 24 months in women ≥ 55 years’ (p1409).

**Start date and maturity of programme under evaluation:** 1977, 20 years

**NB. Between 1978 and 1985, ~50% of women were screened as part of an RCT (and other 50% were controls). After 1985, all were invited.**

**Participation rate:** 85% or higher (p1409)

**Study design:** Before and after study (retrospective historical cohort). Difference between attendees and non-attendees within the screening period also investigated (retrospective concurrent cohort).

**Entry criteria:** N/A

**Recruitment strategy:** N/A

**Nature of Comparison Groups:**

<table>
<thead>
<tr>
<th>Group 1 (Reference)</th>
<th>Group 2 (Intervention)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(N= 120,778 women aged 40–69 years, on average)</td>
<td>(N= 119,938 women aged 40–69 years, on average)</td>
</tr>
<tr>
<td>Pre-screening period, 1958–77, but only including breast cancer diagnosed within this period</td>
<td>Screening period, 1978–97, but only including breast cancer diagnosed within this period</td>
</tr>
</tbody>
</table>

**Outcome/s of interest reported:** Breast cancer mortality and all-cause mortality.

### 4. Bias reduction and statistical methodology

**Data required in order to conduct primary analyses:** Breast cancer deaths, time of diagnosis and age at diagnosis. Individual screening status also required for analyses by attendance.

**At what time points?:** Annual for 40 years

**Statistical analyses performed:** In first instance calculated raw unadjusted relative risks of breast cancer death between the two periods. Then estimated relative risks adjusted for age by Poisson regression, for changes in incidence of breast cancer relative to that in the earlier period, and when relevant, for bias from self-selection for attendance at screening (p1406). Results stratified by age group at diagnosis. (NB. Also looked at all-cause mortality)

**Adjustments performed for inter-group differences:** Adjusts for age and breast cancer incidence. Change in incidence adjustment used data from unscreened women in both periods to avoid artificial increases in incidence as a result of screening. Adjusts for self-selection bias assuming that non-attendees have a 36% higher risk of dying than those in general population (based on previous data).

**Investigating effect of extent of screening exposure:** Able to differentiate between attendees and non-attendees, but no analysis by number of screens.

**Other methods of bias reduction (incl. loss to follow-up):** Uses a weighted average of the reduction in mortality in the non-attending group and a younger age group to estimate the impact from factors other than screening. Subtracts this estimate from the calculation mortality reduction.
6. Study strengths and weaknesses

Study assumptions:

Strengths: Diagnosis and death had to occur within the periods of interest. Able to identify actual screening attendance amongst the individual women in the screened period.

Weaknesses: Uses a weighted average of the reduction in mortality in the non-attending group and a younger age group to estimate the impact from factors other than screening (treatment changes, management changes, increased breast awareness and self-examination). This may or may not be an accurate estimate.

NB. Results of this study may not be entirely generalisable to service screening as RCT for first 8 years.

Adjustments for changes in incidence by age at diagnosis may have introduced bias, as this could have been affected by screening.

7. Other comments

TABAR et al, 2003b

1. Article Details

Study No.: 824
Title: Mammography screening: A key factor in the control of breast cancer.
Author & year: Tabar et al, 2003b
Citation (Jnl, Vol(Num):Pgs.): Cancer J: 9:15–27
Impact Factor (2005): 4.800

Largely replicates the data of Tabar et al, 2003a (Study No. 687) and Paci et al, 2002 (Study No. 1014)

TAYLOR et al, 2004

1. Article Details

Study No.: 487
Title: Mammography screening and breast cancer mortality in New South Wales, Australia.
Author & year: Taylor et al, 2004
Citation (Jnl, Vol(Num):Pgs.): Cancer Causes Control: 15(6):543–50
Impact Factor (2005): 3.195

2. Description of the study

Objectives: To investigate the relationship between the utilisation of service mammography screening and breast cancer mortality in NSW women.

Location of study: 174 municipalities in NSW, Australia
3. Study Design and Population

<table>
<thead>
<tr>
<th>Nature of screening programme &amp; mammography:</th>
<th>BreastScreen NSW, state-based mammography screening service. Two view mammography read by two radiologists.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Target age for screening participants:</td>
<td>50–69 years</td>
</tr>
<tr>
<td>Target screening interval:</td>
<td>2 years</td>
</tr>
<tr>
<td>Start date and maturity of programme under evaluation:</td>
<td>Progressively introduced from 1988, reached full geographic coverage by 1996 (abstract)</td>
</tr>
<tr>
<td>Participation rate:</td>
<td>See Fig 1, exceeded 50% in ~1996</td>
</tr>
<tr>
<td>Study design:</td>
<td>Descriptive investigation of temporal trends only for whole population.</td>
</tr>
<tr>
<td></td>
<td>Focuses upon retrospective analysis relating mortality to intensity of screening on a municipality level (after allowing for time lag)</td>
</tr>
<tr>
<td>Entry criteria:</td>
<td>N/A</td>
</tr>
<tr>
<td>Recruitment strategy:</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Nature of Comparison Groups:

<table>
<thead>
<tr>
<th>Group 1 (Reference):</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2 (Intervention):</td>
<td>(N= 554,000 women in NSW aged 50–69 years)</td>
</tr>
<tr>
<td>Outcome/s of interest reported:</td>
<td>Breast cancer mortality relative to screening participation</td>
</tr>
</tbody>
</table>

4. Bias reduction and statistical methodology

| Data required in order to conduct primary analyses: | Breast cancer mortality by municipality. Participation by municipality. Age. Demographic indicators by municipality. |
| At what time points?: | Annual |
| Statistical analyses performed: | Breast cancer mortality for 1997–2001 was examined in relation to lagged participation rates by municipality, adjusted for age, socioeconomic and geographic indicators, and breast cancer incidence. Detail of Poisson regression model is presented on p545. |
| Adjustments performed for inter-group differences: | Other confounders such as socioeconomic status, geography of municipalities and incident rate of breast cancer were included in final model. |
| Investigating effect of extent of screening exposure: | Relates mortality to participation on a municipality level (but not extent of individual screening i.e., volume) |
| Other methods of bias reduction (incl. loss to follow-up): | |

6. Study strengths and weaknesses

| Study assumptions: | Assumes a single women mortality statistic and screening participation will be attributed to the same municipality. Assumes that the reduction in mortality is due to mammography rather than other factors. |
|Strengths: | |
|Weaknesses: | Association between breast cancer mortality and screening participation may be influenced by other factors that may follow a similar pattern across municipalities, but be unrelated to mammography participation (eg. increased breast awareness and self-examination, accessing of treatments etc). May not be sufficient time lapsed since programme maturity to detect impact upon mortality. Includes deaths diagnosed before the screening period. Excludes mammography screening performed elsewhere, or via Medicare (when referred by doctor). Potential for the extent of adjustment for breast cancer incidence between municipalities to have been affected by screening. |

7. Other comments
### Study No.: 2290

**Title:** A population-based case-cohort evaluation of the efficacy of mammographic screening for breast cancer.

**Author & year:** Thompson *et al.*, 1994

**Citation (Jnl, Vol(Num):Pgs.):** Am J Epidemiol.: 140(10):889–901

**Impact Factor (2005):** 5.068

Also included as a methodological publication

### 2. Description of the study

**Objectives:** Illustrates how case-control methodology can be used to perform efficient assessment of screening efficacy in large cohorts, while eliminating or controlling for bias.

**Location of study:** Puget Sound, WA, USA

### 3. Study Design and Population

**Nature of screening programme & mammography:** Managed care organisation mammography screening programme, by invitation

**Target age for screening participants:** ≥ 40 years

**Target screening interval:** Interval is variable, based upon response to risk questionnaire and age. (83% of women ≥ 40 years are invited once in a 3-year period).

**Start date and maturity of programme under evaluation:** 1985, 3 years

**Participation rate:** NR

**Study design:** Case-cohort (i.e., cases are not linked to specific controls, but rather the entire cohort at risk). The entire cohort is represented by a 'sub-cohort', that has been randomly selected from the cohort (before identification of the cases - therefore a case is also a member of the cohort, and may be a member of the sub-cohort. But covariate information only needs to be collected in the sub-cohort, rather than the whole cohort (decreases costs and efforts) (p892)

**Entry criteria:** N/A

**Recruitment strategy:** N/A

**Nature of Comparison Groups:**

**Group 1 (Cohort):** (N= 2,237)

‘Sub-cohort’ from same managed care organisation. NB, can include cases.

**Group 2 (Cases):** (N= 126)

Breast cancer cases diagnosed between 1982 and 1988, with death occurring up to end 1988, who had been diagnosed during the period of belonging to the managed care organisation.

**Outcome/s of interest reported:** Screening history (ever screened vs. never screened)

### 4. Bias reduction and statistical methodology

**Data required in order to conduct primary analyses:** Breast cancer deaths and date of diagnosis. Screening history and risk factor data for cases and ‘sub-cohort’, extracted from medical records. Age.

**At what time points?:** Annual, 1982–88.

**Statistical analyses performed:** An age-stratified proportional hazards analysis compared the screening exposure of the cases with sub-cohort. The analysis mimics a Cox regression analysis, except as described in detail in the publication (p894). Complex analyses (see p 894 for detail).

**Adjustments performed for inter-group differences:** Family history of breast cancer, previous biopsy and parity were included in the model to control for potential effects of selection bias.
<table>
<thead>
<tr>
<th>Investigating effect of extent of screening exposure:</th>
<th>None (just never vs. ever screened)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other methods of bias reduction (incl. loss to follow-up):</td>
<td>Data on risk factors and potential confounders was extracted from medical records.</td>
</tr>
</tbody>
</table>

### 6. Study strengths and weaknesses

<table>
<thead>
<tr>
<th>Study assumptions:</th>
<th>Assumes ‘sub-cohort’ is representative of the whole cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengths:</td>
<td>Breast cancers diagnosed before start of programme excluded. Collects risk factor information from medical records for cases and sub-cohort to enable adjustment. Screening history extracted from records so not subject to recall bias.</td>
</tr>
<tr>
<td>Weaknesses:</td>
<td>Programme not sufficiently mature to detect impact of mammography upon mortality (entire study period 1982–1988). Study may be limited by women leaving and joining the managed care organisation. Relatively small number of cases. Sub-cohort may not be representative of the entire cohort with respect to risk factors (2237/94656 women) - no way of knowing this.</td>
</tr>
</tbody>
</table>

### 7. Other comments

Investigates impact of actual screening (i.e., attending screening), rather than availability of a screening programme (i.e., invitation).

---

**THRELFALL et al., 2003**

### 1. Article Details

<table>
<thead>
<tr>
<th>Study No.:</th>
<th>3946</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title:</td>
<td>Impact of NHS breast screening on advanced disease and mortality from breast cancer in North West of England.</td>
</tr>
<tr>
<td>Author &amp; year:</td>
<td>Threlfall et al, 2003</td>
</tr>
<tr>
<td>Citation (Jnl, Vol(Num):Pgs.):</td>
<td>Br J Cancer. 89: 77–80</td>
</tr>
<tr>
<td>Impact Factor (2005):</td>
<td>4.115</td>
</tr>
</tbody>
</table>

### 2. Description of the study

| Objectives: | To describe outcomes from breast cancer in women who were aged 54 or younger when they were first invited for NHS breast screening. |
| Location of study: | Wigan and Manchester, England |

### 3. Study Design and Population

| Nature of screening programme & mammography: | Breast screening programme by invitation in Wigan and Manchester in North West England. Publication does not state what the breast screen programme involved (i.e., does not state mammography), but part of NHS programme which is known to include mammography. |
| Target age for screening participants: | First this study: <54 years, when first invited to screening (although not reported, assumed to be 50–54 years) |
| Target screening interval: | NR, assumed to be 3 years |
| Start date and maturity of programme under evaluation: | 1989, ~11 years, ~4 screening rounds |
| Participation rate: | Manchester: 84% attended once, 51% attended all screening invitations. Wigan: 92% attended once, 79% attended all screening invitations. |
| Study design: | Parallel cohorts (differing by virtue of screening attendance). |
| Entry criteria: | N/A |
| Recruitment strategy: | N/A |
**Nature of Comparison Groups:**

<table>
<thead>
<tr>
<th>Group 1:</th>
<th>(N=10,750) Manchester women &lt; 54 years when they were first invited to screening between 1 Jan 1989 and 30 Sept, 1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 2:</td>
<td>(N=5125) Wigan women &lt; 54 years when they were first invited to screening between 1 Jan 1989 and 30 June, 1990</td>
</tr>
</tbody>
</table>

**Outcome/s of interest reported:** Breast cancer mortality

**4. Bias reduction and statistical methodology**

<table>
<thead>
<tr>
<th>Data required in order to conduct primary analyses:</th>
<th>Uses record linkage on an individual basis to link screening histories and mortality. Cancer status at time of first screening round is known. A breast cancer death was recorded if recorded in section 1a, 1b or 1c of the death certificate.</th>
</tr>
</thead>
<tbody>
<tr>
<td>At what time points?:</td>
<td>Not reported. Assumed to be 1989–2000</td>
</tr>
<tr>
<td>Statistical analyses performed:</td>
<td>Population mortality analyses conducted with and without exclusion of women with cancer at time of first screening round. Analyses are per person-years of follow-up. Chi-squared test to compared mortality rates between Manchester and Wigan.</td>
</tr>
<tr>
<td>Adjustments performed for inter-group differences:</td>
<td>None. Mean age of women with the two cohorts is not reported. Age-standardisation not reported. Demographic differences between groups not reported, therefore not adjusted for.</td>
</tr>
<tr>
<td>Investigating effect of extent of screening exposure:</td>
<td>Because of linked data, analyses can be conducted by extent of screening i.e., first screen, re-screen, interval, lapsed attender, non-attender and total</td>
</tr>
<tr>
<td>Other methods of bias reduction (incl. loss to follow-up):</td>
<td>Data linkage allows exclusion of cancers diagnosed before the screening programme from the mortality result.</td>
</tr>
</tbody>
</table>

**6. Study strengths and weaknesses**

<table>
<thead>
<tr>
<th>Study assumptions:</th>
<th>Assumes impact of any treatment changes 1989–2000 was equal in both cohorts. Assumes other demographic factors that may influence breast cancer are comparable between groups. Appears to assume that age is equal in both groups, and there is no adjustment or standardisation for this (i.e., theoretically possible that all Manchester patients were 53 years and all Wigan patients were 50, which could explain higher mortality in Manchester.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strengths:</td>
<td>Excludes deaths where the diagnosis occurred before the commencement of the programme. Able to disaggregate results by extent of screening exposure.</td>
</tr>
<tr>
<td>Weaknesses:</td>
<td>Assumes impact of any contemporaneous treatment changes during 1989–2000 was equal in both cohorts. Assumes other demographic factors that may influence breast cancer are comparable between groups. Appears to assume that age is equal in both groups, and there is no adjustment or standardisation for this (i.e., theoretically possible that all Manchester patients were 53 years and all Wigan patients were 50, which could explain higher mortality in Manchester. Overall mortality result relies on difference in attendance rate between Manchester and Wigan to impute effect of mammography, when the difference is modest. Although analyses by extent of screening exposure is informative. But, both have potential to be heavily influenced by selection bias.</td>
</tr>
</tbody>
</table>

**7. Other comments**
1. Article Details

Study No.: 3947
Author & year: Tornberg et al, 2006
Citation (Jnl, Vol(Num):Pgs.): Acta Oncologica 45: 528–535
Impact Factor (2005): 2.362

Majority of data has been reported elsewhere, but not comparatively as reported here.

2. Description of the study

Objectives: To find out if the decrease in breast cancer mortality in relation to the availability of population-based breast cancer screening programmes, already described elsewhere, would also be possible to demonstrate by using routine, easily accessible statistics (p529).

Location of study: 4 Nordic capitals: Helsinki, Stockholm, Copenhagen, Oslo

3. Study Design and Population

Nature of screening programme & mammography: Nationwide programmes (only capital cities of interest here) all with 2 view mammography and double radiologist reading

Target age for screening participants:
- Helsinki: 50–59 years
- Stockholm: 50–69 years
- Copenhagen: 50–69 years
- Oslo: 50–69 years

Target screening interval:
- Helsinki: 2 years
- Stockholm: 2 years
- Copenhagen: 2 years
- Oslo: 2 years

Start date and maturity of programme under evaluation: see Table 1:
- Helsinki: 1986, 12 years at time of this analysis
- Stockholm: 1989, 9 years at time of this analysis
- Copenhagen: 1991, 7 years at time of this analysis
- Oslo: 1996, 2 years at time of this analysis

Participation rate: see Fig 2:
- Helsinki: average 82%
- Stockholm: 70–75%
- Copenhagen: ~60%
- Oslo: ~70%

Study design: Descriptive only. Does not directly compare mammography screening with no screening. Indicates time of mammography programme commencement on graphical presentation of mortality.

Entry criteria: N/A
Recruitment strategy: N/A

Nature of Comparison Groups:

Group 1 (Reference group): N/A, descriptive only
Group 2 (Intervention group): N/A, descriptive only
Outcome/s of interest reported: Breast cancer mortality
4. Bias reduction and statistical methodology

Data required in order to conduct primary analyses: Breast cancer deaths. Age group.

At what time points?: Mortality rate presented annually from 1970 to 1998

Statistical analyses performed: Descriptive only. Does not directly compare mammography screening with no screening. Indicates time of mammography programme commencement on graphical presentation of mortality.

Adjustments performed for inter-group differences: Not clear if mortality is age-standardised.

Investigating effect of extent of screening exposure: None

Other methods of bias reduction (incl. loss to follow-up): None

6. Study strengths and weaknesses

Study assumptions: Descriptive only, but basic premise is linking temporal pattern of breast cancer mortality to availability of mammography. Does not directly compare mammography screening with no screening

Strengths:

Weaknesses: Weak design to address impact of mammography upon mortality, and short follow-up.

7. Other comments

VAN DIJCK et al, 1997

1. Article Details

Study No.: 1973

Title: Breast-cancer mortality in a non-randomized trial on mammographic screening in women over age 65

Author & year: Van Dijck et al, 1997

Citation (Jnl, Vol(Num):Pgs.): Int J Cancer: 70(2):164–8

Impact Factor (2005): 4.700

Same methodology as Broeders et al, 2001 (Study No. 1307), which investigated women over 50 years, therefore data not extracted here.

2. Description of the study

Objectives: To evaluate whether including elderly women in a screening programme affects breast cancer mortality

Location of study: Nijmegen and Arnhem, The Netherlands

VAN DIJCK et al, 1996

1. Article Details

Study No.: 2088

Title: Mammographic screening after the age of 65 years: Evidence for a reduction in breast cancer mortality

Author & year: Van Dijck et al, 1996

Citation (Jnl, Vol(Num):Pgs.): Int J Cancer: 66, 727–731

Impact Factor (2005): 4.700

Superseded by Broeders et al, 2002 (Study No. 1154) with same methodology
<table>
<thead>
<tr>
<th>Study No.</th>
<th>Title</th>
<th>Author &amp; year</th>
<th>Citation</th>
<th>Impact Factor (2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2353</td>
<td>Efficacy of mammographic screening of the elderly: A case referent study in the Nijmegen program in The Netherlands</td>
<td>Van Dijck et al, 1994</td>
<td>JNCI 86 (12): 934–938</td>
<td>15.171</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Superseded by Broeders et al, 2002 (Study No. 1154) with same methodology</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Superseded by Broeders et al, 2002 (Study No. 1154) with same methodology</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Superseded by Broeders et al, 2002 (Study No. 1154) with same methodology</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX A-11: RANKING OF STUDY DESIGNS BY EXPERT TECHNICAL GROUP

<table>
<thead>
<tr>
<th>Member ID</th>
<th>Study no.</th>
<th>Value of Evidence</th>
<th>Ease of implementation</th>
<th>Risk of unsuccessful implementation</th>
<th>Anticipated cost</th>
<th>Time to complete study</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>3.5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4.5</td>
<td>4.5</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>3.5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>3.5</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>3.5</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3.5</td>
<td>4</td>
<td>2.5</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>1.5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>2.5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>2.5</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2.5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>3.5</td>
<td>4</td>
<td>4</td>
<td>2.5</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>4</td>
<td>3.5</td>
<td>4</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>4.5</td>
<td>3.5</td>
<td>3.5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>4.5</td>
<td>4.5</td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>3</td>
<td>4.5</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
APPENDIX B-1: METHODOLOGICAL PUBLICATIONS RELATING TO EVALUATING THE IMPACT OF MAMMOGRAPHY SERVICE SCREENING UPON POPULATION MORTALITY

Ref ID: 89
Ref ID: 432
Ref ID: 1175
Ref ID: 144
Ref ID: 3935
Ref ID: 3709
Ref ID: 2616
Ref ID: 1781
Ref ID: 2181
Ref ID: 3890
Ref ID: 3897
Ref ID: 3913

Ref ID: 1810

Ref ID: 3941

Ref ID: 189

Ref ID: 3922

Ref ID: 3938

Ref ID: 2251

Ref ID: 1016

Ref ID: 2024

Ref ID: 399

Ref ID: 3942

Ref ID: 3932

Ref ID: 2572

Ref ID: 3933
Ref ID: 1663

Ref ID: 177

Ref ID: 859

Ref ID: 3931

Ref ID: 1176

Ref ID: 3940

Ref ID: 3731

Ref ID: 679

Ref ID: 1025

Ref ID: 3621

Ref ID: 3732

Ref ID: 1139

Ref ID: 3944

Ref ID: 3868
Ref ID: 851

Ref ID: 288

Ref ID: 3934

Ref ID: 743

Ref ID: 3929

Ref ID: 2761

Ref ID: 37

Ref ID: 3722

Ref ID: 2073

Ref ID: 1609

Ref ID: 3887

Ref ID: 2261

Ref ID: 3101

Ref ID: 2306
Ref ID: 753

Ref ID: 3936

Ref ID: 581

Ref ID: 2542

Ref ID: 3928

Ref ID: 3930
APPENDIX B-2: PUBLICATIONS MODELLING IMPACT OF MAMMOGRAPHY SERVICE SCREENING UPON POPULATION MORTALITY

Ref ID: 1980

Ref ID: 1885

Ref ID: 1299

Ref ID: 2831

Ref ID: 3836

Ref ID: 514

Ref ID: 3824

Ref ID: 1646

Ref ID: 3889

Ref ID: 1354
APPENDIX B-3: PUBLICATIONS REPORTING CANCER SURVIVAL BUT NOT POPULATION MORTALITY

Ref ID: 111

Ref ID: 1129

Ref ID: 503

Ref ID: 2343

Ref ID: 128

Ref ID: 2578

Ref ID: 1783

Ref ID: 68

Ref ID: 580

Ref ID: 682

Ref ID: 706

Ref ID: 1056

Ref ID: 679
Ref ID: 2305

Ref ID: 1640

Ref ID: 141

Ref ID: 2750

Ref ID: 2875

Ref ID: 3948

Ref ID: 2867

Ref ID: 101

Ref ID: 2344

Ref ID: 1935

Ref ID: 1011

Ref ID: 2034