This copy has been supplied by the Library of the University of Otago on the understanding that the following conditions will be observed:

1. To comply with s56 of the Copyright Act 1994 [NZ], this thesis copy must only be used for the purposes of research or private study.

2. The author's permission must be obtained before any material in the thesis is reproduced, unless such reproduction falls within the fair dealing guidelines of the Copyright Act 1994. Due acknowledgement must be made to the author in any citation.

3. No further copies may be made without the permission of the Librarian of the University of Otago.
EVALUATION OF A PILOT BREAST CANCER SCREENING PROGRAMME

Ann Kathleen Richardson

A thesis submitted for the degree of
Doctor of Philosophy
at the University of Otago, Dunedin,
New Zealand

September 1996
ABSTRACT

A pilot mammographic breast cancer screening programme was established in 1991 for women aged fifty to sixty-four living in Otago and Southland. The pilot programme was evaluated to measure the acceptability, effectiveness, and economic efficiency of breast cancer screening in New Zealand.

A series of targets related to the performance of the screening programme during its first screening round was used in this evaluation. The targets were derived from screening programmes that had been shown to reduce breast cancer mortality in randomised controlled trials. Methods were developed for measuring these targets in the Otago-Southland pilot programme and the evaluation largely consisted of monitoring these targets, together with carrying out surveys of women in the eligible population and their general practitioners. The results from the evaluation were reported to the programme staff at regular intervals and contributed to the design and management of the pilot programme. Information provided by this evaluation also contributed to the New Zealand government's decision in June 1995, to establish a national breast cancer screening programme which will be implemented over three years.

The early results from the pilot programme were encouraging, with over ninety percent of women aged fifty to sixty-four in Otago and Southland having been identified so they could be invited to be screened. Seventy-four percent of the target population were screened in the first screening round. In the first eighteen months of the programme 7,182 women were screened. Satisfaction with the screening programme was very high, with ninety-four percent of screened women planning to continue in the programme. As a result of screening 832 women (11.6 percent of those screened) were referred to the assessment clinic. This referral rate did not meet the target that had been set, and as a result the specificity of the programme was low, at eighty-nine percent. Fortunately, despite the high referral rate, few women underwent surgical biopsies and the benign : malignant ratio was excellent, at 0.9 : 1. Breast cancer was diagnosed in seventy-three women; a detection
rate of 10.2 per thousand women screened. Sensitivity was excellent, at ninety-two percent. The economic analysis showed that breast cancer screening can be carried out in New Zealand at a similar cost to breast cancer screening programmes in other countries.

The results of the evaluation suggest that the pilot programme will reduce breast cancer mortality, among eligible women, but methods to lower the referral rate and improve specificity should be sought. Arising from the evaluation of the pilot programme, several recommendations are made for implementing a national breast cancer screening programme in New Zealand.
PREFACE

This thesis documents the evaluation of the effectiveness and acceptability of the Otago-Southland pilot programme, with a summary and discussion of the results of the evaluation, including the associated economic analysis of the pilot programme. The evaluation of the pilot programme was directed by Mark Elwood (Professor of Cancer Epidemiology, University of Otago) and was coordinated by the author. The evaluation team also included Sheila Williams (Biostatistician), Bronwen McNoe (Research Assistant), Nancy Devlin (Health Economics Consultant) and Arun Menon (Public Health Medicine Registrar). The author developed the targets for the evaluation and methods to measure the targets, and designed the surveys and questionnaires that were used in the evaluation. The author was also responsible for the day to day running of the evaluation until leaving New Zealand in January 1994 to take up a position in England. Where other members of the evaluation team carried out work that has been discussed in the thesis, this has been acknowledged in the body of the text, or the work has been included as an appendix to the text.

Structure of the thesis

Chapter One is a review of relevant literature, with the published studies of mammographic breast cancer screening described in detail. This chapter presents the evidence about breast cancer screening and discusses the strengths and weaknesses of different study designs in assessing the efficacy of breast cancer screening. It also addresses the issue of the efficacy of screening according to age. Chapter Two covers the introduction of population-based breast cancer screening in New Zealand, including the epidemiology of breast cancer in New Zealand, factors leading up to the introduction of population-based screening, and the recommendations that were made about pilot programmes and their evaluation. This chapter also compares mammographic screening with other strategies to reduce breast cancer mortality in New Zealand. The risks and
benefits of breast cancer screening are also addressed. Methods for evaluating breast cancer screening programmes are described in Chapter Three, with reasons given for the choice of the evaluation method that was used for the Otago-Southland pilot programme. Chapter Four gives a detailed description of the evaluation method including the derivation of the targets used. The methods and results of the evaluation of the first eighteen months of the Otago-Southland pilot programme are presented and discussed in Chapters Five and Six. Chapter Seven is a summary and discussion, including an analysis of the strengths and weaknesses of the evaluation method, the successes and failures of the pilot programme in terms of its effectiveness, acceptability, and economic efficiency, and the implications for a national breast cancer screening programme in New Zealand.

Acknowledgements

I am very grateful to my supervisors Mark Elwood and Charlotte Paul, for supervising my work and for reading and making helpful comments on drafts of this thesis. It has been a privilege to have worked with and to have learned from them. I am very grateful to Keith Hunter, Mary Richardson, Jane Chetwynd, and Elisabeth Wells, who also read drafts of this thesis and made many valuable comments and suggestions. I am very grateful to Caroline Lintott and Ruth Helms for their encouragement and advice about PhD theses. Also I wish to thank Sheila Williams, Brian Cox, Peter Herbison, Bronwen McNoe, Margaret Bahr, Trish Smith, Nancy Devlin, Arun Menon, Elizabeth Bang and the staff of the Otago-Southland Breast Cancer Screening Programme and the general practitioners and the women who took part in the surveys. The evaluation of the Otago-Southland pilot programme was funded by the Cancer Society of New Zealand and the Health Research Council of New Zealand.
The following papers, reporting findings from the evaluation of the Otago and Southland breast cancer screening pilot programme have been published or are in press:


TABLE OF CONTENTS

ABSTRACT

PREFACE

Structure of the thesis iv
Acknowledgements v
Author's note vi

CHAPTER ONE

The efficacy of mammographic screening

Introduction 1
1.1 Evaluating the efficacy of breast cancer screening 1
1.2 Randomised controlled trials of breast cancer screening 3
1.3 Non-randomised studies of breast cancer screening 13
1.4 Breast cancer screening and age 17
1.5 Summary 24
CHAPTER TWO
The establishment of population-based breast cancer screening in New Zealand

Introduction 25
2.1 Breast cancer in New Zealand 25
2.2 Comparing strategies to reduce breast cancer mortality in New Zealand 30
2.3 Breast cancer screening: risks and benefits 35
2.4 Population-based screening 47
2.5 The introduction of breast cancer screening in New Zealand 48
2.6 Summary 52

CHAPTER THREE
Deciding on the evaluation method

Introduction 53
3.1 Evaluation 54
3.2 Comparing the possible methods of evaluation 60
3.3 Comparing the randomised controlled trials of breast cancer screening 60
3.4 Summary 77
CHAPTER FOUR

The derivation of targets for the evaluation

Introduction 80
4.1 Measuring the effectiveness of the pilot programme 83
4.2 Measuring the acceptability of the pilot programme 93
4.3 The economic evaluation of the pilot programme 94
4.4 Data collection for the evaluation 95
4.5 Summary 99

CHAPTER FIVE

The effectiveness of the pilot programme

Introduction 102
5.1 Identifying and inviting eligible women 103
5.2 The screening test 114
5.3 Referral for assessment 116
5.4 Breast cancer diagnosis in the pilot programme 121
5.5 Summary 137
CHAPTER SIX

The acceptability of the pilot programme

Introduction
6.1 Pre-screening survey
6.2 Participation in screening
6.3 Acceptability of screening
6.4 Acceptability of assessment
6.5 Survey of women who did not attend after an invitation to screening
6.6 Summary

CHAPTER SEVEN

Summary and Discussion

Introduction
7.1 Method strengths and limitations
7.2 Principal findings; effectiveness
7.3 Principal findings; acceptability
7.4 Principal findings; economic efficiency
7.5 Implications
7.6 Recommendations for the national breast cancer screening programme
7.7 Conclusions

REFERENCES
APPENDICES

APPENDIX ONE


APPENDIX TWO


APPENDIX THREE

Download ASCII file from pilot programme computer system

APPENDIX FOUR

General Practitioner questionnaire

APPENDIX FIVE

Consent form and symptom questionnaire

APPENDIX SIX

Screening questionnaire

APPENDIX SEVEN

Assessment questionnaire

APPENDIX EIGHT

Non-attender questionnaire
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Randomised controlled trials of breast cancer screening</td>
<td>12</td>
</tr>
<tr>
<td>1.2</td>
<td>Case-control studies of breast cancer screening</td>
<td>16</td>
</tr>
<tr>
<td>1.3</td>
<td>Results of randomised controlled trials by age: younger women</td>
<td>19</td>
</tr>
<tr>
<td>1.4</td>
<td>Results of randomised controlled trials by age: older women</td>
<td>20</td>
</tr>
<tr>
<td>1.5</td>
<td>Measures of effectiveness of screening by age in the Swedish</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Two-Counties trial</td>
<td>22</td>
</tr>
<tr>
<td>1.6</td>
<td>Comparing the first year of screening for 10,000 women aged 40-49 with 10,000 women aged 50-64</td>
<td>23</td>
</tr>
<tr>
<td>2.1</td>
<td>Breast cancer registrations and deaths in New Zealand women</td>
<td>27</td>
</tr>
<tr>
<td>2.2</td>
<td>Stage distribution of breast cancer diagnosed in New Zealand women aged fifty to sixty-four in ten years from 1979 to 1988</td>
<td>29</td>
</tr>
<tr>
<td>2.3</td>
<td>Breast cancer incidence and mortality in New Zealand compared with other countries</td>
<td>31</td>
</tr>
<tr>
<td>2.4</td>
<td>A comparison of the maximum numbers of breast cancer deaths that could be prevented each year in New Zealand</td>
<td>36</td>
</tr>
<tr>
<td>2.5</td>
<td>Benefits and risks of breast cancer screening</td>
<td>38</td>
</tr>
<tr>
<td>2.6</td>
<td>Comparison of assessment and biopsy rates before and after the introduction of a breast cancer screening programme</td>
<td>42</td>
</tr>
<tr>
<td>2.7</td>
<td>Biopsy and detection rates in the second round of screening compared with the rates before the introduction of screening in the Netherlands</td>
<td>44</td>
</tr>
<tr>
<td>2.8</td>
<td>Essential features of an organised breast cancer screening programme</td>
<td>46</td>
</tr>
<tr>
<td>2.9</td>
<td>Principles of screening</td>
<td>49</td>
</tr>
<tr>
<td>2.10</td>
<td>Recommendations for pilot programmes in New Zealand</td>
<td>51</td>
</tr>
<tr>
<td>3.1</td>
<td>Summary of randomised controlled trials</td>
<td>62</td>
</tr>
<tr>
<td>3.1</td>
<td>Mortality end-points used in the randomised controlled trials</td>
<td>66</td>
</tr>
</tbody>
</table>
3.3 Percentage of cancers in localised stage, Connecticut Cancer Registry, 1940 to 1979 67
3.4 Health Insurance Plan histologically confirmed breast cancers 69
3.5 UK Screening Programme first screen (mammography and clinical examination) 70
3.6 Adjustment for self-selection in randomised controlled trials 75
3.7 Maximum breast cancer detection for each trial 76
3.8 Breast cancer detection and mortality reduction in randomised controlled trials 78
4.1 Characteristics of the first screening round of established mammographic screening programmes, and targets for the pilot programme 84
4.2 A comparison of two methods of estimating sensitivity in randomised controlled trials 90
4.3 Sensitivity and coverage in screening programmes 91
4.4 Action on failure to reach evaluation targets 100
5.1 Recommendations for mammography in Otago and Southland general practices in 1991 (before the start of the pilot programme) 106
5.2 The effect of letters from general practitioners 109
5.3 Comparison of telephoned with postal reminders 112
5.4 Assessment results for the first eighteen months of screening in the Otago and Southland pilot programme 118
5.5 Outcome of screening by reason for referral 119
5.6 Referral rates from September 1991 to March 1993 for the two main film-readers 122
5.7 Breast cancer incidence in women aged 50-64 in Otago and Southland before the establishment of the screening programme (1979-1988) 124
5.8 Sensitivity, specificity and predictive value in the first eighteen months of the Otago and Southland pilot programme 130
5.9 Stage distribution of tumours detected in the Otago and Southland programme 132
5.10 Breast cancers detected in the Otago and Southland prevalence screen compared with the Swedish Two-Counties prevalence screen 133
5.11 Suggested targets for size, histological grade, and nodal status for tumours detected in breast cancer screening programmes 136
6.1 Demographic characteristics of eligible respondents and women aged forty to seventy in Otago and Southland at the 1986 census 141
6.2 Factors affecting intention to participate in mammographic screening 142
6.3 Preferred location for a screening centre (women who intend to be screened) 143
6.4 Principal reason suggested by respondents that women may not participate in screening 144
6.5 Participation rate in response to an invitation in the first eighteen months of screening 151
6.6 Demographic characteristics of urban and rural participants in the Otago and Southland pilot programme 154
6.7 Anxiety related to screening 156
6.8 Anxiety related to screening by history of breast cancer in family or close friend 157
6.9 Anxiety while awaiting results, and time until results received 158
6.10 Anxiety on receiving invitation by anxiety while awaiting results 159
6.11 Experience of the screening test 160
6.12 Anxiety before screening and experience of screening 161
6.13 Anxiety while awaiting results and experience of screening 162
6.14 Amount women would be prepared to pay if the screening programme could not continue as a free service 164
6.15 Demographic characteristics of women attending the assessment clinic 169
6.16 Anxiety related to assessment appointment compared with screening invitation in 125 women who attended the assessment clinic 170
6.17 Anxiety related to being screened and being assessed 171
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.18</td>
<td>Time until results received</td>
<td>173</td>
</tr>
<tr>
<td>6.19</td>
<td>Anxiety while awaiting results of assessment by time until results received</td>
<td>174</td>
</tr>
<tr>
<td>6.20</td>
<td>Intention to participate in future</td>
<td>175</td>
</tr>
<tr>
<td>6.21</td>
<td>Reasons for attendance or non-attendance at screening</td>
<td>179</td>
</tr>
<tr>
<td>6.22</td>
<td>Practicalities of screening for women invited to screening</td>
<td>181</td>
</tr>
<tr>
<td>6.23</td>
<td>Anxiety on being invited to screening, and perceived risk of breast cancer</td>
<td>182</td>
</tr>
<tr>
<td>6.24</td>
<td>Intention to participate in future screening</td>
<td>183</td>
</tr>
<tr>
<td>7.1</td>
<td>Characteristics of the first screening round and targets for the pilot programme</td>
<td>192</td>
</tr>
<tr>
<td>7.2</td>
<td>Cost per quality adjusted life year (QALY) of competing treatments</td>
<td>208</td>
</tr>
<tr>
<td>7.3</td>
<td>Suggested targets for the incidence screens of the Otago and Southland pilot programme</td>
<td>212</td>
</tr>
</tbody>
</table>
CHAPTER ONE

The Efficacy of Mammographic Screening

Introduction

The first mammography screening programme was set up in the United States in 1956, where 1,312 asymptomatic women aged thirty-five and over volunteered to undergo six-monthly mammograms. The results were encouraging but their interpretation was limited because of the lack of an appropriate comparison group (Gershon-Cohen et al. 1967). The first randomised controlled trial of mammographic screening for breast cancer was started in 1963 (Shapiro 1977). Since then the results from several randomised controlled trials and numerous other studies of breast cancer screening have been published. These trials are described below (sections 1.2 - 1.4).

1.1 Evaluating the efficacy of breast cancer screening

As with other types of screening, evaluation of the efficacy of breast cancer screening can be affected by bias (Miller 1978). The four types of bias which cause particular problems for evaluating screening are lead time bias, length bias, overdiagnosis bias, and selection bias:

Lead time bias.
The survival time for women with breast cancer is the time from diagnosis until death. In screened women the diagnosis is made earlier than it would have been in the absence of screening; this is known as the "lead time" obtained by screening. The true effect of screening on breast cancer mortality cannot be assessed merely by comparing survival times for screened with unscreened women because longer survival times in the screened women will be at least partly due to lead time.
Length bias.
Breast tumours grow at different rates and therefore remain for differing periods in the presymptomatic screen-detectable phase. With each screening round the probability of detecting slow growing tumours is greater than the probability of detecting fast growing tumours, as slower growing tumours remain in the presymptomatic screen-detectable phase for longer. There will be fewer fast growing tumours in the screened compared with the unscreened group, and differences in outcome between groups may be partly due to this rather than to an effect of screening. Length bias is a greater problem with the first (prevalence) screen than with subsequent screens, since some of the tumours detected at the prevalence screen could have been in the pre-symptomatic screen-detectable state for longer than the interval between screens.

Overdiagnosis bias.
Screening detects very early lesions and it is possible that some of these cancers would never affect a woman in her lifetime (with the woman remaining asymptomatic and dying from some cause other than breast cancer). There is also the possibility that screening detects tumours that are not really malignant, but which are pathologically indistinguishable from early malignant breast cancers. Because these cancers are more likely to be found in a screened group than an unscreened group, comparisons of outcome could favour the screened group irrespective of any real effect of screening.

Selection bias.
Not all women who are invited, take part in screening. It may be that women who choose to take part have a different underlying risk of developing or dying from breast cancer. For instance women with a family history of breast cancer may decide to take part because they perceive themselves to be at higher risk. It is possible that this would be an even greater problem in screening programmes where women are not invited, but self-refer, since women who self-refer may differ more markedly in their risk of breast cancer. A particular problem with selection bias is that it can operate in two directions; if low-risk women are more likely to be screened, then breast cancer mortality is likely to be lower anyway in this group, and the effect of screening will be overestimated, whereas if high-risk women are more likely to be screened the effect of screening could be underestimated.
With each of these biases the problem lies in comparing dissimilar groups. If the groups are intrinsically different in time of diagnosis, type of cancer detected, and/or underlying risk of breast cancer, any difference in outcome between the groups cannot be attributed solely to screening. These biases can affect comparisons of tumour size, nodal involvement at diagnosis, and post-diagnosis survival time. The first three biases will cause overestimation of the benefits of screening, while the fourth, selection bias, could result in either overestimation or underestimation of the effect of screening.

The best information about the efficacy of breast cancer screening comes from randomised controlled trials with breast cancer mortality as the outcome measure. This is because with such a study design the groups compared are similar (due to randomisation) and the outcome measure (death from breast cancer) is independent of the time of diagnosis. To avoid selection bias the outcome for all the women in the intervention group is measured even though not all of these women will have accepted the offer of screening. This means that the efficacy of offering screening to a group of women is being measured rather than the efficacy of screening a group of women. Trials which are designed to measure the efficacy of screening enroll women who have already agreed to take part, and then randomly allocate them to be screened or not. This has not been a common approach in randomised controlled trials of breast cancer screening, with only two trials (the Canadian trials) designed in this way (section 1.2.6).

1.2 Randomised controlled trials of breast cancer screening

Several randomised controlled trials of mammographic screening have been carried out during the last thirty years. All of these trials used breast cancer mortality as the outcome measure. The trials are described in chronological order and the most recently published results from the trials are summarised in Table 1.1

1.2.1 New York Health Insurance Plan Trial

The Health Insurance Plan (HIP) trial was the first randomised controlled trial of mammographic screening for breast cancer (Shapiro 1977, Shapiro et al 1982, Shapiro et al 1988). In December
1963 62,000 female members of the HIP medical insurance scheme aged forty to sixty-four years, who had not previously been diagnosed with breast cancer were individually randomised to either an intervention or a control group of 31,000 women each. The women in the intervention group were invited to be screened while the women in the control group received their usual medical care. The trial was designed to be able to detect at least a twenty percent decrease in breast cancer mortality.

The first screening round began in 1963 and continued until June 1966. Annual rescreening ended in June 1970. In the intervention group sixty-five percent attended for an initial screen and they were offered screening annually for the next three years. Of these women, eighty-eight percent had at least one subsequent screen. Screening consisted of two-view mammography, and a clinical examination carried out by a physician. The two screening modalities were carried out independently so that the unique contribution of each modality could be assessed. Women were referred for assessment to a “large number of hospitals in which surgery was performed” (Shapiro 1977). All histology was reviewed by one pathologist who was blind to the intervention or control group status of the women.

The methods of follow up used were identical for intervention and control group women. Follow up consisted of contacting the women (or their next of kin), searches of hospital insurance and death records, and surveys of all women in the trial whose status was unknown, at five year intervals. Death from breast cancer was defined as death with breast cancer as the underlying cause. Deaths without histological confirmation (clinical or autopsy evidence only) were included.

By ten years after entry (about six years after screening ended) there were thirty percent fewer deaths from breast cancer in the intervention group than in the control group (RR 0.71, 95% CI 0.56-0.89). The reduction in breast cancer mortality was confined to women who were aged fifty or more at the time of diagnosis. Eighteen years after the trial began (Shapiro 1994) the relative risk was 0.77 (95% CI 0.61-0.97). The amount of mammography, physical examination, or breast self-examination in either the intervention or control group after the end of the trial is unknown, and the amount in the control group during the trial has not been reported.
1.2.2 Kaiser-Permanente Trial

In 1964 a randomised trial was set up to evaluate the effect of multiphasic health checks (Dales et al 1979). The study subjects were all men and women aged thirty-five to fifty-four who lived in the San Francisco Bay area and had been members of the Kaiser-Permanente Health Plan for at least two years. The subjects were randomly allocated to an intervention and a control group. The intervention group was offered health checks while the control group was not. One of the health checks offered was annual mammography for the next eleven years for women aged forty-eight and over. There were fewer than three thousand women in this category and three quarters of them were aged less than fifty when the study began. Breast cancer mortality was the same in the intervention and control groups after eleven years follow up, with fourteen deaths from breast cancer in each of the equal sized groups.

1.2.3 Swedish Two-Counties Trial

A randomised controlled trial of mammographic screening was started in Sweden in 1977 (Tabar and Gad 1981, Tabar et al 1985, Tabar et al 1992, Tabar et al 1995). In this trial 162,981 women in the counties of Östergötland and Kopparberg were randomly allocated to control or intervention groups. Randomisation occurred at community level (according to area of residence) rather than by individual. Women in the control group received their usual medical care. Screening was by single-view mammography only, and was repeated about every twenty-four months for women aged forty to forty-nine and about every thirty-three months for women aged fifty or more at entry. Of the women who were invited to be screened, eighty-nine percent attended the first screening round. The participation rate in women aged over seventy-four was less than fifty percent and published reports from this trial are confined to women aged forty to seventy-four at randomisation. Women in the control group were offered screening in 1985 (once three screening rounds had been completed in women aged fifty and over at entry, and after four rounds in women aged less than fifty at entry). Thirteen percent of the women in the control group had a mammogram during the study period. It is not stated in any of the published reports that any particular surgical protocol was followed, only that “ductography, pneumocystography, cytologic examination, or surgical excision are performed as required” (Tabar and Gad 1981).
According to published reports on this trial death from breast cancer was defined as due to breast cancer after review of clinical and pathological records. However this is inconsistent with a report from the Stockholm randomised controlled trial (Frisell et al 1991) where it was stated that death with breast cancer present at death was used as the end point in order to be consistent with the Two Counties trial.

After seven years there were thirty-one percent fewer deaths from breast cancer in the intervention group (RR 0.69, 95% CI 0.51 to 0.92). The reduction in breast cancer mortality was seen in women aged fifty or more (RR 0.61, 95% CI 0.44 to 0.84) but not for women under fifty (RR 1.26, 95% CI 0.56 to 2.84). The difference between intervention and control groups began to emerge after the fourth year of screening. By twelve years of follow up the overall result remained about the same (Tabar et al 1995), with a thirty percent reduction (RR 0.69, 95% CI 0.57 to 0.84) in breast cancer mortality in the women invited to screening, but in younger women by this time there was a non-significant reduction (RR 0.87, 95% CI 0.54 to 1.41 for women aged forty to forty-nine). There were differences in the results for the two counties (Tabar et al 1995), with Kopparberg county showing a greater reduction in breast cancer mortality (RR 0.60, 95% CI 0.46 to 0.79) compared with Östergötland county (RR 0.78, 95% CI 0.57 to 0.84), but these results are not significantly different (Breslow-Day test for homogeneity of effect, \( \chi^2 = 2.38, \ p = 0.13 \)).

### 1.2.4 Malmo Trial

A randomised controlled trial was also started in Malmö, Sweden in 1977 (Andersson et al 1988). The 42,000 women in Malmö aged forty-five to sixty-nine years were identified from the Malmö population registry and were individually randomised to intervention and control groups. Women who had been previously diagnosed with breast cancer were included in this trial but it was not reported whether such women were equally distributed between the intervention and control groups or by age. Staging was done according to the tumour diagnosed through screening, irrespective of the stage of any tumour previously diagnosed. Analyses which excluded women with previously diagnosed breast cancer did not appreciably alter the results (Andersson et al 1988).
Women in the intervention group were offered screening by mammography only, approximately every twenty-one months. In the first two screening rounds women were offered two-view mammography. In subsequent screening rounds women were offered single-view or two-view mammography depending on their type of breast parenchymal pattern. At the first screen seventy-four percent of women participated. Virtually all women who required treatment were treated at one hospital by a team specialising in breast diseases. The cause of death in patients with breast cancer was determined by an independent committee consisting of a pathologist and an oncologist who were blind to the intervention or control group status of the patients and who independently assessed the cause of death for all women who died after having been diagnosed with breast cancer.

After ten years (five completed screening rounds) there was no significant difference in breast cancer mortality between the two groups (RR 0.96, 95% CI 0.68-1.35). There was a twenty-one percent reduction in breast cancer mortality among intervention group women aged fifty-five and over; this reduction was not significant (RR 0.79, 95% CI 0.51 to 1.24). There was a non-significant excess mortality of similar magnitude among intervention group women under fifty-five (RR 1.29, 95% CI 0.74 to 2.25). By twelve years after the trial began (Shapiro 1994) there was a nineteen percent reduction in breast cancer mortality (RR 0.81, 95% CI 0.62 to 1.07). Nearly a quarter of the control women had a mammogram during the study period, and twenty percent of the control women who were diagnosed with breast cancer were diagnosed by mammography.

1.2.5 Edinburgh Trial
A randomised controlled trial of breast cancer screening was started in Edinburgh in 1979 (Roberts et al 1989). 45,130 women registered with eighty-four general practices in the city and aged forty-five to sixty-four who had not previously been diagnosed with breast cancer were entered into the trial. Screening was by two-view mammography and clinical examination at the first round. After this women were invited to be screened annually with clinical examination alone, in years two, four, and six, and single-view mammography and clinical examination in years three, five, and seven. Of the women invited sixty-one percent attended the first screening round. Women diagnosed with breast cancer were seen at a surgical review clinic. Death from breast cancer was defined as death where the death certificate recorded breast cancer as primary or secondary cause of death.
After seven years there was a non-significant seventeen percent reduction in breast cancer mortality among women offered screening (RR 0.83, 95% CI 0.58 to 1.18). This reduction was almost completely confined to women over fifty. In women aged fifty to sixty-four there was a non-significant reduction of twenty percent (RR 0.80, 95% CI 0.54 to 1.17). After ten years (Shapiro 1994) the reduction was sixteen percent (RR 0.84, 95% CI 0.63 to 1.12).

1.2.6 Canadian Trials
Two randomised controlled trials were started in Canada in 1980 (Miller et al 1992a, Miller et al 1992b). The women in the Canadian trials were volunteers who had no history of breast cancer and had not had a mammogram in the previous twelve months. These women were randomly allocated to control and intervention groups after having agreed to take part in the trial. Since all the women in the trial had agreed to take part in screening before they were allocated to either the intervention or control group, these trials were different from the other previously described trials; they investigated the effect of screening rather than the effect of an invitation to screening.

Trial I (women aged fifty to fifty-nine years)
The purpose of this trial was to measure the additional contribution of annual mammographic screening to screening by physical examination alone in women aged fifty to fifty-nine on entry. There were 39,405 women aged fifty to fifty-nine in this trial. All the women received an initial physical examination by a trained nurse-examiner or physician and were taught breast self examination. The women in the control group received annual physical examinations, while the intervention group received both annual physical examinations and annual two-view mammography. After eight years there was little difference in breast cancer mortality between groups (RR 0.97, 95% CI 0.62 to 1.52).

Trial II (women aged forty to forty-nine years)
This trial was designed to find out whether annual screening with mammography and physical examination reduces mortality from breast cancer in women aged forty to forty-nine on entry. There were 50,430 women aged forty to forty-nine in this trial. All the women received an initial physical examination by a trained nurse-examiner or physician and were taught breast self
examination. They were then randomly allocated to a control group or an intervention group. The women in the intervention group were offered annual two-view mammography and physical examination while the women in the control group were not offered further screening but were followed up annually by mail. After eight years there were more deaths from breast cancer among the intervention group than the control group but this difference was not statistically significant (RR 1.36, 95% CI 0.84 to 2.21).

In both of these trials referral for surgical treatment was the responsibility of each woman's physician rather than the screening programme, so there was no standard surgical protocol for the trials. Whether death was a result of breast cancer was determined by a panel which was blind to the intervention or control group status of the women. The panel reviewed the death certificates and case notes for any woman whose death could have been from breast cancer. This included women whose death certificates mentioned cancer of the breast, liver, or colon, and women whose cause of death was unknown.

There has been considerable controversy over the results of the Canadian trials and about the way the trials were conducted (Kopans 1990, Stacey-Clear et al 1992, Rasuli 1993, Rose 1993, Burhenne 1993, Miller 1993, Mettlin and Smart 1993, Baines 1994, Miller 1994, Tarone 1995). The trials were criticised for the quality of their mammography (Kopans 1990, Rasuli 1993, Burhenne 1993), although much work was done to monitor screening quality (Miller 1993, Baines 1994). The randomisation process was also questioned, since there was a significant excess of women with advanced breast cancer at the initial screen in the intervention group of the Canada II trial (Tarone 1995), however the randomisation was very carefully carried out and demographic variables, and breast cancer risk factors were equally distributed across the two groups (Baines 1994). Also there was no imbalance in advanced breast cancers at the initial screen in the fifty to fifty-nine age group in the Canada I trial, which used the same randomisation procedure (Mettlin and Smart 1993). It appears that the imbalance in women with advanced cancer occurred by chance (Mettlin and Smart 1993, Baines 1994).

One paper criticising the Canada II study was based on a comparison of survival times in women under fifty with breast cancer diagnosed by mammography compared with women whose cancers were diagnosed clinically (Stacey-Clear et al 1992). The paper concluded from the increased
survival time in the women diagnosed by mammography, that the Canada II study must be flawed. However, lead time bias could explain some, if not all, of the difference in survival between the two groups in the study by Stacey-Clear et al.

Probably the most important issue with respect to the Canadian trials is that the follow-up period was relatively short at the time the results were reported. Both trials could be affected by this. The Canada I trial differed from other trials in that it compared two screening programmes (mammography and physical examination compared with physical examination only) whereas other trials compared screening with no screening. It is likely that in this situation the difference between groups would be smaller, and could take longer to emerge (Baines 1994). The trial was designed to detect a forty percent reduction in breast cancer mortality. If the actual effect of adding mammography to physical examination is thirty percent or less, it will take longer follow-up, with more breast cancer deaths, to show a significant effect (Miller et al 1992).

In the Canada II trial there were relatively few breast cancer deaths at the time of publication of the results, and the trial did not achieve its planned power (Baines 1994). The results of the Canada II trial are not inconsistent with other trials, as several of the other trials showed an initial non-significant excess mortality in women under fifty in the first eight years of follow-up (Tabar et al 1985, Tabar et al 1992, Andersson et al 1988, Frisell et al 1986). Further results from both trials will be published after longer follow-up (Miller et al 1992).

1.2.7 Stockholm Trial
This randomised controlled trial started in Stockholm in Sweden in 1981 (Frisell et al 1986, Frisell et al 1991). 60,000 women aged forty to sixty-four were randomly allocated to an intervention group and a control group. No women were excluded from either the intervention or the control group. Screening was with single-view mammography and the screening interval was about twenty-eight months. At the first round eighty-one percent of invited women took part. After five years women in the control group were invited to take part in a single screening round. Once this screening round was completed no further diagnoses of cancer were included in the trial. The criteria for surgery were the same for women in the intervention and control groups. Death from breast cancer was defined as death with breast cancer present at death.
After 7.4 years follow up there was a non-significant twenty-nine percent reduction in breast cancer mortality in the women offered screening (RR 0.71, 95% CI 0.4 to 1.2). The mortality reduction in women over fifty was forty-three percent and was not significant (RR 0.57, 95% CI 0.3 to 1.1), while in women under fifty there was a non-significant nine percent increase in breast cancer mortality (RR 1.09, 95% CI 0.4 to 3.0). By eight years (Shapiro 1994) there was a twenty percent reduction in breast cancer mortality in the women offered screening (RR 0.80, 95% CI 0.53 to 1.22).

1.2.8 Gothenburg Trial
This trial was started in 1982. The results from this trial have only been published as part of an overview of the Swedish randomised trials of breast cancer screening (Nystrom et al 1993). In the Gothenburg trial 49,533 women aged forty to fifty-nine were randomly allocated to intervention and control groups. Women aged forty to forty-nine were individually randomised while the women aged fifty to fifty-nine were randomised by cluster, according to day of birth. Screening was by two-view mammography at eighteen month intervals. Participation in the first screening round was eighty-four percent. Women in the control group were invited to take part in screening in 1987. There is no information published about any surgical protocol used in the Gothenburg trial. After seven years follow up there was a non-significant fourteen percent reduction in breast cancer mortality among the women offered screening (RR 0.86, 95% CI 0.54 to 1.37). There are no published results comparing women under fifty with women aged fifty years and over.

1.2.9 Combined analyses of randomised controlled trials
There have been several combined analyses of the randomised controlled trials of mammographic screening; meta-analyses based on the published results of all the trials (Elwood et al 1993), or of the trials and case-control studies (Kerlikowske et al 1995) and an overview (combined analysis of the raw data) of all the Swedish trials (Nystrom et al 1993). Two meta-analyses focussing specifically on women under fifty have also been carried out (Glasziou et al 1995, Smart et al
### TABLE 1.1

Randomised controlled trials of breast cancer screening

<table>
<thead>
<tr>
<th>Trial</th>
<th>Year started</th>
<th>Age range</th>
<th>Study group</th>
<th>Control group</th>
<th>Screen modality</th>
<th>Screen interval in months</th>
<th>Years of follow up</th>
<th>Relative Risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIP</td>
<td>1963</td>
<td>40-64</td>
<td>31,000</td>
<td>31,000</td>
<td>2vM + P</td>
<td>12</td>
<td>18</td>
<td>0.77 (0.61-0.97)</td>
</tr>
<tr>
<td>S2C</td>
<td>1977</td>
<td>40-74</td>
<td>77,000</td>
<td>56,000</td>
<td>1vM</td>
<td>24 *</td>
<td>12</td>
<td>0.69 (0.57-0.84)</td>
</tr>
<tr>
<td>MAL</td>
<td>1977</td>
<td>45-69</td>
<td>21,000</td>
<td>21,000</td>
<td>2vM</td>
<td>21</td>
<td>12</td>
<td>0.81 (0.62-1.07)</td>
</tr>
<tr>
<td>EDB</td>
<td>1979</td>
<td>45-64</td>
<td>23,000</td>
<td>23,000</td>
<td>2vM + P</td>
<td>12</td>
<td>10</td>
<td>0.84 (0.63-1.12)</td>
</tr>
<tr>
<td>CAN I #</td>
<td>1980</td>
<td>50-59</td>
<td>19,700</td>
<td>19,700</td>
<td>2vM + P</td>
<td>12</td>
<td>8</td>
<td>0.97 (0.62-1.52)</td>
</tr>
<tr>
<td>CAN II</td>
<td>1980</td>
<td>40-49</td>
<td>25,200</td>
<td>25,200</td>
<td>2vM + P</td>
<td>12</td>
<td>8</td>
<td>1.36 (0.84-2.21)</td>
</tr>
<tr>
<td>STK</td>
<td>1981</td>
<td>40-64</td>
<td>40,000</td>
<td>20,000</td>
<td>1vM</td>
<td>28</td>
<td>8</td>
<td>0.80 (0.53-1.22)</td>
</tr>
<tr>
<td>GTH</td>
<td>1982</td>
<td>40-59</td>
<td>24,700</td>
<td>24,700</td>
<td>2vM</td>
<td>18</td>
<td>7</td>
<td>0.86 (0.54-1.37)</td>
</tr>
</tbody>
</table>

**Notes:**
- HIP: Health Insurance Plan
- EDB: Edinburgh
- GTH: Gothenburg
- S2C: Swedish Two-Counties
- MAL: Malmö
- CAN: Canadian trials
- STK: Stockholm
- 2vM: Two-view mammography
- P: Physical

* 24 months for women aged less than 50 years
** 33 months for women aged 50 and over
# CAN I compared mammography and physical examination with physical examination.
The first meta-analysis (Elwood et al 1993) showed a reduction in breast cancer mortality of thirty-four percent (RR 0.66, 95% CI 0.55 to 0.79) for women over fifty at entry, and no reduction in breast cancer mortality for younger women (RR 1.08, 95% CI 0.85 to 1.39). The second meta-analysis (Kerlikowske et al 1995) showed a twenty-six percent reduction (RR 0.74, 95% CI 0.66 to 0.83) for women over fifty, and a non-significant seven percent reduction (RR 0.93, 95% CI 0.76 to 1.13) for women under fifty. One meta-analysis of published data on screening in women under fifty (Glasziou et al 1995) found a non-significant reduction (RR 0.95, 95% CI 0.77 to 1.18), while the other (Smart et al 1995), which had a longer follow-up of between seven and eighteen years found a non-significant sixteen percent reduction for all the trials (RR 0.84, 95% CI 0.69 to 1.02) but a significant twenty-four percent (RR 0.76, 95% CI 0.62 to 0.95) reduction if the Canadian trial was excluded. The meta-analysis by Smart et al includes data that has not yet been published, but was presented at conferences. It also combines trials at different lengths of follow up, which wrongly assumes a constant breast cancer mortality rate over time.

The Swedish overview found a reduction in breast cancer mortality of twenty-nine percent (RR 0.71, 95% CI 0.57 to 0.89) among women aged fifty to sixty-nine at entry and a non-significant thirteen percent reduction (RR 0.87, 95% CI 0.63 to 1.20) for women aged forty to forty-nine. Further information about differences in the efficacy of screening for women of different age groups is given in section 1.4.

### 1.3 Non-randomised studies of breast cancer screening

Several non-randomised studies of screening have also been carried out. These studies are more likely to be affected by the biases associated with the evaluation of screening (section 1.1) and so are not as useful for measuring the true effect of mammographic screening on breast cancer mortality.

#### 1.3.1 UK Trial of Early Detection of Breast Cancer

The UK Trial of Early Detection of Breast Cancer (UKTEDBC) was a non-randomised comparison of methods of early detection of breast cancer carried out in eight districts in the United
Kingdom (UK Trial of Early Detection of Breast Cancer Group 1988a). Two districts (Edinburgh and Guildford) offered women mammographic screening, two (Huddersfield and Nottingham) offered women instruction in breast self-examination (BSE), and the other four (Dundee, Oxford, Southmead, and Stoke-on-Trent) were comparison districts. In Edinburgh women were randomly allocated to be offered screening or to receive their usual care; Edinburgh therefore conducted a randomised controlled trial (section 1.2.5) but also contributed to the non-randomised UK Trial of Early Detection of Breast Cancer.

The UKTEDBC started in 1971 and 45,841 women between forty-five and sixty-four were offered mammographic screening in Guildford and Edinburgh. There were 127,117 women in the comparison districts. Screening was by two-view mammography (Edinburgh) or single-view mammography (Guildford) and clinical examination at the first round. Of the women invited sixty-one percent in Edinburgh, and seventy-two percent in Guildford attended the first screening round. After this women were invited to be screened annually with clinical examination alone in years two, four, and six, and single-view mammography and clinical examination in years three, five, and seven.

Death from breast cancer was defined as the death of any woman in the study where the death certificate gave breast cancer as the primary cause. (In the Edinburgh randomised controlled trial the definition of death from breast cancer included deaths where breast cancer was given as the primary or the secondary cause). There was a non-significant reduction of twenty percent (RR 0.80, 95% CI 0.64 to 1.01) in the risk of dying from breast cancer among these women compared with women in the comparison population for whom no extra services were offered.

In the reports from this trial (UK Trial of Early Detection of Breast Cancer Group 1988a, UK Trial of Early Detection of Breast Cancer Group 1988b) it is recognised that the results should be interpreted with greater caution than those from the randomised controlled trials. However some comparisons were made between the UKTEDBC results and those from previously published randomised controlled trials (HIP and Swedish Two-Counties). The reasons suggested for the smaller benefit observed in the UKTEDBC included difficulties with identifying and inviting eligible
women, low participation rates, self-selection of lower risk women for screening, and lower sensitivity of the screening test (UK Trial of Early Detection of Breast Cancer Group, 1988b). There were also differences in treatment protocols between the districts (UK Trial of Early Detection of Breast Cancer Group, 1988b).

1.3.2 Breast Cancer Detection Demonstration Project
The Breast Cancer Detection Demonstration Project (BCDDP) was set up in America in 1972 (Seidman et al 1987). Its purpose was to show that breast cancer screening for women aged thirty-five to seventy-four could be implemented on a large scale. The BCDDP was a demonstration project rather than a scientific study and it did not include a comparison group. The women who took part were screened with annual clinical examination, mammography, and in some centres thermography. The women were also taught breast self examination (BSE). The records for women who took part were centrally processed at a data management centre in Philadelphia. There was some controversy about the relevance of the BCDDP results and a working group was set up to review the BCDDP. This working group reported in 1979 (Working Group to Review the National Cancer Institute-American Cancer Society Breast Cancer Detection Demonstration Projects 1979) concluding that results from the BCDDP could not be used to provide evidence on the efficacy of screening because of the lack of a suitable comparison group.

1.3.3 Case-control studies
Several case-control studies have been carried out to measure the effectiveness of established breast cancer screening programmes. These studies reported a reduced risk of death from breast cancer for women who had taken part in screening (Verbeek et al 1984, Verbeek et al 1985, Collette et al 1984, Collette et al 1992, Palli et al 1986, Palli et al 1989). In these studies the cases were women eligible for screening who had died from breast cancer after their first screening invitation. The control women were age-matched women who were eligible for screening. The screening history for both cases and controls was taken up to and including the date of diagnosis of the case. The results from these case-control studies are summarised in Table 1.2.
TABLE 1.2

Case-control studies of breast cancer screening

<table>
<thead>
<tr>
<th>Study</th>
<th>Odds Ratio (95% CI)</th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nijmegen 35-65 1975</td>
<td>0.48 (0.33-1.0)</td>
<td>Women diagnosed after their first invitation and who died before 1981</td>
<td>All women invited, who had not died from breast cancer at the time when the case died (five age-matched controls per case)</td>
</tr>
<tr>
<td>Utrecht 1 50-64 1975</td>
<td>0.30 (0.03-0.7)</td>
<td>Women whose diagnosis occurred after the start of the programme and who died before 1981</td>
<td>Utrecht women of the same age as the case (three age-matched controls per case)</td>
</tr>
<tr>
<td>Utrecht 2 50-64 1975</td>
<td>0.52 (0.44-0.79)</td>
<td>Women whose diagnosis occurred after the start of the programme and who died before 1987</td>
<td>Utrecht women of the same age as the case (three age-matched controls per case)</td>
</tr>
<tr>
<td>Florence 40-70 1970</td>
<td>0.53 (0.29-0.95)</td>
<td>Women diagnosed after their first invitation who died in the years 1977-84</td>
<td>Florence women matched for age and residence (five controls per case)</td>
</tr>
</tbody>
</table>
1.3.4 Case-control analysis of Guildford data

Two case-control studies have been carried out related to one of the screening districts (Guildford) and a comparison district (Stoke-on-Trent) from the UKTEDBC (Moss et al 1992). The first study (study A) estimated the risk of death from breast cancer in women in the screening district relative to the control district, where screening had not been offered, and the second (study B) estimated the relative risk for women who had ever been screened compared with women who had never been screened, in the screening district alone. Study B was carried out to assess the extent of selection bias in a case-control study carried out among women who had been offered screening. The results of this study, and the effect of selection bias on case-control analyses of breast cancer screening are discussed in Chapter Three (section 3.1.3). The relative risk in study A was 0.76 (95% CI 0.54 to 1.08), suggesting a twenty-four percent reduction in breast cancer mortality in the screening district compared with the comparison district. The women in the study were aged forty-five to sixty-four. The results for women under fifty were not examined separately.

1.3.5 Case-control studies and the efficacy of breast cancer screening

Case-control studies cannot measure the efficacy of screening very accurately because of the likelihood of selection bias. The studies may overestimate or underestimate the effect of screening, depending on the direction of selection bias. The studies are consistent however in finding a reduction in breast cancer mortality for women over fifty. The two studies which included women under fifty were unable to demonstrate a significant protective effect of screening for women under fifty.

1.4 Breast cancer screening and age

Even in the earliest randomised controlled trials of breast cancer screening it was noticed that the reduction in breast cancer mortality was largely confined to women who were over fifty at entry into the studies. However there was a problem with investigating this because, apart from the
Canada II trial, none of the trials was designed to look at the issue of screening for women under fifty, so the numbers of women aged less than fifty in the trials were small. Whether it is reasonable to analyse sub-groups within the trials has been questioned (Kopans 1994), but there is significant heterogeneity between the results for older and younger women in the trials (Elwood et al 1993), suggesting that it is legitimate to analyse the results in the two age groups separately. The one trial that was designed to look at the question of screening for women under fifty, the Canada II trial, (section 1.2.6) found no evidence of benefit.

Meta-analyses have been carried out (section 1.2.9) showing that breast cancer screening for women over fifty results in about a thirty-four percent reduction in breast cancer mortality. For women under fifty there is no good evidence of benefit. Since the first meta-analysis (Elwood et al 1993) was published several others have been unable to demonstrate any benefit of screening for women under fifty. An overview of the Swedish randomised controlled trials (Nystrom et al 1993) found a non-significant thirteen percent reduction in mortality for women under fifty, but a recent analysis estimated that about seventy percent of this mortality reduction was likely to be due to screening in these women after they had reached the age of fifty (de Koning et al 1995a). This issue also affects the interpretation of the meta-analysis carried out by Smart et al since it included results of trials after long follow-up periods of ten years or more. There is a trial under way in the UK to investigate this issue, but the results will not be available for several years. In this trial women are randomised to be offered screening from the age of forty and then to enter the national screening programme at fifty, or to enter the national screening programme at fifty with no prior offer of screening. The trial will investigate whether there is any additional benefit of screening women from forty to forty-nine in addition to screening from fifty to sixty-four (Murphy and Odartchenko 1993).

Tables 1.3 and 1.4 compare the results from the randomised controlled trials at seven years follow-up, according to the age at entry of the women screened. The Canadian trials (section 1.2.6) were not included since they differed from the other trials in assessing the efficacy of screening, rather than the efficacy of offering screening. The Gothenburg trial (section 1.2.8) is not included because the results from this trial have not been published (except as part of the overview of Swedish trials). Unfortunately none of these trials was specifically designed to provide
TABLE 1.3

Results of Randomised Controlled Trials by Age:
Younger Women

<table>
<thead>
<tr>
<th>Trial</th>
<th>Age group</th>
<th>Person-years of follow up</th>
<th>Breast cancer deaths</th>
<th>Mortality ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
</tr>
<tr>
<td>HIP</td>
<td>40-49</td>
<td>93,878</td>
<td>94,714</td>
<td>30</td>
</tr>
<tr>
<td>Malmö</td>
<td>45-54</td>
<td>55,145</td>
<td>55,938</td>
<td>17</td>
</tr>
<tr>
<td>S2C</td>
<td>40-49</td>
<td>136,459</td>
<td>107,836</td>
<td>26</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>45-49</td>
<td>40,851</td>
<td>40,009</td>
<td>13</td>
</tr>
<tr>
<td>Stockholm</td>
<td>40-49</td>
<td>99,155</td>
<td>54,446</td>
<td>16</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>425,488</td>
<td>352,943</td>
<td>102</td>
</tr>
</tbody>
</table>

* Mantel-Haenszel odds ratio

### TABLE 1.4

**Results of Randomised Controlled Trials by Age:**  
**Older Women**

<table>
<thead>
<tr>
<th>Trial</th>
<th>Age group</th>
<th>Person-years of follow up</th>
<th>Breast cancer deaths</th>
<th>Mortality ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Intervention</td>
<td>Control</td>
<td>Intervention</td>
</tr>
<tr>
<td>HIP</td>
<td>50-64</td>
<td>112,898</td>
<td>115,024</td>
<td>39</td>
</tr>
<tr>
<td>Malmo</td>
<td>55-69</td>
<td>88,278</td>
<td>88,221</td>
<td>27</td>
</tr>
<tr>
<td>S2C</td>
<td>50-69</td>
<td>315,262</td>
<td>223,523</td>
<td>75</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>50-64</td>
<td>117,095</td>
<td>107,845</td>
<td>55</td>
</tr>
<tr>
<td>Stockholm</td>
<td>50-64</td>
<td>171,092</td>
<td>92,927</td>
<td>23</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>804,625</td>
<td>627,540</td>
<td>219</td>
</tr>
</tbody>
</table>

* Mantel-Haenszel odds ratio

information on the efficacy of screening for women at different ages, and the small numbers of women involved together with a lower risk of breast cancer in the younger age group mean that only large differences in mortality would have been statistically significant. But there is significant heterogeneity between the results shown in Tables 1.4 and 1.5 for the younger and older groups (Breslow-Day test of homogeneity of effect $\chi^2 = 4.34$, df = 1, $p = 0.04$), and although the individual trials had inadequate power to detect small reductions in mortality for younger women the meta-analysis by Elwood et al. had adequate statistical power to detect a twenty percent reduction in mortality in the younger group. None of the combined analyses found convincing evidence of a reduction in breast cancer mortality resulting from screening women under fifty, but they were consistent in finding that screening reduced mortality by about a third for women over fifty.

The difference in outcome for the two age-groups may be related to the performance of the screening test. It is clear from the Swedish Two-Counties trial (Tabar et al. 1992) that mammography produces different results in the two age-groups (Table 1.5). In Table 1.5 “sojourn time” refers to the amount of time that a cancer remains pre-clinical but screen-detectable (this can also be interpreted as the mean lead-time gained by screening), and “predictive value” refers to the percentage of breast cancers diagnosed in the prevalence screen that would later have come to diagnosis in the absence of screening. A low predictive value suggests overdiagnosis due to screening, and this was found particularly in women under fifty. The differences for younger women seen in Table 1.5 may also be related to their pre-menopausal status. Most women aged less than fifty are premenopausal and therefore have more glandular tissue in their breasts. This tends to make their breast tissue denser, so that mammography may be less sensitive (Tabar et al. 1995).

Some authors have been reluctant to accept that screening could be less effective for women under fifty (Kopans 1994). It has been suggested, on the basis of the shorter mean sojourn time for women under fifty, that the screening test should be carried out more frequently in younger women, and recommendations of annual mammography for women aged forty to forty-nine have been made (Kopans 1994, Tabar et al. 1995). However, if the results from the trials are applied to the actual situation that would result from offering annual screening to women under fifty, it is clear
TABLE 1.5

Measures of effectiveness of screening by age in the Swedish Two-Counties trial

<table>
<thead>
<tr>
<th>Age group</th>
<th>40-49</th>
<th>50-59</th>
<th>60-69</th>
<th>70-74</th>
<th>50-69 *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incidence in controls per 1000 woman-years</td>
<td>1.05</td>
<td>1.87</td>
<td>2.50</td>
<td>2.99</td>
<td></td>
</tr>
<tr>
<td>Prevalence in first screen per 1000</td>
<td>2.09</td>
<td>4.67</td>
<td>8.80</td>
<td>12.15</td>
<td></td>
</tr>
<tr>
<td>Ratio prevalence to incidence</td>
<td>1.99</td>
<td>2.50</td>
<td>3.52</td>
<td>4.06</td>
<td></td>
</tr>
<tr>
<td>Incidence in year after screen (% of expected)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1</td>
<td>46</td>
<td>10</td>
<td>17</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Year 2</td>
<td>53</td>
<td>28</td>
<td>27</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>60%</td>
<td>86%</td>
<td>86%</td>
<td>95%</td>
<td>86%</td>
</tr>
<tr>
<td>Mean sojourn time</td>
<td>1.25</td>
<td>3.03</td>
<td>3.89</td>
<td>3.41</td>
<td>3.47</td>
</tr>
<tr>
<td>“Predictive value”</td>
<td>38%</td>
<td>100%</td>
<td>95%</td>
<td>80%</td>
<td>97%</td>
</tr>
</tbody>
</table>

* combined

### TABLE 1.6

Comparing the first year of screening for 10,000 women aged 40-49 with 10,000 women aged 50-64

<table>
<thead>
<tr>
<th>Age group</th>
<th>40-49</th>
<th>50-59</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening interval (years)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Number of women screened per year</td>
<td>10,000</td>
<td>5,000</td>
</tr>
<tr>
<td>Number referred per year</td>
<td>770</td>
<td>360</td>
</tr>
<tr>
<td>Number of biopsies per year</td>
<td>42</td>
<td>50</td>
</tr>
<tr>
<td>Women diagnosed with breast cancer each year as a result of screening</td>
<td>21</td>
<td>32</td>
</tr>
<tr>
<td>Benign : malignant biopsy ratio</td>
<td>2 : 1</td>
<td>0.6 : 1</td>
</tr>
<tr>
<td>Percentage of the cancers expected in the next year that were detected early by screening</td>
<td>60%</td>
<td>85%</td>
</tr>
<tr>
<td>Evidence that screening reduces breast cancer mortality</td>
<td>NO</td>
<td>YES</td>
</tr>
</tbody>
</table>

that not only are the benefits of screening not proven for this group, but the risks of false positive and false negative tests are greater (Table 1.6). Because mammography is less sensitive in younger women, there will be more false negative tests than in women over fifty. There are also more false positive tests in younger women, and the benign to malignant biopsy ratio is higher than in women over fifty. All of these risks are compounded by the fact that the women would be screened twice as often as women over fifty (who are offered screening every two years). Finally, and most importantly, these risks are not balanced by any evidence of benefit in women under fifty. The risks and benefits of screening are discussed further in Chapter Two.

1.5 Summary

There have been many studies of the efficacy of mammographic screening for breast cancer. The most useful studies are the randomised controlled trials because their results are least likely to have been affected by bias or confounding. The non-randomised studies, while requiring more cautious interpretation, have been useful in supplementing the results from the randomised controlled trials and have been used to explore the practical aspects of population-based screening. Together, the results show that under the best conditions, mammographic screening for breast cancer can reduce breast cancer mortality in women over fifty years by about a third. At present there is not good evidence that mammographic screening is efficacious for women under the age of fifty.
CHAPTER TWO

The Establishment of Population-Based Breast Cancer Screening in New Zealand

Introduction

This chapter examines the context in which population-based breast cancer screening programmes were introduced in New Zealand. The epidemiology of breast cancer in New Zealand, and the way data on breast cancer incidence are collected are described in section 2.1. The likely effect of breast cancer screening is compared with other possible strategies to reduce breast cancer mortality in section 2.2. The risks and benefits of breast cancer screening are investigated, and the characteristics of an organised screening programme are described in section 2.3. Population-based screening is discussed in section 2.4 and the factors leading up to the introduction of population-based breast cancer screening in New Zealand, and the establishment of the pilot programme and its evaluation are outlined in section 2.5.

2.1 Breast cancer in New Zealand

New Zealand has a population of about three and a half million people. Eighty percent of the population are of European descent, ten percent are Maori (New Zealand Health Information Service 1993). Breast cancer is an important cause of morbidity and mortality in New Zealand. It is the most common cause of death from cancer in women and causes eighteen percent of all female deaths from cancer and five percent of all female deaths. In 1991 the age-standardised incidence rate (standardised to Segi’s world population) for New Zealand women was 74.3 per hundred thousand per year. Breast cancer incidence
rates for Maori and non-Maori women are similar (New Zealand Health Information Service 1994). The incidence of breast cancer increases with age, and in New Zealand seventy percent of breast cancer registrations are in women over fifty years old (Table 2.1). From 1980 to 1991 there was a ten percent increase in breast cancer registrations but the breast cancer mortality rate remained unchanged over this time at about twenty-six deaths per hundred thousand per year (New Zealand Health Information Service 1994). Mortality rates for Maori women were lower than for non-Maori women, at about twenty-two per hundred thousand per year. Each year about fifteen hundred New Zealand women are diagnosed with breast cancer and about six hundred women die from it (New Zealand Health Information Service 1994).

Information about cancer incidence and mortality in New Zealand is collected by the New Zealand Cancer Register. This Register was established in 1948 within the National Health Statistics Centre of the Department of Health. Initially the Register consisted of information about people treated for cancer in public hospitals only. Information on cancers of all sites was initially collected, but in 1958 because of resource constraints, registration of skin cancers (apart from malignant melanoma) was discontinued. In 1972 cancer registration was extended to private hospitals. The Cancer Registry also received notice of all deaths occurring in New Zealand every year and used this information to update its records and create new records if necessary. Collection of data for the Cancer Register was carried out routinely by hospital coders in public hospitals, but in private hospitals it was the responsibility of the medical practitioner treating the patient, and was not compulsory. This led to problems with under-reporting of cancers where people were treated outside the public hospital system, and concern that the proportion of all cancers registered by the Cancer Registry had varied over time (Cox et al 1989). In 1988 a working group recommended that legislation should be enacted to make it compulsory for all pathology laboratories (both public and private) to supply a copy of any pathology report with the diagnosis of cancer and related conditions to the New Zealand Cancer Registry (Cancer Registration Working Group, 1988). It was hoped that this would increase coverage, and also provide legal protection for medical practitioners who were
### TABLE 2.1

Breast cancer registrations and deaths in New Zealand women

<table>
<thead>
<tr>
<th>Age group</th>
<th>Number of registrations in 1991</th>
<th>Age-specific rate (per 100,000)</th>
<th>Number of deaths in 1990</th>
<th>Age-specific mortality rate (per 100,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5-9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>10-14</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>15-19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20-24</td>
<td>2</td>
<td>1.5</td>
<td>-</td>
<td>1.4</td>
</tr>
<tr>
<td>25-29</td>
<td>16</td>
<td>11.5</td>
<td>2</td>
<td>7.9</td>
</tr>
<tr>
<td>30-34</td>
<td>39</td>
<td>28.2</td>
<td>11</td>
<td>17.9</td>
</tr>
<tr>
<td>35-39</td>
<td>85</td>
<td>68.1</td>
<td>22</td>
<td>17.9</td>
</tr>
<tr>
<td>40-44</td>
<td>158</td>
<td>132.8</td>
<td>37</td>
<td>32.4</td>
</tr>
<tr>
<td>45-49</td>
<td>159</td>
<td>170.5</td>
<td>57</td>
<td>62.9</td>
</tr>
<tr>
<td>50-54</td>
<td>172</td>
<td>215.4</td>
<td>45</td>
<td>57.6</td>
</tr>
<tr>
<td>55-59</td>
<td>153</td>
<td>223.0</td>
<td>66</td>
<td>94.6</td>
</tr>
<tr>
<td>60-64</td>
<td>171</td>
<td>243.7</td>
<td>79</td>
<td>112.6</td>
</tr>
<tr>
<td>65-69</td>
<td>174</td>
<td>261.7</td>
<td>73</td>
<td>112.2</td>
</tr>
<tr>
<td>70-74</td>
<td>161</td>
<td>288.6</td>
<td>71</td>
<td>132.8</td>
</tr>
<tr>
<td>75-79</td>
<td>119</td>
<td>264.4</td>
<td>69</td>
<td>153.8</td>
</tr>
<tr>
<td>80-84</td>
<td>101</td>
<td>339.3</td>
<td>53</td>
<td>180.0</td>
</tr>
<tr>
<td>85+</td>
<td>84</td>
<td>382.9</td>
<td>50</td>
<td>237.8</td>
</tr>
<tr>
<td>Total</td>
<td>1594</td>
<td>635</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


reluctant to notify the registry because of concerns about patient confidentiality. Little progress was made however, and by 1989 private medical practitioners in Auckland, New Zealand's largest city, had refused to register their patients, and data collection at the registry had fallen several years behind (Skegg 1989). It was suggested that if this downward trend continued the Cancer Register would no longer be accurate enough to be used to evaluate cancer control initiatives such as cervical and breast cancer screening (Skegg 1989).

In 1991 a review committee was set up (Cancer Registry Working Group 1991) which reiterated the need for legislation to make registration compulsory, since the existing system could miss people who were treated only at outpatient clinics or at private hospitals, and also suggested a minimum delay of eighteen months between collection of data and its publication by the registry. Later the chair of the review committee reported on progress since the review, a bill having been introduced in Parliament in mid-1992 to make cancer registration compulsory (Elwood 1992). This legislation, the Cancer Registration Act, was passed in 1993 to take effect from 1 July 1994. It is believed that cancer registrations for the period 1980 to 1991 represented at least ninety-five percent of all tumours (National Health Statistics Centre 1989, New Zealand Health Information Service 1994).

From 1979 to 1988, when there was little use of mammography for screening in New Zealand, the stage distribution of breast cancer at diagnosis (New Zealand Cancer Registry 1990) for women aged fifty to sixty-four was as shown in Table 2.2. As would be expected in the absence of widespread screening, there was very little ductal carcinoma in situ (DCIS) diagnosed, fifty-one percent of women were diagnosed with localised breast cancer, thirty-seven percent with regional or node involvement, and only eight percent of women had distant metastases at the time of diagnosis. The stage at diagnosis was not stated for three percent.

It is interesting to compare the stage-distribution at diagnosis for New Zealand women in the absence of screening with the control groups in randomised controlled trials of screening since this approximates the baseline from which screening will have an effect.
TABLE 2.2

Stage distribution of breast cancer diagnosed in New Zealand women aged fifty to sixty-four in ten years from 1979 to 1988

<table>
<thead>
<tr>
<th>Stage</th>
<th>in situ</th>
<th>localised disease</th>
<th>regional or node involvement</th>
<th>distant metastases</th>
<th>not stated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>42</td>
<td>2,289</td>
<td>1,677</td>
<td>368</td>
<td>122</td>
</tr>
<tr>
<td>Percent</td>
<td>1</td>
<td>51</td>
<td>37</td>
<td>8</td>
<td>3</td>
</tr>
</tbody>
</table>


Stage-distribution data: Dr J Elisabeth Wells, Department of Public Health and General Practice, Christchurch School of Medicine
The percentage of women diagnosed with localised disease in New Zealand prior to the introduction of screening was higher than in the control group in the Health Insurance Plan (HIP) trial where forty-six percent of control women had node negative tumours, but close to that for control group women in the Swedish Two-Counties trial, where fifty-four percent of women had node-negative tumours (Shapiro et al 1988, Tabar et al 1992). The Swedish Two-Counties trial showed a significant reduction in breast cancer mortality when screening was offered to a population where, in the absence of screening, there was a similar stage-distribution at diagnosis to that of New Zealand women in the absence of screening. This suggests that screening in New Zealand has the potential (assuming a screening programme of equivalent quality) to have a similar impact to the Swedish Two-Counties trial. The issue of stage-distribution at diagnosis in the absence of screening and its likely effect on the effectiveness of screening are discussed in more detail in Chapter Three (section 3.6.3). There are large inter-country differences in breast cancer incidence (Table 2.3), with rates in the highest risk countries being five times the rates in countries with the lowest risk (Parkin et al 1992). New Zealand has among the highest breast cancer incidence and mortality rates in the world (Parkin et al 1992, Boring et al 1993). The inter-country differences in breast cancer incidence are further discussed under prospects for the primary prevention of breast cancer in section 2.2.1.

2.2 Comparing strategies to reduce breast cancer mortality in New Zealand.

Breast cancer is a major health problem in New Zealand and it is important to find ways to reduce its impact on New Zealand women. This section focuses on breast cancer mortality. There are three preventive strategies which could reduce breast cancer mortality. They are removal of or protection from the cause(s) of breast cancer (primary prevention), early detection by screening (secondary prevention) and improvements in treatment (tertiary prevention). In this section the three strategies are compared. The comparison is restricted to women over fifty years because screening has proven efficacy only in this age-range.
TABLE 2.3

Breast cancer incidence and mortality in New Zealand compared with other countries

<table>
<thead>
<tr>
<th>Country</th>
<th>Age standardised breast cancer incidence per 100,000 *</th>
<th>Age standardised breast cancer mortality per 100,000 #</th>
</tr>
</thead>
<tbody>
<tr>
<td>New Zealand</td>
<td>74.3</td>
<td>26.8</td>
</tr>
<tr>
<td>USA</td>
<td>89.2 white</td>
<td>22.4</td>
</tr>
<tr>
<td></td>
<td>65.0 black</td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>71.1</td>
<td>24.2</td>
</tr>
<tr>
<td>Denmark</td>
<td>68.6</td>
<td>28.3</td>
</tr>
<tr>
<td>Sweden</td>
<td>62.5</td>
<td>17.4</td>
</tr>
<tr>
<td>Australia</td>
<td>59.0</td>
<td>20.6</td>
</tr>
<tr>
<td>England</td>
<td>56.0</td>
<td>29.3</td>
</tr>
<tr>
<td>Finland</td>
<td>52.5</td>
<td>17.3</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>32.3</td>
<td>8.4</td>
</tr>
<tr>
<td>Japan</td>
<td>25.0</td>
<td>6.0</td>
</tr>
<tr>
<td>China</td>
<td>21.5</td>
<td>4.7</td>
</tr>
</tbody>
</table>

* Standardised to Segi's world population

# Standardised to WHO world standard population
2.2.1 Primary prevention

The ideal strategy would be to prevent breast cancer from occurring (primary prevention). Since the cause of breast cancer is unknown, complete prevention of the disease is not possible. Despite this, it might seem that there is some potential to reduce breast cancer incidence as migrant studies suggest that the differences in breast cancer incidence between countries are due to environmental rather than genetic factors (Buell 1973, Shimizu et al 1991, Kelsey and Horn-Ross 1993, Zeigler et al 1993). If all the environmental factors were identified and if they were all modifiable, it would be possible to reduce breast cancer incidence everywhere to the level found in low incidence countries. But it is clear that all the factors that contribute to inter-country differences have not been identified, as these international differences cannot be explained on the basis of differences in known risk factors between countries (Thomas 1991, Thomas 1993).

Even with respect to the recognised risk factors primary prevention is a distant goal; most of the women who develop breast cancer could not have been previously identified as being at higher risk of the disease (Love 1995). Also many of the recognised risk factors are not easily modifiable; for instance family history, age at menarche, and age at menopause (Bush and Helzlsouer 1993, Brinton 1994). Others such as age at first pregnancy or the number of children that a woman has are usually determined by considerations other than the wish to reduce risk of breast cancer. Dietary modifications are a possibility but studies investigating the risk associated with a high fat diet or with alcohol have been inconclusive (Brinton 1994, Lund 1994, Toniolo et al 1994). Two large cohort studies (Willett et al 1987, Jones et al 1987) failed to show an association between high fat intake and breast cancer risk. A randomised controlled trial to investigate the effect of a low fat diet on breast cancer incidence is presently underway in Canada (Love 1995), and a similar trial, the Women's Health Initiative Dietary Modification Trial, is underway in the United States (Henderson 1995) but the results will not be available for several years. A high level of physical activity may slightly reduce the risk of breast cancer (Brinton, 1994). Genetic testing for women with a strong family history is now a possibility. Some women with a strong family history and a very high risk of breast cancer can be offered genetic testing to see whether they are at risk (Vogelstein and Kinzler 1994); some of these women opt for prophylactic mastectomy, but this is not a desirable option for most women, where the risk
is very much less. Genetic testing has the potential to prevent about two percent of women from developing breast cancer (Easton et al 1993, Easton et al 1994) although this would be likely to affect women younger than fifty since familial breast cancer tends to occur earlier than sporadic breast cancer. Chemoprevention is another possible means of prevention and a trial is presently under way to find out whether tamoxifen taken regularly can reduce the incidence of breast cancer among high risk women aged thirty-five and over (Bush and Helzlsouer 1993). The results from this trial will not be available until 1997.

Estimates of the possible reduction in breast cancer that could be achieved by modifications in lifestyle have been made, with the highest estimate being forty-five percent (Miller 1992) but incorporated in this estimate was the effect of diet, in particular dietary fat, which was estimated by Miller to be about twenty-seven percent but it may be much less (Lund, 1994, Toniolo et al, 1994). It is now widely accepted that primary prevention of breast cancer is not yet feasible because most of the established risk factors do not increase risk greatly, and most are not modifiable (Kelsey 1993, Brinton 1994, Love 1995). Another important feature of primary prevention as a strategy to reduce breast cancer mortality in women over fifty is that it appears that the risk of breast cancer may be determined early in life (Bjarnason 1974, de Waard 1991, Ekbom et al 1992, Kelsey and Horn-Ross 1993) so even if primary prevention were possible, it could take several decades to have any effect.

2.2.2 Screening

The efficacy of screening for breast cancer has been discussed in Chapter One. The aim of screening is to detect disease at an earlier stage when treatment may be more successful. Randomised controlled trials of breast cancer screening using mammography have provided clear evidence that screening can reduce breast cancer mortality by about thirty percent in a population of women over fifty who are offered screening. In New Zealand in 1984, before there was much use of mammography for screening, 195 women aged from fifty-five to sixty-nine died from breast cancer. From the results of the randomised controlled trials screening has the potential to reduce this by thirty percent (about fifty-eight women). The thirty percent reduction has been applied to the fifty-five to sixty-nine age group since
breast screening has a delayed effect on mortality, beginning about five years after entry to screening (Rutqvist et al 1990).

2.2.3 Treatment
The third strategy would be to improve the treatment of breast cancer (tertiary prevention). This strategy is worthwhile, but so far the gains from treatment of breast cancer have been relatively small. A systematic overview of the treatment of early breast cancer (stage one and two) showed that appropriate treatment (with tamoxifen and adjuvant chemotherapy compared with treatments that did not include tamoxifen and adjuvant chemotherapy) improved five-year survival from about sixty-five percent to seventy-two percent for women aged over fifty (Early Breast Cancer Trialists’ Collaborative Group 1990). In 1984 (before the introduction of widespread mammography) there were 433 women aged between fifty and sixty-four diagnosed with breast cancer in New Zealand (National Health Statistics Centre, 1989). The distribution of stage at diagnosis for fifty to sixty-four year old women in the ten years from 1979 to 1988 is shown in Table 2.2 (section 2.1). Up to eighty-eight percent of women aged fifty to sixty-four were diagnosed with “early breast cancer”. This equates to 381 women diagnosed with “early breast cancer” from the total of 433 women diagnosed with breast cancer in New Zealand in 1984. From the overview of treatment of early breast cancer (Early Breast Cancer Trialists’ Collaborative Group 1990), treatment including tamoxifen and adjuvant chemotherapy can improve the five year survival of women over fifty by up to nine percent, compared with treatment that does not include tamoxifen and adjuvant chemotherapy. It is likely that some of the 381 women aged fifty to sixty-four diagnosed with early breast cancer would have received treatment with tamoxifen and adjuvant chemotherapy (Mason et al 1994), so adding a nine percent improvement in five year survival (from about sixty-five percent without tamoxifen and adjuvant chemotherapy to seventy-two percent with), is probably an overestimate, but means that after five years, of the 381 women, the number surviving would rise from 248 (sixty-five percent) to 274 (seventy-two percent), and so up to twenty-six extra women would be alive five years after diagnosis.

For women with more advanced breast cancer five year survival is lower and the results of treatment have not been as encouraging (Peto and Easton 1989), and a similar systematic
overview has not been done. As a way of estimating the maximum effect that could be obtained from optimal treatment, a nine percent increase in five year survival from sixty-five percent to seventy-two percent can be applied to all 433 women aged fifty to sixty-four who were diagnosed with breast cancer. The number of women surviving five years would then rise from 281 to 311, or thirty extra women. This is an overestimate of the true potential of treatment, since the actual five year survival for women with advanced disease is lower than for women with early breast cancer.

Table 2.4 shows the number of deaths that could be prevented in New Zealand in a year by secondary prevention (mammographic screening), and tertiary prevention (ensuring the optimum treatment) for women who had been diagnosed with breast cancer when they were between fifty and sixty-four years old. To make this comparison the number of deaths that could be prevented by each of these strategies in a year has been calculated for women aged fifty-five to sixty-nine as the effects of both treatment and screening on breast cancer mortality are delayed. The comparison shows that mammographic screening has the greatest potential for reducing breast cancer mortality. Employing both strategies would have the greatest impact on breast cancer mortality.

2.3 Breast cancer screening: risks and benefits

For women over fifty mammographic screening is the most effective way to reduce breast cancer mortality. Even though its impact is relatively small (potentially preventing the deaths of fifty-eight of the six hundred women who die from breast cancer in New Zealand every year) it is the most effective of the three strategies to reduce breast cancer mortality. Obviously if an effective method of primary prevention, or a cure for breast cancer could be found there would be no need for screening. In the meantime it has the greatest potential to reduce breast cancer mortality in New Zealand women aged fifty and over. But breast cancer screening has risks as well as benefits. These risks and benefits are outlined in Table 2.5 (adapted from Chamberlain 1984). The benefits associated with screening will be examined first, in section 2.3.1, and then the risks of screening will be discussed in section 2.3.2.
TABLE 2.4

A comparison of the maximum numbers of breast cancer deaths that could be prevented each year in New Zealand, using different preventive strategies to reduce breast cancer mortality for women aged 50-64 at diagnosis.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Primary prevention</th>
<th>Secondary prevention (screening)</th>
<th>Tertiary prevention (treatment)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of breast cancer deaths that could be prevented each year *</td>
<td>not applicable</td>
<td>58</td>
<td>30</td>
</tr>
</tbody>
</table>

* this annual reduction in mortality would not begin until about five years after the introduction of the preventive strategy, as the effects of both treatment and screening on breast cancer mortality are delayed.
2.3.1 Benefits of breast cancer screening

The benefits of breast cancer screening depend upon the age-range of the women to whom it is offered. There is evidence that screening can reduce breast cancer mortality for women over fifty but no clear evidence for women under fifty (Chapter One, section 1.4). This is important, as there must be clear evidence of benefit before screening is offered:

We believe that there is an ethical difference between everyday medical practice and screening. If a patient asks a medical practitioner for help, the doctor does the best he [or she] can. He [or she] is not responsible for defects in medical knowledge. If, however, the practitioner initiates screening procedures he [or she] is in a very different situation. He [or she] should, in our view, have conclusive evidence that screening can alter the natural history of the disease in a significant proportion of those screened. (Cochrane and Holland 1971).

Screening should not be offered to women under fifty, as there is no clear evidence of benefit. Although mammography can detect breast cancer in younger women it is not clear that doing this actually reduces breast cancer mortality (Chapter One, section 1.4). For women over fifty there is evidence of benefit from screening. It is clear from the literature that breast cancer screening can reduce breast cancer mortality by about a third in a population of women aged fifty and over who are offered screening. The benefit is greater than this for women who take part in screening. Although the reduction in breast cancer mortality is about a third in the randomised controlled trials, this includes women who were offered screening but did not take part, and is therefore a diluted estimate of the effect of screening. For women who take part in screening the benefit has been estimated at forty to fifty percent (Day and Chamberlain 1988, Day 1991). So for every hundred women who are diagnosed with breast cancer in a screening programme about forty to fifty will have increased survival as a result. Some of these women will survive many years longer than they would have otherwise and some may be cured (Joensuu and Toikkanen 1995).

But it is important to note that this benefit is available only to the women who are diagnosed with breast cancer, and these women are quite a small proportion of the population offered screening. The risk of dying from breast cancer in the next twenty years for a New Zealand woman aged fifty is only about two percent, or one chance in fifty.
TABLE 2.5

Benefits and risks of breast cancer screening

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Risks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved prognosis for some women diagnosed by screening</td>
<td>Longer morbidity for women whose prognosis is unaltered</td>
</tr>
<tr>
<td>Less radical treatment which cures some women with early breast cancer</td>
<td>Over-treatment of questionable abnormalities</td>
</tr>
<tr>
<td>Resource savings</td>
<td>Resource costs</td>
</tr>
<tr>
<td>Reassurance for those with negative test results</td>
<td>False reassurance for women with false negative results</td>
</tr>
<tr>
<td></td>
<td>Anxiety and sometimes morbidity for women with false positive results</td>
</tr>
<tr>
<td></td>
<td>Hazard of screening test</td>
</tr>
</tbody>
</table>

Taking part in a breast cancer screening programme could, at the most, halve this risk (to one chance in a hundred). Although this absolute effect of screening is rather small, it is nevertheless the largest reduction in the risk of dying from breast cancer that is available to women aged fifty to sixty-four at present.

Another benefit of screening is the possibility of less radical treatment for women whose cancers are detected early, and it has been shown that the introduction of a breast cancer screening programme results in less invasive treatment of breast cancer (Moody et al 1994, de Koning et al 1990). Less invasive treatment is also associated with less psychological morbidity (Schain et al 1994).

There are economic issues that must also be considered. Is the cost of providing a screening programme excessive in relation to the benefit that it can produce? This also depends on the population being considered. Where women under fifty are included, the costs are extremely high (Carter et al 1993, Wright and Mueller 1995). But analyses of breast cancer screening programmes for women over fifty have found that the cost per year of life saved is not excessive, being within the realms of other interventions already provided by the health service (Forrest 1987, de Koning et al 1991, Hall et al 1992, Carter et al 1993). The costs of screening will differ depending on the country in which the programme is carried out, the structure of the health service, and the acceptance of screening in the eligible population. Some of the costs of setting up and running a screening programme may be offset later by a reduction in the costs of treating women with advanced breast cancer. In Sweden after five years of breast cancer screening it became cheaper to continue the screening programme than to discontinue it because the programme resulted in fewer women needing treatment for advanced disease (Tabar 1987). The economic aspects of breast cancer screening are discussed in detail in Chapter Seven.

The final benefit attributed to screening is reassurance for women with negative tests. Whether this can legitimately be claimed as a benefit is uncertain, since an invitation to screening may engender anxiety which is later relieved by a negative test (a net zero benefit). But some women are anxious about breast cancer even in the absence of a screening programme, and these women may feel less anxious after a negative test. It is important not to place too much emphasis on this benefit however, since breast cancer can
develop in the interval after a negative screen, and women should be aware of this so they do not delay seeking advice if they develop breast symptoms after a negative screen.

Given that screening should only be offered to women aged fifty and over it is important to determine whether the benefits available for women aged over fifty outweigh the possible risks associated with screening. As outlined above, the proportion of screened women who could benefit from screening is low, since most women do not develop breast cancer. In the Swedish Two-Counties study nineteen deaths from breast cancer were averted for every ten thousand women aged fifty to sixty-four screened (Chen et al 1995). Since many women are screened for every woman who benefits it is extremely important to minimise the risk of harm from being screened.

2.3.2 Risks of breast cancer screening

No screening test is perfect, and although most women will be correctly identified as having breast cancer or not, some women will have false positive or false negative tests. The women who have false negative tests may be wrongly reassured so that they present later than they would have otherwise. There is very little information on this, except that the number of women experiencing false negative tests is low (as would be expected, since the incidence of breast cancer is relatively low), at about five per ten thousand women screened (Shapiro et al 1982, Tabar et al 1985, Roberts et al 1989, Frisell et al 1991). It is possible to calculate the false negative rate for a given screening programme by subtracting the sensitivity for that programme from one. It is not certain how many women with false negative screening tests delay seeking advice, nor what effect such delay has.

The women who have false positive tests undergo unnecessary investigations with associated anxiety. It has been shown that women who have had routine breast cancer screening examinations do not experience increased psychiatric morbidity (Dean et al 1986, Ellman et al 1989, Baines et al 1990). But women requiring further assessment have significantly higher anxiety scores than do those attending for routine screening (Lerman et al 1991, Ellman et al 1989 Gram et al 1990). There is conflicting information on the delayed effects of assessment, with two studies finding no increase in anxiety, even for
women with false positive results (Ellman et al 1989, Gram et al 1990), while another study found that women who had false positive tests were more anxious about breast cancer afterwards (Bull and Campbell 1991). As well as the likelihood of increased anxiety, women with false positive tests suffer unnecessary investigations with associated morbidity. Obviously it is very important to keep this risk as low as possible in a screening programme.

But it is also important to recognise that some of the risks associated with screening are not new; women experience similar risks in the absence of a screening programme. Even without a screening programme many women are concerned about breast cancer and visit their general practitioners because of breast problems. In the Netherlands, without mass screening, nine percent of women aged over forty visited their general practitioners for breast assessment in the course of a year (de Koning et al 1990). Some of the women who visit their general practitioners are then referred on for further assessment. These women may have discovered a breast lump themselves, the general practitioner may have discovered a problem at routine breast examination, or they may be asymptomatic women who have had mammography in the absence of organised screening. Women who undergo assessment at surgical outpatients clinics, but who do not have breast cancer have been described as “the worried well” (Gravelle et al 1982) and in a study carried out in England before the introduction of the national breast cancer screening programme 4.7 percent of women aged forty and over fell into the “worried well” category in a year (Gravelle et al 1982).

These women experience the anxiety and morbidity resulting from unnecessary investigations in just the same way that women in the “false positive” group in a screening programme do, and the number of such women is not inconsiderable. Table 2.6 compares biopsy rates per thousand women before and after the introduction of a breast cancer screening programme in Australia (Hall et al 1992, Salkeld and Gerard 1994). It is clear from this table that in the first round of screening considerably more women were referred
TABLE 2.6

Comparison of assessment and biopsy rates before and after the introduction of a breast cancer screening programme.

<table>
<thead>
<tr>
<th></th>
<th>With screening programme</th>
<th>Without screening programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of women</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Referred for assessment</td>
<td>78</td>
<td>33</td>
</tr>
<tr>
<td>Excision biopsy</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Breast cancer detected</td>
<td>7</td>
<td>4</td>
</tr>
</tbody>
</table>


for assessment and the biopsy rate increased by fifty percent. But for the extra five biopsies per thousand women an extra three women were diagnosed with breast cancer. So for each woman diagnosed with breast cancer the number undergoing unnecessary biopsies was lower in an organised screening programme than without a screening programme.

Once the screening programme is established, the likelihood of false positive results and unnecessary biopsies falls. The first (prevalence) round of a screening programme results in more women being referred for assessment than the second and subsequent rounds because in the later rounds most of the women taking part have already been screened two years earlier (in an earlier round), and their previous films are available for comparison. The later rounds will largely detect cancers that have arisen since the previous screen. In the first screen there will be a higher prevalence of cancers, because some cancers may be detected at this screen that would have been detectable earlier, had there been a previous screen. In the Netherlands (de Koning et al 1990, de Koning et al 1995b) the biopsy rate in the second screening round was actually less than the biopsy rate before the breast screening programme was introduced (Table 2.7).

It is possible that radiation exposure during mammography could increase the risk of breast cancer; it is known that radiation exposure associated with the atomic bombs in Japan, and repeated x-ray treatment of women with tuberculosis increased breast cancer risk (Boice and Monson 1977, Tokunaga et al 1982, Adami et al 1990). But the level of radiation associated with mammography is much less. For a woman undergoing regular mammograms in a breast cancer screening programme the radiation risk is very small, and has been estimated at less than a hundredth of a woman's underlying risk of developing breast cancer (Gohagan et al 1986). Another risk of screening is discomfort associated with the screening test. Investigations of the level of this risk show that most women find mammography uncomfortable rather than painful (Stomper et al 1988, Jackson et al 1988, Wolosin 1989, Rutter et al 1992). This issue is examined further in Chapter Six, along with other aspects of the risk and acceptability of breast cancer screening.
### TABLE 2.7

Biopsy and breast cancer detection rates in the second round of screening compared with the rates before the introduction of screening in the Netherlands.

<table>
<thead>
<tr>
<th></th>
<th>With screening programme</th>
<th>Without screening programme</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of women</td>
<td>1000</td>
<td>1000</td>
</tr>
<tr>
<td>Referred for assessment</td>
<td>30</td>
<td>33</td>
</tr>
<tr>
<td>Excision biopsy</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Breast cancer detected</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>


2.3.3 Summary of benefits and risks

There is good evidence that screening can reduce mortality from breast cancer by about a third in women aged fifty and over. The evidence for women under fifty is not clear. In the absence of this evidence it would be unethical to offer screening to women under fifty. For women over fifty for whom a benefit has been shown, the risks of screening must also be considered. Screening is not unique in having risks, as most treatments have risks or side-effects, but the risks associated with screening can affect a larger number of women, since compared with treatment, screening is offered to a much larger group. The aim for any screening programme should be to keep the risks as low as possible without losing the benefits of screening. The best way to minimise risks and ensure that screening can reduce breast cancer mortality is to offer screening in a coordinated fashion, so that each stage of the screening process is of high quality. It is possible for a well-organised screening programme to reduce some of the risks associated with the investigation of breast problems, that would have occurred in the absence of screening (Table 2.5). The essential characteristics of an organised screening programme have been described for cervical screening programmes (Hakama et al 1985). These characteristics can be adapted for a breast cancer screening programme, since similar considerations apply. Table 2.8 shows the essential characteristics of an organised breast cancer screening programme (modified from Hakama et al 1985).

The essential characteristics of a properly organised programme enhance the possibility of benefit while reducing the risks associated with screening. The first three characteristics relate to identifying and inviting eligible women. The possible impact of a programme is greatest if all eligible women are invited to be screened. No matter how good a programme is, its impact will be poor if many of the eligible women do not receive invitations. An organised screening programme will have a policy on the appropriate age-range for screening, whereas in the absence of an organised programme women (for instance those aged less than fifty) may be screened inappropriately. The fourth and fifth characteristics relate to the facilities for screening, assessment and treatment. In particular, adequate facilities for assessment and treatment are essential, especially in the first screening round where the assessment and biopsy rates are higher than in subsequent rounds. If assessment facilities are inadequate women with positive tests will have to go on a waiting list. This
TABLE 2.8

Essential features of an organised breast cancer screening programme:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>The target population has been identified</td>
<td></td>
</tr>
<tr>
<td>The individual women are identifiable</td>
<td></td>
</tr>
<tr>
<td>Measures are available to guarantee high coverage and attendance, such as a personal letter of invitation</td>
<td></td>
</tr>
<tr>
<td>There are adequate facilities for carrying out the screening test</td>
<td></td>
</tr>
<tr>
<td>Adequate facilities exist for diagnosis and for appropriate treatment</td>
<td></td>
</tr>
<tr>
<td>There is a carefully designed and agreed referral system so that women are followed up appropriately</td>
<td></td>
</tr>
<tr>
<td>There is an organised quality control system</td>
<td></td>
</tr>
</tbody>
</table>

delay increases anxiety (Fentiman 1988, Marteau 1990, Smith et al 1991). The sixth essential for organised breast cancer screening is adequate follow up and coordination between those offering the screening test and those carrying out assessments. Without good coordination there is the risk that women with positive tests will not be followed up and referred for assessment. Organised screening programmes have specialist assessment clinics, which can reduce biopsy rates. In the Netherlands the biopsy rate after the introduction of an organised programme was less than the biopsy rate before the breast screening programme was introduced (de Koning et al 1990, de Koning et al 1995b).

With properly organised screening there is also less potential for waste. In the United States there are already more mammography machines than would be required to screen every woman in the United States according to the national screening guidelines. This means that many mammography facilities are under-utilised, and calls into question whether the staff taking and reading the films have adequate throughput to maintain their skills (Skolnick 1994). Lastly, it is also easier to build evaluation and quality control into organised screening programmes, so that the risks and benefits of screening can be monitored. Organised screening programmes, where women are identified and invited to take part tend to be offered to all women in a defined population. This type of screening has been described as “population-based”. Population-based screening will be discussed in the next section, with respect to the introduction of population-based screening in New Zealand.

2.4 Population-based screening

Population-based screening or mass screening is screening “pertaining to a general population defined by geopolitical boundaries, with this population being the denominator and/or the sampling frame” (Last, 1988). Screening is offered to all eligible people within a defined population. Eligibility may be determined by certain characteristics such as gender, and age, but everyone in the eligible population will then be offered screening. Ten principles of screening were developed by the World Health Organisation as guidelines to help determine whether population-based screening for a particular disease would be appropriate (Wilson and Jungner, 1968). The principles relate partly to the characteristics
of the population and the disease but also to the provision of an organised screening programme (as outlined in section 2.3.3). The ten principles of screening are listed in Table 2.9.

According to these principles, population-based breast cancer screening for women aged fifty to sixty-four was found to be an appropriate public health strategy for Britain (Breast Cancer Screening Working Group, 1987). But in making this assessment the British working group had information from the screening programmes that had been set up as part of the United Kingdom Trial of Early Detection of Breast Cancer. Although many of the British findings apply equally to New Zealand, the acceptability and the effectiveness of breast cancer screening, and its costs (principles six, seven, nine and ten) could not be addressed in New Zealand without first collecting local information.

2.5 The introduction of breast cancer screening in New Zealand

By 1987 the results from randomised controlled trials of breast screening were well known, many countries had local or regional population-based breast cancer screening programmes and Sweden, Finland and Iceland were developing national programmes. The British government had set up a working group to investigate the prospects for breast cancer screening (Breast Cancer Screening Working Group, 1987), had accepted its recommendations, and was planning to set up a national programme which would offer single-view mammography every three years to women aged fifty to sixty-four. At this time there were no population-based breast cancer screening programmes in New Zealand, although there were a few private radiology clinics offering screening mammography.

In response to these developments a working group was set up by the Cancer Society and the Department of Health to make recommendations about mammographic screening in New Zealand. The working group chaired by Professor David Skegg, reported in 1988 (Skegg et al, 1988). After reviewing the available strategies for reducing breast cancer
## TABLE 2.9

### Principles of screening

1. The condition should pose an important health problem
2. The natural history of the disease should be well understood
3. There should be a recognisable early stage
4. The treatment of the disease at an early stage should be of more benefit than treatment started at a later stage
5. There should be a suitable test
6. The test should be acceptable to the population
7. There should be adequate facilities for the diagnosis and treatment of abnormalities detected
8. Screening should be repeated at intervals determined by the natural history of the disease
9. The chance of physical or psychological harm to those screened should be less than the chance of benefit
10. The cost of a screening programme should be balanced against the benefit it provides

morbidity and mortality the group decided that breast screening was the most effective strategy available at the time. But the working group recognised that there were problems with introducing population-based screening in New Zealand. At the time of the report there were not enough skilled people in the country to organise and carry out a national screening programme. Neither were there adequate systems for inviting and recalling women.

We therefore recommend that decisions about routine screening should be postponed until at least two pilot studies have been completed. The purpose of these studies would not be to assess the efficacy of mammographic screening, which has already been established overseas. The purpose would be to determine whether screening can be introduced in New Zealand in a way that is effective (in terms of medical outcome), efficient (from an economic viewpoint), and acceptable to women. (Skegg et al. 1988)

Several recommendations about pilot programmes were made by the working group (Table 2.10). The Ministry of Health agreed to fund two pilot programmes according to the guidelines suggested in the Skegg report. The Cancer Society of New Zealand agreed to fund independent evaluations of the pilot programmes which would provide information about the acceptability, effectiveness, and economic efficiency of population-based mammographic screening in New Zealand. The Health Research Council of New Zealand provided scientific assessment of the proposals for pilot programmes together with their associated evaluation protocols, which were submitted from different areas in New Zealand. In December 1989, after the proposals submitted from various regions in New Zealand had been assessed, the proposal for a mammographic screening programme in Otago and Southland was accepted.

Otago and Southland are the southernmost regions in the South Island of New Zealand and their combined population is about 270,000 people. There are two main centres; Dunedin with a population of about 100,000 and Invercargill with a population of about 55,000. Large parts of Otago and Southland are rural. The population density of Otago and Southland is low by international standards, ranging from about 376 per square kilometre in Dunedin to under three per square kilometre in parts of Southland.
TABLE 2.10

**Recommendations for pilot programmes in New Zealand**

1. the eligible age-range should be 50-64
2. screening should be offered every two years
3. ways to achieve high participation should be explored
4. only dedicated mammography equipment should be used
5. screening should be by two-view mammography
6. the possibility of preliminary reading by non-radiologists should be investigated
7. the results of screening should be sent to women and their general practitioners, and any further assessment should be carried out at specialist assessment clinics
8. there should be adequate training for the staff of the pilot programmes
9. each pilot programme should be directed by one person, with managerial skills and knowledge of population-based screening
10. the pilot programmes should be carefully monitored
11. a cost-benefit analysis of the pilot programmes should be carried out
12. the acceptability of screening should be assessed

In the proposed programme women aged fifty to sixty-four and living in Otago and Southland were to be invited for screening with two-view mammography every two years. Three screening units were subsequently established; two fixed units based at Dunedin and Invercargill, and a mobile unit which travelled to rural areas. Assessment clinics were set up at Dunedin and Invercargill. The associated evaluation was designed to ensure that the programme would benefit the women who took part (since the benefit to risk ratio is highest in a well-organised and monitored programme), and to provide information about the acceptability, effectiveness, and economic efficiency of population-based mammographic screening in New Zealand. Screening at the fixed units in Dunedin and Invercargill started in September 1991. Shortly after this a second pilot programme was started in Waikato, to offer two yearly two-view mammography to women aged fifty to sixty-four, and to provide similar information about acceptability, effectiveness, and economic efficiency, and in particular to address the acceptability of breast cancer screening for Maori women (two percent of the eligible women in Otago and Southland were Maori whereas sixteen percent of the eligible women in Waikato were Maori).

2.6 Summary

This chapter has described the situation leading up to the introduction of population-based breast cancer screening in the form of pilot programmes in New Zealand. The evaluation of the Otago and Southland pilot programme was carried out in accordance with the recommendations of the Skegg report (Skegg et al. 1988), focussing on the acceptability, effectiveness, and economic efficiency of the pilot programme. Methods for evaluating pilot programmes are discussed in Chapter Three and Chapter Four describes the development of the evaluation protocol for the Otago and Southland pilot programme. The methods and results from the evaluation of the Otago and Southland pilot programme are reported in Chapters Five (the effectiveness of the pilot programme) and Six (the acceptability of the pilot programme), and these results, together with the results from the economic evaluation of the pilot programme, are summarised and discussed in Chapter Seven.
CHAPTER THREE

Deciding on the Evaluation Method

Introduction

The evaluation of the Otago and Southland pilot programme was carried out in accordance with the recommendations of the Skegg report (Skegg et al 1988), focussing on the acceptability, effectiveness, and economic efficiency of the pilot programme. The purpose of the evaluation was to find out whether population based breast cancer screening could be carried out in New Zealand in a way that would be effective, acceptable, and economically efficient. The results from the evaluation had to be available within a short time so that they could be used to help make decisions about a national breast cancer screening programme in New Zealand.

In this chapter techniques and methods for evaluating pilot programmes are discussed in section 3.1. The evaluation methods considered were experimental or quasi-experimental methods (a randomised controlled trial or a non-randomised comparison study), a case-control study, or the goal-attainment method of evaluation (evaluation using pre-set goals). In section 3.2 the methods are compared in terms of their ability to meet the requirements for the evaluation of the Otago and Southland pilot programme. Some key results from the randomised controlled trials of breast cancer screening are examined in section 3.3 to determine which of the trials should be used to develop targets against which the performance of the pilot programme could be measured.
3.1 Evaluation

Evaluation is the process that attempts to determine as systematically and objectively as possible the relevance, effectiveness, and impact of activities in the light of their objectives (Last 1988). A key concept in evaluation is that the activities being evaluated are compared against some standard of acceptability (Downie et al 1991).

Three approaches to the evaluation of medical care have been defined (Donabedian 1966); evaluation of outcomes, evaluation of processes, and evaluation of the settings in which medical care takes place. These can be linked to three commonly used evaluation methods; the experimental paradigm, which measures outputs, the goal attainment model, which focuses on processes, and the elucidation model which examine settings (Robinson 1984, Turner et al 1992). The experimental approach measures outcomes. This approach to evaluation has been the traditional approach and usually involved classical experimental design (for instance a randomised controlled trial) or a quasi-experimental design with a comparison group but without randomisation. The goal attainment model involves measuring the extent to which a programme attains goals which have been specified before the start of the programme. It is sometimes difficult to develop measurable goals for evaluation, and elucidation models of evaluation do not necessarily involve specific goals, but describe the way a programme is carried out.

There were several possible methods of evaluating the pilot breast cancer screening programme. These methods are discussed in this chapter. The possibilities included experimental or quasi-experimental methods (a randomised controlled trial, or a non-randomised comparison study), or a case-control study, and the goal-attainment method (evaluation using pre-set goals). The elucidation method was not considered, for two reasons. First, the focus of the elucidation method is not on measurement (Robinson 1984), but the evaluation was required to “monitor the process and outcome of mammography” including “the ratio of benign to malignant biopsies and assessment of the cancer detection rate” (Skegg et al 1988). Secondly, a major criticism of the elucidation method is that this approach rarely allows generalisation, (Robinson 1984), but one of the main purposes of the evaluation of the Otago and Southland programme was to collect information that could be used in a national breast cancer screening programme. Each of the
evaluation methods considered, with its strengths and limitations is discussed as an option for evaluating the Otago and Southland pilot programme.

3.1.1 Experimental method
The classical experimental method for evaluating breast cancer screening is a randomised controlled trial, with breast cancer mortality as the outcome measure. But the purpose of evaluating the pilot programme was not to find out whether screening could reduce mortality from breast cancer. This had already been shown in several randomised controlled trials (Chapter One). Rather, the purpose of the evaluation of the Otago and Southland pilot programme was to provide information about the acceptability, effectiveness, and economic efficiency of population based mammographic screening in New Zealand. The evaluation had to provide rapid results based on relatively small numbers of women. A randomised controlled trial would have been inappropriate in this context for two reasons. First, the efficacy of breast cancer screening had already been demonstrated. Secondly, the number of women and the period of follow up required would make it impossible to use this method of evaluation, given the size of the eligible population and the time available for evaluating the Otago and Southland pilot programme.

3.1.2 Quasi-experimental methods
It would have been possible to evaluate the pilot programme by comparing breast cancer mortality in the pilot area with a comparison area. But there are problems with this approach. Firstly any differences in breast cancer mortality between the two areas could be due to factors other than the presence or absence of the screening programme. Such factors might include differences in access to health care and the availability of appropriate treatment, and differences in the underlying risk of breast cancer in the two areas. This was addressed in reports from the UKTEDBC (United Kingdom Trial of Early Detection of Breast Cancer group 1988a, United Kingdom Trial of Early Detection of Breast Cancer Group 1988b). The authors of those reports stated that their results should be interpreted with greater caution than results from randomised controlled trials because the comparison was non-randomised, and this meant “it was not possible to adjust for all the factors other than the early detection interventions which may have contributed to differences in breast cancer mortality”. The second problem with this method of evaluation is that it requires large
numbers of women and several years of follow up in order to detect changes in breast cancer mortality that could be attributable to screening. The time available for the evaluation of the Otago and Southland pilot programme was too short to allow this approach.

3.1.3 Case-control study
Several case-control studies have been carried out to try to measure the effectiveness of established breast cancer screening programmes. These case-control studies, summarised in Chapter One, reported a reduced risk of death from breast cancer for women who had taken part in screening (Verbeek et al 1984, Verbeek et al 1985, Collette et al 1984, Collette et al 1992, Palli et al 1986, Palli et al 1989).

A case-control study might appear to be a better method for evaluating the pilot programme than the options discussed previously, since it could be carried out with smaller numbers and over a shorter time period. But there are limitations of using a case-control study as an evaluation method. The first limitation is selection bias. The women who take part in screening may have a different risk of developing or dying from breast cancer, compared with the women who do not take part. If women who are at low risk of dying from breast cancer are more likely to participate in screening, a case-control study would overestimate the protective effect of screening. (If high-risk women were less likely to accept screening then the controls would have had a greater exposure to screening even if screening actually had no effect on breast cancer mortality).

In the HIP study, women at higher risk took part in screening (Shapiro 1977, Shapiro et al 1982) while in the Swedish and UK randomised trials women at lower risk took part (Tabar et al 1985, Andersson et al 1988, Roberts et al 1989). A case-control study of the women offered screening in the HIP study using a sample of invited women as controls has been shown to underestimate the benefit of screening because of this self-selection of high-risk women (Freidman and Dubin 1991). It is impossible to adjust for this except in the context of a randomised controlled trial where there is a control group to compare with the non-attenders. In the HIP study the reason for the lower incidence of breast cancer among non-attenders could not be explained by differences in known risk factors such as age, parity, and history of benign breast disease between the attenders and the non-attenders (Weiss 1983). This means that it may be impossible to remove the effect of selection bias.
completely, even by adjusting for known risk factors, in a case-control study of breast cancer screening.

The effect of selection bias in case-control studies has also been demonstrated in two case-control analyses carried out within the UKTEDBC study (Moss 1991, Moss et al 1992). The UKTEDBC study was a non-randomised study which compared breast cancer mortality in districts where women had been offered mammography with breast cancer mortality in comparison districts where screening had not been offered (section 1.3.1). In the first case-control analysis carried out by Moss et al, the cases and controls were taken from two combined districts (a screening district and a comparison district). In the second case-control analysis the cases and controls were taken from the screening district only. The first analysis estimated the risk of death from breast cancer in women in the screening district relative to the control district (analogous to estimating the risk of death from breast cancer in the intervention group compared with the control group in a randomised controlled trial of breast cancer screening). The second analysis estimated the relative risk for women who had at some time been screened compared with women who had never been screened in the screening district alone. The second study was carried out to assess the extent of selection bias in a case-control study carried out in a population where all the women had been offered screening. The relative risk in the first study was 0.76 (95% CI 0.54-1.08), suggesting a twenty-four percent reduction in breast cancer mortality in the screening district compared with the comparison district. The relative risk in the second study was 0.51 (95% CI 0.27-0.98).

A cohort analysis of breast cancer incidence and mortality in the two districts was carried out which produced a relative risk of breast cancer death for the screening district compared with the comparison district of 0.79. The incidence of breast cancer in the comparison district was the same as that in non-attenders, but breast cancer mortality was higher in the non-attenders, suggesting that women at higher risk of dying from breast cancer had been less likely to take part in screening. The authors concluded that a case-control study carried out within a population that had been offered screening gave a biased estimate of the overall benefit of screening to the population.

The study by Moss et al suggested that, in order to reduce the effect of selection bias in a case-control study, it would be necessary to include information from a comparison district where screening had not been offered. But this approach is limited in the same way as the quasi-
experimental approach (section 3.1.2) by the problem that differences in breast cancer mortality between the two districts could be due to factors other than the presence or absence of the screening programme.

The second major limitation is that, although a case-control study could be carried out over a shorter time period than either a randomised controlled trial or a non-randomised trial of breast cancer screening, the time required would still have been considerable. Because breast cancer screening takes about five years to reduce breast cancer mortality (Rutqvist et al 1990), a case-control study to evaluate the effectiveness of the pilot programme could not be carried out earlier than five years from the start of the programme.

3.1.4 Goal-attainment method

Once randomised controlled trials had shown that breast cancer screening was efficacious the possibility arose that new programmes could be evaluated by comparing them with the screening programmes that were studied in the randomised controlled trials.

Another approach to assessment is to examine the process of care itself rather than its outcomes. This is justified by the assumption that one is interested not in the power of medical technology to achieve results, but in whether what is now known to be “good” medical care has been applied. Judgements are based on considerations such as the appropriateness, completeness and redundancy of information obtained through clinical history, physical examination and diagnostic tests; justification of diagnosis and therapy; technical competence in the performance of diagnostic and therapeutic procedures including surgery; evidence of preventive management in health and illness; coordination and continuity of care; acceptability of care to the recipient and so on. This approach requires that a great deal of attention be given to specifying the relevant dimensions, values and standards to be used in assessment. The estimates of quality that one obtains are less stable and less final than those that derive from the measurement of outcomes. They may, however, be more relevant to the question at hand: whether medicine is properly practiced. (Donabedian 1966).

It was suggested that early characteristics of programmes such as sensitivity and specificity, stage distribution, and detection rates could be used as surrogate endpoints for reduced breast cancer mortality. (Day 1989, Day et al 1989, Verbeek et al 1991). These surrogate endpoints could then
be used to set goals for the evaluation of a pilot programme. The advantages of a goal-attainment evaluation of the pilot programme are outlined below:

(1) breast cancer screening could be offered with adequate monitoring to ensure that the benefits demonstrated in randomised controlled trials were delivered by the pilot programme. Screening is unlike other medical interventions because it is offered to people who are well, with the understanding that they will benefit from taking part. This meant that breast cancer screening should not have been offered to women unless the programme was organised and evaluated in such a way that the benefits could be attained. The effectiveness of screening could be evaluated against a standard devised from the results of programmes that had been shown to reduce breast cancer mortality.

(2) the quality of the programme could be monitored so that the risks associated with screening could be minimised. This included ensuring that women received their results rapidly, and investigating ways to reduce the number of women recalled unnecessarily, thus minimising the anxiety related to positive tests.

(3) the choice of characteristics to be monitored could be tailored to the particular requirements of the evaluation of the pilot programme. For instance aspects other than the likely effect on mortality could be measured, and compared with the results from existing programmes. Such aspects required in the evaluation of the Otago and Southland pilot programme included information on the acceptability and the economic efficiency of the programme.

(4) the results from this method of evaluation were available rapidly and did not require years of follow up. This also meant that if corrective action is needed it could take place early in the operation of the pilot programme.
3.2 Comparing the possible methods of evaluation

Randomised controlled trials are essential for assessing the efficacy of screening. Once efficacy had been demonstrated it was inappropriate and also impractical, because of the numbers of women and the time required, to use this method to evaluate the effectiveness of pilot programmes. A non-randomised comparison was also rejected because factors other than the effect of the screening programme could affect the results. Also this approach could not provide rapid results. A case-control study was not used because of the potential for selection bias. This could not be adjusted for easily in a case-control study because it is impossible to measure the direction or magnitude of selection bias except in a randomised controlled trial. Unless the study included women who had been offered screening and women who had not been offered screening (eg: an appropriate comparison district as well as the screening district) selection bias would have been a problem. Also a case-control study could not show any effect on breast cancer mortality earlier than about five years from the start of the programme, since it takes this long before a reduction in breast cancer mortality is detectable (Rutqvist et al 1990). A goal-attainment evaluation using surrogate endpoints appeared to be the most practical and appropriate method for evaluating pilot programmes. Because of this, and recognising the limitations of other approaches, the evaluation of the Otago and Southland pilot programme used goals based on surrogate endpoints to predict the effect of the pilot programme on breast cancer mortality among eligible women. The evaluation also included the collection of information on acceptability and economic issues to supplement the goal-attainment method. Targets were selected based on the programmes which had been shown to be efficacious (programmes which had been evaluated in randomised controlled trials). The trials were examined in detail (section 3.6) to select goals (targets) that would suit the particular aims of the evaluation of the Otago and Southland pilot programme.

3.3 Comparing the randomised controlled trials of breast cancer screening. Which trials should be used for the development of targets?

Breast cancer screening programmes that were part of randomised controlled trials of screening could provide very useful information for designing an evaluation using surrogate markers. Once a programme had been shown in a randomised controlled trial to reduce breast cancer mortality, then
that programme could be used as a gold standard for other programmes. But it was essential that
the programmes were sufficiently alike in their design and execution for it to be realistic to compare
them in order to develop targets for the pilot programme. The randomised trials were examined in
detail to find out whether it would be appropriate to include them all in developing targets for the
evaluation, whether any adjustments needed to be made when comparing the trials, and whether
more emphasis should have been placed on the results from some trials than others.

Table 3.1 gives some characteristics of the randomised controlled trials. The Canadian trial is not
included in this table because this trial differed from the others in its design which was to test the
efficacy of screening rather than the efficacy of the offer of screening.

3.3.1 Exclusion criteria in the trials
There were differences in the exclusion criteria used in the randomised controlled trials. In the
Swedish Two-Counties trial the total population of women aged forty and over was entered into
the trial and no exclusions were documented, but most analyses have been restricted to women
with breast cancer diagnosed after the date of randomisation. In the Malmö trial women with
previously diagnosed breast cancer were included and staging was carried out according to the
stage of the cancer diagnosed during the trial, irrespective of any previous diagnosis of breast
cancer. In the Stockholm trial women with breast cancer diagnosed after randomisation but before
screening were analysed separately. No information is given about women with breast cancer
diagnosed prior to randomisation. In the Swedish overview all women with breast cancer
diagnosed before the date of randomisation were excluded.

Women who already have breast cancer may have a different prognosis from other women who
take part in screening. Provided randomisation has evenly distributed these women between study
and control groups, including or excluding them should not affect the results. In the Malmö study
women with previously diagnosed breast cancer were included but separate analyses which
excluded these women did not alter the results (Andersson et al 1988).
### TABLE 3.1

Summary of Randomised controlled trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>HIP</th>
<th>S2C</th>
<th>MAL</th>
<th>EDB</th>
<th>STK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Randomisation</strong></td>
<td>Individual</td>
<td>Cluster</td>
<td>Individual</td>
<td>Cluster</td>
<td>Day of birth</td>
</tr>
<tr>
<td><strong>Age range</strong></td>
<td>40-64</td>
<td>40-74</td>
<td>45-69</td>
<td>45-64</td>
<td>40-64</td>
</tr>
<tr>
<td><strong>Study group</strong></td>
<td>31000</td>
<td>77000</td>
<td>21000</td>
<td>23000</td>
<td>40000</td>
</tr>
<tr>
<td><strong>Control group</strong></td>
<td>31000</td>
<td>56000</td>
<td>21000</td>
<td>23000</td>
<td>20000</td>
</tr>
<tr>
<td><strong>Screening method</strong></td>
<td>2v M P</td>
<td>1v M</td>
<td>2v M</td>
<td>2v M P</td>
<td>1v M</td>
</tr>
<tr>
<td><strong>Screening interval (months)</strong></td>
<td>12</td>
<td>24 - 36*</td>
<td>21</td>
<td>12</td>
<td>28</td>
</tr>
<tr>
<td><strong>1st screen sensitivity (%)</strong></td>
<td>82</td>
<td>96</td>
<td>78</td>
<td>92</td>
<td>86</td>
</tr>
<tr>
<td><strong>1st round participation (%)</strong></td>
<td>65</td>
<td>89</td>
<td>74</td>
<td>61</td>
<td>81</td>
</tr>
</tbody>
</table>

(continues on next page)
<table>
<thead>
<tr>
<th>Trial</th>
<th>HIP</th>
<th>S2C</th>
<th>MAL</th>
<th>EDB</th>
<th>STK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of follow up</td>
<td>18</td>
<td>12</td>
<td>12</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>% reduction in BC mortality</td>
<td>23</td>
<td>31</td>
<td>19</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>95% CI</td>
<td>13-39</td>
<td>16-43</td>
<td>'7-38</td>
<td>'12-37</td>
<td>'22-47</td>
</tr>
<tr>
<td>Women with previous breast cancer included in trial</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Non-attenders invited to later screens</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Defined surgical protocol</td>
<td>no</td>
<td>no</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Controls invited to be screened</td>
<td>no</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Key: 2v M = two view mammography  P = physical examination

* the screening interval was 24 months for women under 50 and 36 months for women over 50
3.3.2 Definition of mortality outcomes in the trials

Different end points for breast cancer mortality were used in the randomised controlled trials (Table 3.2). Some trials defined death from breast cancer as “death with breast cancer as the primary cause given on the death certificate” while others used less rigorous definitions such as “death with breast cancer as the underlying cause” or “death with breast cancer present at death”. If death with breast cancer present at death is used as the definition, many more women could be included (for instance women who died in motor vehicle accidents or women who died prematurely from other diseases although they had also been diagnosed with breast cancer). The deaths of these women would not have been influenced by screening, but there could be more such women in the intervention group because screening diagnosed breast cancer earlier; (women in the control group who died from other causes may also have had breast cancer but not had it diagnosed because they had not been screened). Also if the less rigorous definition is used it is likely that larger numbers of women would be classified as having died from breast cancer. This could result, given the same number of subjects, in these trials producing results at a greater level of significance than if the more rigorous definition had been used. This problem was addressed in the Swedish overview (Nystrom et al 1993) and in a later paper (Nystrom et al 1995), where an independent review committee carried out a blinded assessment of mortality. Two definitions were used; “breast cancer present at death” and “breast cancer as the underlying cause of death”. The results for “breast cancer as the primary cause of death” are not reported. Predictably, use of “breast cancer present at death” produced more deaths but the relative risks and ninety-five percent confidence intervals were similar with a relative risk of 0.78 (95% CI 0.68 to 0.89) for “breast cancer present at death” and 0.77 (95% CI 0.67 to 0.88) for “breast cancer as the underlying cause of death” (Nystrom et al 1993).

3.3.3 Comparing the HIP trial with subsequent randomised trials of breast cancer screening

The HIP trial differs from subsequent trials of breast cancer screening, in particular in its results for women under fifty. It is recognised that the quality of mammography in the early 1960’s in the HIP trial was probably not as good as that in later trials (Smart 1994); but despite this the reduction in mortality found in the HIP trial was as good as or better than that found in the other trials. A possible explanation for this is that, in the absence of screening, the clinical diagnosis of breast cancer at the time of the HIP trial occurred later than it did when the other trials were carried out.
This could mean that even inferior mammography would be capable of detecting cancer much earlier than it was generally detected during the 1960's. There is evidence for this, since the distribution of stage at diagnosis changed markedly from 1940 to 1979. By the 1970's a higher percentage of breast cancers was detected in a localised stage even though mammography was not widely available. Table 3.3 shows the changing percentage of breast cancers registered in a localised stage with the Connecticut Cancer Registry from 1940 to 1979 (Roush et al 1987).

It is consistent with this explanation that although the age-adjusted incidence of breast cancer has increased by fifty percent from 1965 to 1975 in America, the age-adjusted mortality has remained almost unchanged since 1960 (Fox 1979). Age-specific mortality rates for fifty to sixty-nine year old women have remained constant and those for forty to forty-nine year old women have declined since the mid-1960's even though the age-specific incidence increased during this time (Devesa et al 1987, Avila and Walker 1987).

The stage distribution at diagnosis reported for control women in the HIP study is similar to that for control women in later trials, although forty-six percent of control women in the HIP trial had node negative tumours compared with fifty-four percent of control women in the Swedish Two-Counties trial (Shapiro et al 1988, Tabar et al 1992). Larger differences might have been expected, but the validity of such comparisons has been questioned because of changes in stage classification and also because staging in the HIP study was based on retrospective analysis of hospital records which is known to be inaccurate (Forrest 1990). Also it has been pointed out by Peto and Easton (Peto and Easton 1989) that diagnostic improvements tend to alter staging over time, so that a tumour diagnosed twenty years ago might be allocated a higher stage if it were diagnosed now, because it is easier to detect smaller tumours and micrometastases. So the tumours diagnosed in the HIP study could have been more advanced even though the reported stage distribution is similar to that for the tumours diagnosed in later trials.
TABLE 3.2

Mortality end-points used in the randomised controlled trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>End point</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAL</td>
<td>death with breast cancer as the primary cause of death</td>
</tr>
<tr>
<td>CAN</td>
<td></td>
</tr>
<tr>
<td>HIP</td>
<td>death with breast cancer as the underlying cause of death</td>
</tr>
<tr>
<td>EDB</td>
<td></td>
</tr>
<tr>
<td>S2C</td>
<td>death with breast cancer present at death</td>
</tr>
<tr>
<td>STK</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3.3

Percentage of Cancers in Localised Stage
Connecticut Cancer Registry, 1940 to 1979

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All ages</td>
<td>37.7</td>
<td>41.8</td>
<td>48.6</td>
<td>51.0</td>
</tr>
<tr>
<td>55-64</td>
<td>36.4</td>
<td>39.8</td>
<td>46.5</td>
<td>48.2</td>
</tr>
</tbody>
</table>

A similar explanation has been given for the low mortality from breast cancer among Swedish women compared with other women in Europe. Although the age adjusted breast cancer incidence in Sweden is similar to that in Denmark, the Netherlands, and the UK (about sixty per hundred thousand per year) mortality is considerably lower (seventeen per hundred thousand per year compared with about twenty-eight per hundred thousand per year). The stage-distribution of breast cancer at diagnosis is different in Sweden, where women seem to present earlier and there are fewer women diagnosed with stage three and stage four cancers (Gad et al 1984).

Later detection of breast cancer in the United States at the time of the HIP trial may have affected the results for women under fifty in particular. No randomised trial of breast cancer screening has provided convincing evidence that screening is beneficial for women under fifty. The HIP trial is the only one to have shown an effect even approaching the magnitude of that seen for women over fifty. But in the HIP trial, the effect of screening on breast cancer mortality in women under fifty may have been mostly due to clinical examination (table 3.4). This effect would not be seen in later trials if breast cancer was by this time being diagnosed earlier (either due to greater awareness leading to earlier presentation or to better diagnosis) even in unscreened women. Table 3.4 shows the mode of detection of histologically confirmed cancers in the HIP trial (Shapiro et al 1988) compared with the mode of detection for cancers in the Edinburgh and Guildford screening programmes shown in Table 3.5 (UK Trial of Early Detection of Breast Cancer Group 1988b). It is clear that the proportion of breast cancers detected by mammography had increased in the later programmes compared with the HIP trial.

The only trial to have shown an appreciable effect on breast cancer mortality for women under 50 is the HIP trial. However the HIP trial was carried out a decade earlier than subsequent trials. The HIP results may be due to the relatively delayed diagnosis of breast cancer that was occurring in the United States in the 1960’s (Table 3.3), so that the lead time achieved by screening younger women was sufficient to make an impact on mortality. Earlier diagnosis since then may mean that the lead time now produced by screening is no longer sufficient to reduce breast cancer mortality in younger women.
**TABLE 3.4**

Health Insurance Plan trial histologically confirmed breast cancers

<table>
<thead>
<tr>
<th>Age at entry</th>
<th>Number of breast cancers</th>
<th>Clinical only</th>
<th>Mammography only</th>
<th>Clinical and Mammography</th>
</tr>
</thead>
<tbody>
<tr>
<td>40-49</td>
<td>40.0</td>
<td>57.5</td>
<td>25.0</td>
<td>17.5</td>
</tr>
<tr>
<td>50-59</td>
<td>67.0</td>
<td>40.3</td>
<td>38.8</td>
<td>20.9</td>
</tr>
<tr>
<td>60-64</td>
<td>25.0</td>
<td>36.0</td>
<td>32.0</td>
<td>32.0</td>
</tr>
</tbody>
</table>

### TABLE 3.5

UK Screening Programme First screen (mammography and clinical examination)

<table>
<thead>
<tr>
<th>Centre</th>
<th>Number of breast cancers</th>
<th>Clinical only</th>
<th>Mammography only</th>
<th>Clinical and Mammography</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edinburgh</td>
<td>81.0</td>
<td>6.0</td>
<td>32.0</td>
<td>62.0</td>
</tr>
<tr>
<td>Guildford</td>
<td>99.0</td>
<td>7.0</td>
<td>27.0</td>
<td>63.0</td>
</tr>
</tbody>
</table>

3.3.4 The effect of self-selection for screening on comparisons of randomised controlled trials

A randomised controlled trial with breast cancer mortality as the outcome measure is the best way to measure the efficacy of breast cancer screening, because the results will not be affected by bias. However when randomised controlled trials are compared with one another (for instance when setting targets for screening programmes), bias can be introduced. For example, in the Health Insurance Plan (HIP) trial the incidence of breast cancer in unscreened study group women was 1.45 per thousand, while in control group women it was 1.87 per thousand (Shapiro 1977, Shapiro et al 1982). It has been suggested that women with a higher risk of cancer tended to self-select themselves for screening (Shapiro 1977, Shapiro et al 1982). A trial where the screened women have a higher underlying risk of breast cancer is likely to detect more early cancers and produce a greater mortality reduction than a trial which had a similar participation rate but no self-selection of high-risk women. Self-selection of high risk women did not occur in the other randomised trials of breast cancer screening, in fact in some of the trials the opposite occurred and there was self-selection of low-risk women (Shapiro 1977, Shapiro et al 1982, United Kingdom Trial of Early Detection of Breast Cancer Group 1988a, Roberts et al 1989, Frisell et al 1986, Frisell et al 1991, Andersson et al 1988). This means that comparisons of the trials will be misleading because other trials with similar participation rates could not expect to match the results of the HIP trial unless they too had self-selection of high risk women as attenders. For instance the Edinburgh trial (Roberts et al 1989), with similar participation, sensitivity, and underlying breast cancer incidence showed self-selection that was in the opposite direction to the HIP trial (the non-attenders in the Edinburgh trial were at higher risk of breast cancer), and this may be one of the reasons that the Edinburgh trial did not achieve a comparable reduction in breast cancer mortality.

Before comparing randomised controlled trials of breast cancer screening it would be useful to be able to adjust for differences in self-selection between the trials. An “adjusted participation rate” (the expected participation rate in the absence of self-selection) should be used. A method of calculating an adjusted participation rate was developed by the author and is described below.

By comparing breast cancer mortality in the non-attenders and the controls the expected breast cancer mortality in screened women (in the absence of screening) can be calculated. In the absence of screening, because the two groups were randomly assigned, breast cancer mortality would have
been similar in the two groups. Since breast cancer mortality in the control group and in non-attenders is known, it is possible to calculate what the breast cancer mortality would have been among attenders, in the absence of screening.

If there is no self-selection \( m(\text{na}) = m(\text{ctl}) \) where \( m = \) breast cancer mortality
\( \text{na} = \) non-attender
\( \text{ctl} = \) control group

If there is self-selection of high-risk women, then \( m(\text{na}) < m(\text{ctl}) \)

If there is self-selection of low-risk women, then \( m(\text{na}) > m(\text{ctl}) \)

If there had been no screening programme, the expected breast cancer mortality in the intervention group would have been the same as in the control group (since the groups were produced by randomisation). If \( m(\text{ctl}) \) is the number of deaths per thousand control group women, then the expected number of deaths from breast cancer among a thousand intervention group women (in the absence of screening) should be the same as in a thousand control group women.

\[ m(\text{ctl}) = \hat{m}(\text{int}) \] where \( \text{int} = \) intervention group
\[ \hat{m} = \text{expected mortality} \]

The intervention group is comprised of women who attended screening and women who did not, so the number of deaths in the intervention group is the sum of breast cancer deaths in attenders plus non-attenders:

\[ \hat{m}(\text{int}) = \hat{m}(\text{a}) + m(\text{na}) \]
\( \text{a} = \) attenders
\( \hat{m}(\text{int}) \) cannot be observed, but can be estimated from \( m(\text{ctl}) \).

\[
\hat{m}(\text{int}) = m(\text{ctl})
\]

Since allocation to the groups was at random, \( m(\text{ctl}) \) and \( \hat{m}(\text{int}) \) are the same:

\[
m(\text{ctl}) = \hat{m}(a) + m(\text{na})
\]

and therefore the expected number deaths among the attenders (in the absence of screening) is:

\[
\hat{m}(a) = m(\text{ctl}) - m(\text{na})
\]

The participation rate, adjusted for selection, is then:

\[
\frac{\hat{m}(a)}{m(\text{ctl})} \times 100\%
\]

For example, in the HIP trial the cumulative mortality rate (deaths per thousand women entered) was 2.7 in non-attenders and 4.3 in controls. Thirty-five percent of women in the intervention group did not attend screening. There was self-selection of high-risk women for screening in the HIP trial, since \( m(\text{na}) < m(\text{ctl}) \).

For every two thousand women entered in the HIP trial there were a thousand women in the control group with 4.3 breast cancer deaths, and a thousand women in the intervention group, of whom 350 were non-attenders with \( 350 \times 2.7 \) per thousand = 0.95 breast cancer deaths.
The expected number of breast cancer deaths per thousand among attenders (in the absence of screening) was:

\[ \hat{m}(a) = m(ctl) - m(na). \]

\[ = 4.3 - 0.95 \]

\[ = 3.35 \]

The adjusted participation rate is:

\[ \frac{\hat{m}(a)}{m(ctl)} \times \frac{100\%}{1} \]

\[ = \frac{3.35}{4.3} \times \frac{100\%}{1} \]

\[ = 78\% \]

It is clear that women at higher risk of dying from breast cancer were more likely to attend screening, since the thirty-five percent who did not attend screening experienced only twenty-two percent of the expected breast cancer deaths in the intervention group. In the absence of screening, seventy-eight percent of the deaths would have occurred among the other sixty-five percent of the intervention group.

In the example from the HIP trial, because of the self-selection of women with a higher underlying risk of death from breast cancer, it is likely that the HIP screening programme would have had a higher breast cancer detection rate than programmes which had similar participation rates but no self-selection of high-risk women. The HIP results are equivalent to what would be expected if the screening programme had a seventy-eight participation rate but no self-selection of high-risk women, so the "adjusted participation rate" for the HIP trial is seventy-eight percent. Table 3.6 shows adjusted participation rates for other breast cancer screening trials, calculated by the same method. Multiplying the participation rate by the sensitivity for each programme gives the "maximum breast cancer detection" which is the maximum percentage of cancers in the eligible women that it would be possible for each programme to detect (Table 3.7). Sensitivity is here
### TABLE 3.6

Adjustment for self-selection in randomised controlled trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Cumulative breast cancer mortality (Deaths /1000 women entered)</th>
<th>Screening Participation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-attenders</td>
<td>Controls</td>
</tr>
<tr>
<td>HIP</td>
<td>2.7</td>
<td>4.3</td>
</tr>
<tr>
<td>Stockholm</td>
<td>1.8</td>
<td>1.5</td>
</tr>
<tr>
<td>Swedish Two-Counties</td>
<td>4.5</td>
<td>2.1</td>
</tr>
<tr>
<td>Edinburgh</td>
<td>3.2</td>
<td>2.8</td>
</tr>
<tr>
<td>Malmö</td>
<td>5.6</td>
<td>3.1</td>
</tr>
</tbody>
</table>
TABLE 3.7

Maximum breast cancer detection for each trial

<table>
<thead>
<tr>
<th>Trial</th>
<th>Adjusted Participation</th>
<th>Sensitivity*</th>
<th>Maximum Breast Cancer Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIP</td>
<td>78%</td>
<td>82%</td>
<td>64%</td>
</tr>
<tr>
<td>STK</td>
<td>77%</td>
<td>86%</td>
<td>66%</td>
</tr>
<tr>
<td>S2C</td>
<td>77%</td>
<td>96%</td>
<td>74%</td>
</tr>
<tr>
<td>EDB</td>
<td>55%</td>
<td>92%</td>
<td>51%</td>
</tr>
<tr>
<td>MAL</td>
<td>53%</td>
<td>78%</td>
<td>41%</td>
</tr>
</tbody>
</table>

* Sensitivity is defined as screen-detected cancers divided by the sum of screen-detected cancers plus interval cancers in the twelve months after screening.
defined as screen-detected cancers divided by the sum of screen-detected cancers plus interval cancers in the twelve months after screening. The maximum breast cancer detection should correlate with the mortality reduction achieved by the programme. The maximum breast cancer detection calculated using the adjusted rather than the unadjusted participation rate, correlates more closely with the reduction in breast cancer mortality for each trial (Table 3.8).

When using the results from randomised trials to set targets for breast cancer screening programmes it is important to use an adjusted participation rate because otherwise it could seem (from the unadjusted HIP results) that a thirty percent reduction in breast cancer mortality in the eligible population is obtainable even if only sixty-five percent of the population participate in screening. In fact this would only be the case if self-selection of high risk women occurred in all screening programmes and it is clear from other programmes that this is not so.

3.4 Summary

The published literature about mammographic breast cancer screening has been reviewed in Chapter One, and the background to and requirements for the evaluation of the Otago and Southland pilot programme were described in Chapter Two. Possible methods of evaluating a pilot programme have been discussed in this chapter. The randomised controlled trials have also been compared in detail in this chapter to establish whether it is appropriate to consider all the trials when developing targets for the pilot programme, whether any adjustments should be made when comparing the trials, and whether more emphasis should be placed on the results from some trials than others.

Although the trials differed in their endpoints for breast cancer mortality and in the criteria for excluding women, adjusting for the differences does not affect the results appreciably. This suggests that despite differences in endpoints and exclusion criteria the various trials can be compared and may be used to develop targets for the pilot programme. It is likely that in the HIP trial the time of diagnosis of breast cancer was brought forward relative to the usual time of diagnosis in the United States in the 1960’s, which was probably later than the usual time of diagnosis by the time the subsequent trials were underway. This may partly explain the success of
TABLE 3.8

Breast cancer detection and mortality reduction in randomised controlled trials

<table>
<thead>
<tr>
<th>Trial</th>
<th>Mortality Reduction (%)</th>
<th>Breast Cancer Detection (%)</th>
<th>Adjusted Breast Cancer Detection (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>S2C</td>
<td>31</td>
<td>85</td>
<td>74</td>
</tr>
<tr>
<td>HIP</td>
<td>29</td>
<td>53</td>
<td>64</td>
</tr>
<tr>
<td>STK</td>
<td>24</td>
<td>70</td>
<td>66</td>
</tr>
<tr>
<td>EDB</td>
<td>17</td>
<td>56</td>
<td>51</td>
</tr>
<tr>
<td>MAL</td>
<td>4</td>
<td>58</td>
<td>41</td>
</tr>
</tbody>
</table>
the HIP trial, and also explains some of the differences between the HIP and the other trials. Because of this, less emphasis should be placed on the HIP trial relative to the later trials, especially when developing targets for the pilot screening programme.

When setting a target for participation based on the trials it was important to adjust for the self-selection that occurred among the participants in all the trials, so that the target for participation in the pilot programme was not inadvertently set too low. It was recognised that self-selection could occur in the pilot programme, and that it would not be measurable, but by recognising this possibility, and considering the lowest and highest levels of participation found in the trials, an appropriate target was set that would still allow for some self-selection in either direction.

Chapter Four describes the derivation of appropriate targets for the evaluation of the Otago and Southland pilot programme. The targets based on the randomised controlled trials of breast cancer screening were chosen to provide a standard against which the performance of the Otago and Southland pilot programme could be evaluated. The targets related to the three key areas of effectiveness, acceptability, and economic efficiency that had been specified for the evaluation (Skegg et al 1988) and these targets are described in Chapter Four, with an explanation for the choice of each target.
CHAPTER FOUR

The Derivation of Targets for the Evaluation

Introduction

The Otago and Southland pilot breast cancer screening programme was established in 1991 (Elwood et al 1991). In this programme two-view mammography is offered every two years to an eligible population of about nineteen thousand women aged between fifty and sixty-four and living in Otago and Southland. The programme has two fixed screening centres at Dunedin and Invercargill hospitals, a mobile screening unit that travels to rural areas, and two specialist assessment clinics (at Dunedin and Invercargill) for women who are referred for further investigations after being screened.

An evaluation of the pilot programme was designed as recommended by the Skegg report (Skegg et al 1988) to provide information about the acceptability, effectiveness, and economic efficiency of breast cancer screening in New Zealand. The evaluation measured the performance of the pilot programme against targets based on the results of successful programmes. The results from the evaluation were regularly reported to the programme staff, and were used in the implementation and management of the pilot programme. The evaluation was designed to provide information about the acceptability, effectiveness, and economic efficiency of population-based breast cancer screening in New Zealand as specified in the Skegg report (Skegg et al 1988). The evaluation protocol was approved by both the Otago Area Health Board and Southland Area Health Board ethics committees.

To try to measure the effect of the pilot programme on breast cancer mortality directly would have taken many years and would have been hampered by small numbers and the lack of a suitable comparison group, as discussed in Chapter Three. Instead it was decided that surrogate measures of breast cancer mortality would be used as targets for a goal-attainment evaluation method. The
evaluation was designed to compare early characteristics of the pilot programme with those of established programmes which had been shown to reduce breast cancer mortality. Results from five randomised controlled trials of screening (Shapiro 1977, Shapiro et al 1982, Tabar and Gad 1981, Tabar et al 1985, Andersson et al 1988, Roberts et al 1989, Frisell et al 1986, Rutqvist et al 1990) were used in this comparison. The Swedish Two-Counties and Health Insurance Plan trials showed significant reductions in breast cancer mortality, while the other three trials showed smaller, non-significant reductions. The aim was for the pilot programme to share the early characteristics of the programmes which had been shown (in randomised controlled trials) to reduce breast cancer mortality in the longer term.

Information that could be used for quality control was also collected. For instance although the specificity of the screening test would not directly affect breast cancer mortality, a small drop in specificity would adversely affect a large number of women (section 4.1.2). At worst, it would be possible for a programme with extremely low specificity to be effective in reducing breast cancer mortality but be unsustainable because of the cost to women and to the programme, of unnecessary investigations.

The evaluation method used targets based on relevant characteristics of the prevalence screen of programmes that had been shown to be effective and efficient in the longer term. Both quality control and the effect of the programme on breast cancer mortality were addressed. If the pilot programme met the targets it would be assumed to be operating as well as successful programmes were at the same stage. Targets were developed for the first screening round (prevalence screen) of the pilot programme (the development of targets for subsequent screening rounds is discussed in Chapter Seven). The targets are described within the three specified evaluation areas of effectiveness, acceptability, and economic efficiency:

Effectiveness

The development of targets to measure the effectiveness of the pilot programme is described in section 4.1. The effectiveness targets included targets for the identification and invitation of eligible women (section 4.1.1), screen sensitivity and specificity (section 4.1.2), the referral rate and the benign to malignant biopsy rate (section 4.1.3), and the ratio of cancers detected at the prevalence screen to that expected in the absence of screening (section 4.1.4). Comparisons between the pilot
programme and published data on tumour size, grade, and stage of prevalence screen cancers were also made. The results of the evaluation of the effectiveness of the pilot programme are described in Chapter Five.

Acceptability
Methods for evaluating the acceptability of screening were also developed (section 4.2). A target participation rate (section 4.2.1) was used to assess the acceptability of screening. Several surveys were conducted to measure acceptability (section 4.2.2), including a pre-screening survey, and surveys of women who took part in screening, and who attended the assessment clinics. Women who declined an invitation to take part in the programme were also surveyed. These surveys are described briefly in this chapter and in detail in Chapter Six.

Economic efficiency
There were two issues to be addressed in the economic evaluation; first whether the pilot programme could provide screening in a way that would not be "wasteful of scarce resources" (Skegg et al 1988), and second to be able to extrapolate from the pilot programmes so that an estimate of the cost of a national breast cancer screening programme could be made. The evaluation of the economic efficiency of the pilot programme is described and the results discussed in detail in Chapter Seven.

The methods of data collection used in the evaluation are described briefly in section 4.4. More detailed methodology is given in Chapters Five and Six. The relevant characteristics of the reference breast cancer screening programmes are shown in Table 4.1 with the targets set for the pilot programme. The reasons for choosing these targets for the pilot programme are discussed in detail below (sections 4.1 to 4.3).
4.1 Measuring the effectiveness of the pilot programme

4.1.1 Identification of eligible women

The aim of a screening programme is to reduce the impact of breast cancer on the eligible population. To achieve this as many eligible women as possible must be identified and invited to take part in the programme. In the Health Insurance Plan (HIP) trial all eligible women could be identified from the HIP insurance records (Shapiro 1977, Shapiro et al 1982). In Sweden and the Netherlands eligible women were identified from population registers which were extremely accurate (Tabar and Gad 1981, Tabar et al 1985, Andersson et al 1988). In Edinburgh eligible women were identified through general practice registers, but these were found to be inaccurate with the eligible population overestimated despite some eligible women not being included (Roberts et al 1989, Bowling and Jacobsen 1989).

In New Zealand the electoral roll is the closest approximation to a population register. It is compulsory for all New Zealand citizens to register on the electoral roll once they reach the age of eighteen. Maori have the option of registering on either the general electoral roll or the Maori electoral roll. In 1991 it became possible to obtain lists of voters within certain age-groups from the electoral roll centre, for health-related purposes. The actual ages or dates of birth of voters are not released, but a lists of names of voters within designated age-ranges can be obtained for each electorate. The electoral roll centre does not separate voters according to gender, so any lists obtained include males and females. The electoral roll is estimated to have only about ninety-three percent coverage (Department of Statistics 1984) which means that not all eligible women could be identified and invited to be screened using this source alone. Some women who are not on the electoral roll could be identified and invited through general practice age-sex registers. The target set for the pilot programme was to identify at least ninety percent of the eligible women. The target was to be measured by comparing the number of eligible women in the census with the number who could be identified, with names and addresses, using a combination of the electoral roll and general practice age-sex registers.

The participation rate was defined as the proportion of invited women who were screened as a result of the invitation. In order to approach or exceed the lowest recorded participation rate for successful programmes, the pilot programme had a target participation rate of seventy-eight
TABLE 4.1

Characteristics of the first screening round of established mammographic screening programmes and targets for the pilot programme

<table>
<thead>
<tr>
<th>Characteristic to be evaluated</th>
<th>Programme</th>
<th>HIP</th>
<th>S2C</th>
<th>STK</th>
<th>MAL</th>
<th>EDB</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of eligible women</td>
<td></td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>90%</td>
<td>90%</td>
</tr>
<tr>
<td>Participation *</td>
<td></td>
<td>78%</td>
<td>77%</td>
<td>77%</td>
<td>53%</td>
<td>55%</td>
<td>78%</td>
</tr>
<tr>
<td>Coverage</td>
<td></td>
<td>78%</td>
<td>77%</td>
<td>77%</td>
<td>53%</td>
<td>50%</td>
<td>70%</td>
</tr>
<tr>
<td>Sensitivity #</td>
<td></td>
<td>82%</td>
<td>96%</td>
<td>86%</td>
<td>78%</td>
<td>92%</td>
<td>85%</td>
</tr>
<tr>
<td>Specificity</td>
<td></td>
<td>-----</td>
<td>96%</td>
<td>95%</td>
<td>97%</td>
<td>97%</td>
<td>95%</td>
</tr>
<tr>
<td>Referral rate</td>
<td></td>
<td>5%</td>
<td>5%</td>
<td>1.5%</td>
<td>1.7%</td>
<td>3%</td>
<td>5-10%</td>
</tr>
<tr>
<td>Biopsy rate</td>
<td></td>
<td>3%</td>
<td>1.5%</td>
<td>----</td>
<td>1.2%</td>
<td>1.0%</td>
<td>&lt; 2%</td>
</tr>
<tr>
<td>Benign : malignant ratio</td>
<td></td>
<td>4 : 1</td>
<td>2 : 1</td>
<td>1 : 2</td>
<td>0.7 : 1</td>
<td>2 : 1</td>
<td>&lt; 3 : 1</td>
</tr>
<tr>
<td>Reduction in mortality</td>
<td></td>
<td>29%</td>
<td>31%</td>
<td>24%</td>
<td>4%</td>
<td>20%</td>
<td>n/a</td>
</tr>
<tr>
<td>Screen prevalence / expected incidence</td>
<td></td>
<td>1.3</td>
<td>3.2</td>
<td>2.0</td>
<td>2.0</td>
<td>3.8</td>
<td>3 +</td>
</tr>
</tbody>
</table>

* Randomised controlled trial results adjusted for selection effect (Chapter Three, section 3.6.3)
# Sensitivity has been estimated using Method One (section 4.1.2)
---- data unavailable
percent or higher in the first screening round.

The participation rate is linked to the identification of eligible women, in that the two combined produce the coverage of the eligible population. With over ninety percent identification of eligible women a participation rate of seventy-eight percent would result in more than seventy percent coverage (seventy percent of women in the target population being screened). Both the HIP and Swedish Two-Counties programmes had over seventy percent coverage, but they had been able to identify and invite one hundred percent of their target population, whereas the pilot programme would not be able to do this. A target of seventy percent coverage was chosen because this seemed the highest possible coverage that could realistically be achieved in a programme that lacked a fully-inclusive population register.

4.1.2 The screening test
It was important for the radiologists in the pilot programme to match the standard of film reading achieved in successful programmes. It was thought that films from established screening programmes (where the outcome from each film is known) could be used as a “gold standard” if such films could be obtained. In the Otago and Southland pilot programme each woman’s films were read by two radiologists. The radiologists divided the screened women into two categories on the basis of their screening films; those who required further investigation, and those who did not. It was decided to assess inter- and intra-rater reliability using coefficient kappa to measure agreement. This had rarely been done in breast cancer screening programmes; in the few studies on inter-rater reliability for mammography, radiologists achieved kappa scores of 0.8 for agreement on the classification of parenchymal pattern (Boyd et al 1986). Reported kappa scores for recommendation for further action were not as high (up to 0.6) however these had been based on four possible recommendations rather than two where it may be possible to achieve greater agreement (Boyd et al 1982, Boyd et al 1986, Jannarone et al 1987). It is important for inter-rater reliability to be high, but this must be combined with high sensitivity and specificity. High inter-rater reliability would be irrelevant if the sensitivity were low (all this would mean is that both radiologists were missing suspicious films).
Repeat mammograms

Poor technical quality of some films could hamper their interpretation. It was important to maintain good technical quality so that women were not recalled unnecessarily (Breast Cancer Screening Working Group 1987). It is possible to keep the number of mammograms that have to be repeated for technical reasons alone as low as 0.4 percent (Sickles et al 1987). Technical quality assurance for the Otago and Southland pilot programme was the responsibility of the Otago University Departments of Radiology and Medical Physics. From the time that the evaluation protocol was first developed, a medical physicist was involved so that information about technical quality was part of the feedback given to the programme staff. The number of extra films taken, and the number of women recalled for technical reasons were recorded in the evaluation, as an indication of the technical quality of mammography in the pilot programme. The aim was to maintain the number of women referred for technical reasons toward the lower end of the range reported for other screening programmes. Technical recall rates in other breast cancer screening programmes ranged from 0.4% in the United States (Sickles et al 1987) to 3.8% in Australia (Rickard et al 1991).

Specificity

It is difficult to measure sensitivity and specificity during the first screening round because some false negative results cannot be identified until the second round. However it was possible to estimate sensitivity and specificity in the shorter term for the evaluation. Specificity is the proportion of women without disease who are correctly identified as negative by the screening test. If a test has high specificity it means that only a very small proportion of women have false positive tests. Achieving high specificity is important in quality control in order to minimise the number of women undergoing unnecessary investigations as a result of false positive tests. Specificity can be estimated early in the screening process. It is possible to estimate specificity before the second screening round by dividing all negative tests (including false negatives) by the sum of all negatives and false positives (Miller 1985, Morrison 1985). This is an adequate estimate of specificity (although false negatives have been included in the numerator and the denominator) because the number of false negatives is very small in relation to the number of true negatives.

The target set for specificity in the screening programme was based on results from the randomised controlled trials of screening (Table 4.1) and European screening programmes. All had specificities
of ninety-five percent (Breast Cancer Screening Working Group 1987, Gad et al 1984) or greater. The target for specificity in the pilot programme was therefore set at ninety-five percent.

High sensitivity is more important than high specificity in reducing breast cancer mortality, since sensitivity is the ability of the test to correctly identify women with breast cancer. A small drop in specificity would not influence mortality greatly but would cause a considerable number of women to undergo unnecessary investigations. High sensitivity is sometimes achieved at the expense of specificity, since lowering the threshold of suspicion in order to miss very few women with breast cancer usually means that the number of false positive tests increases, and specificity drops. This trade-off, and the importance of reducing breast cancer mortality were acknowledged when the evaluation was planned; the target for specificity was set at the lower end of the range published for successful breast cancer screening programmes (Table 4.1).

Sensitivity

It was important to estimate sensitivity because this could be used to help predict the effect of the programme on breast cancer mortality. Sensitivity has been measured in screening programmes in three different ways. The three methods for estimating sensitivity in breast cancer screening programmes are described below:

Method One defines sensitivity as screen-detected cancers divided by the sum of screen-detected cancers plus interval cancers in the following twelve months. The disadvantages of this method are (i) it includes cancers that were not detectable at screening (but also excludes slow-growing cancers that were detectable at screening but were not detected clinically within twelve months), (ii) it may include non-progressive cancers in the numerator, thus overestimating sensitivity, and (iii) it overestimates sensitivity for all programmes except those with a one-year screening interval.

Method Two estimates sensitivity by the formula $1 - PI_1$ where $PI_1$ is the proportional incidence of interval cancers (Day 1985). In the year after screening, screened women should have a reduced incidence of breast cancer compared with the incidence that would have been expected in the absence of screening. This happens if screening has successfully detected breast cancer early (if it has detected cancers that would otherwise have appeared later). One way of estimating this is to compare the number of interval cancers to the number of cancers diagnosed over the same period.
in a control group. In a randomised controlled trial a control group is available, but in other situations an historical comparison group can be used provided there is no evidence of a change in the underlying incidence of breast cancer over the time of the study. If the number of interval cancers is low compared with the expected incidence of cancer it suggests that the screening test is picking up many of the breast cancers that would otherwise have arisen in the following year. Subtracting the proportional incidence from one gives an estimate of the sensitivity of the screening test; if the proportional incidence is five percent it suggests that ninety-five percent of cancers were detected early by screening. Calculating the expected breast cancer incidence in the screened group in the absence of screening (Day et al 1989) is straightforward:

The incidence rate in the total population, \( I_t = P_l + (1-P)I_{na} \) where the \( P \) is the participation rate, 
\( I_l \) is the incidence rate in attenders 
\( I_{na} \) is the incidence rate in non-attenders.

The expected incidence in the screened group, in the absence of screening, is therefore:

\[
I_s = I_t - \frac{(1-P)I_{na}}{P}
\]

\( I_t \) is the incidence in the control group (in a randomised controlled trial) or in an historical comparison group.

The advantage of this method is that it relates interval cancers to the number of cancers that would have been expected to develop in the screened women in the absence of screening (the expected number of cancers is estimated from the incidence in a control group or an historical comparison group), and comparisons between programmes are therefore not affected by differences in prevalence, nor by selection bias. Overdiagnosis bias is avoided because non-progressive cancers are not included in the estimation of sensitivity (since the calculation is based on a comparison of interval cancers with expected cancers).

Method Three defines sensitivity as screen-detected cancers divided by the sum of screen-detected cancers plus cancers missed at screening. This method involves re-reading the previous films for all
women diagnosed with breast cancer in the interval between screens and at re-screening (used in Malmö and Canada). This method cannot be used before the second screening round.

The important issue for the evaluation was not necessarily which method is better at estimating the sensitivity of the screening test, but which could be used to produce reliable early targets for evaluating the programme. Method Three is not useful before the second screening round. Methods One and Two can be used during the first screening round. Method One requires information about cancers detected by screening, and interval cancers. This information is relatively easy to obtain because screened women are usually followed up. Method Two requires information about interval cancers and also cancer incidence in an unscreened group (historical, or women not yet invited for screening) and cancer incidence in women who do not participate in the screening programme. Method Two should give the most accurate estimate of sensitivity but Method One has the advantage of being easily calculated early in the programme. Although Method One overestimates sensitivity it can be used as an interim target. Targets for sensitivity estimated by both methods are given in Table 4.2.

To reduce breast cancer mortality in the eligible population a screening programme must be able to (a) identify a high proportion of the women with early breast cancer, and (b) include enough women for the group as a whole to benefit. The product of sensitivity multiplied by coverage is the maximum percentage of breast cancers that can be detected in a population-based screening programme (Table 4.3). The HIP trial had a sensitivity of eighty-two percent with sixty-five percent coverage (equivalent to seventy-eight percent coverage without self-selection). Both sensitivity and coverage were considerably higher than this in the other trials except Malmö and Edinburgh. The target set for sensitivity in the Otago and Southland pilot programme was eighty-five percent sensitivity (Method One) and the target set for coverage of the eligible population was seventy percent (section 4.1.1). The eighty-five percent target for sensitivity was chosen because it was comparable with sensitivity in the randomised controlled trials, and because together with seventy percent coverage of the eligible population it would give the programme a maximum breast cancer detection rate of sixty percent (eighty-five percent sensitivity multiplied by seventy percent coverage) which is close to the lowest detection rate for a successful programme (Table 4.3).
TABLE 4.2

A comparison of two methods of estimating sensitivity in randomised controlled trials

<table>
<thead>
<tr>
<th>Trial:</th>
<th>Women screened (age range)</th>
<th>BC detected</th>
<th>I.C.</th>
<th>Sensitivity Method</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIP</td>
<td>20,166 (40-64)</td>
<td>55 (2.7/1000)</td>
<td>12</td>
<td>82</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>42 (2.1/1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2C Kopparberg</td>
<td>39,713 (40-74)</td>
<td>278 (7.0/1000)</td>
<td>13</td>
<td>96</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>87 (2.2/1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAL</td>
<td>15,748 (45-69)</td>
<td>118 (7.5/1000)</td>
<td>33*</td>
<td>78</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>58 (in 22 mths)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDB</td>
<td>14,971 (45-64)</td>
<td>92 (6.2/1000)</td>
<td>8</td>
<td>92</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>29 (1.9/1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STK</td>
<td>32,555 (40-64)</td>
<td>128 (4.0/1000)</td>
<td>21</td>
<td>86</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>65 (2.0/1000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(I.C. = interval cancers within twelve months from screening)

* interval cancers in twenty-two months (Malmö)
# TABLE 4.3

Sensitivity and coverage in screening programmes

<table>
<thead>
<tr>
<th>Trial</th>
<th>Coverage of Eligible Population</th>
<th>Test Sensitivity</th>
<th>Maximum Breast Cancer Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIP</td>
<td>78%</td>
<td>82%</td>
<td>64%</td>
</tr>
<tr>
<td>STK</td>
<td>77%</td>
<td>86%</td>
<td>66%</td>
</tr>
<tr>
<td>S2C</td>
<td>77%</td>
<td>96%</td>
<td>74%</td>
</tr>
<tr>
<td>EDB</td>
<td>50%</td>
<td>92%</td>
<td>46%</td>
</tr>
<tr>
<td>MAL</td>
<td>53%</td>
<td>78%</td>
<td>41%</td>
</tr>
</tbody>
</table>

* Participation adjusted for selection (Chapter Three, section 3.6.3)

(Sensitivity calculated by Method One, section 4.1.2)
4.1.3 Referral for assessment

The number of women referred will be influenced by the sensitivity and specificity of the screening test. If referral rates are low it may be that the test accurately distinguishes between women who require further assessment and those who do not, but alternatively it could mean that the sensitivity of the test is low and that women who require further assessment have not been referred. Referral rates in successful programmes have been five percent or less. In the Swedish Two-Counties trial there was a five percent referral rate in the first screening round which decreased to 2.5 percent in later rounds. However the Swedish programme had a two-stage referral process where some women were recalled for more detailed mammograms, and only those with persistent abnormalities were then referred for assessment (Gad et al 1984). It is not clear from published reports how many of the other programmes had similar two-stage referral procedures. In the HIP study the referral rate was five percent but this was averaged across screening rounds. The rate in the first round was probably higher than five percent, since referral rates are usually higher in the first than in subsequent screening rounds (Kopans and Swann 1988). In the pilot programme it was decided that all symptomatic women should be referred irrespective of the results of their mammograms. Because the referral rate in the Swedish Two-Counties trial was five percent in the first screening round after detailed mammography, and possibly higher in the first round in the HIP trial, and because the pilot programme also had a policy of automatically referring all symptomatic women, the target for the first round of the pilot programme was set at between five and ten percent, but with the expectation that it would decline in subsequent screening rounds. The purpose of further assessment is to distinguish between women who require biopsies and those who do not. Because having a biopsy is likely to cause women anxiety and since many biopsies resulting from screening require hook-wire localisation and a general anaesthetic, it is important to minimise the number of unnecessary biopsies. Provided test sensitivity and specificity are adequate, the biopsy rate and benign to malignant ratio indicate how well the assessment process identifies women who require biopsies. Successful screening programmes in Europe have had benign to malignant biopsy ratios between 1.7 : 1 and 3 : 1 (Breast Cancer Screening Working Group 1987). The ratio is usually higher during the establishment of screening programmes (Kopans and Swann 1988, Stockdale et al 1988). In recognition of this the target for the Otago and Southland pilot programme was set at 3 : 1 or less initially, with the expectation that it would decline as the programme continued.
4.1.4 Breast cancer detection

Some measures of clinical outcome are important predictors of the programme's effect on breast cancer mortality. Day et al (1989) suggested, based on the results from the Swedish Two-Counties trial, that the prevalence of breast cancer at the first screening round should be at least three times the expected incidence of breast cancer in the absence of a screening programme, since taken together with the incidence of interval cancers, this gives an indication of the ability of the screening programme to detect cancers early. The stage-distribution of cancers detected by screening can be recorded for comparison with the stage-distribution seen before the introduction of screening. The stage-distribution of screen-detected cancers was suggested as a surrogate endpoint for reduced mortality because it appeared to explain almost all the reduction that occurred in the Swedish Two-Counties trial (Duffy et al 1991, Tabar et al 1992). The problem with using stage-distribution as a surrogate measure is that length and overdiagnosis biases would tend to shift the stage-distribution of screen-detected cancers towards the earlier stages, and it would be possible for a shift in stage distribution to occur due to the lead-time resulting from screening, without a concomitant reduction in mortality. In fact stage-distribution did not appear to explain the reduced mortality in the United Kingdom trial of early detection of breast cancer (Moss et al 1994). To avoid the problem of bias affecting stage-distribution, the absolute rate of advanced breast cancer can be assessed after about the third year of screening, as this is an indicator of the mortality rate expected in future years (Day et al 1989). Stage-distribution can only be used as an approximate indicator of the likely effect of the programme, and must be combined with other surrogate endpoints, especially in the prevalence screen, when it is more likely to be affected by length bias (Duffy et al 1991). Because stage-distribution and the absolute rate of advanced cancers are both better used in the incidence screen, they have not been included as targets for the evaluation of the first round of the pilot programme.

4.2 Measuring the acceptability of the pilot programme

4.2.1 Participation in screening

The lowest participation rate for a successful programme was sixty-five percent in the HIP trial. Because all the women in the trial were identified from the HIP insurance records this represented sixty-five percent of all the eligible women. However in the HIP trial women with a higher risk of cancer tended to self-select themselves for screening (Shapiro 1977, Shapiro et al 1982). This
means that the HIP screening programme detected more women with breast cancer than would be expected from a sixty-five percent participation rate. In other programmes the selection effect was different, but it is possible to adjust participation to take account of this as described in Chapter Three, section 3.6.3 (Tables 3.6 to 3.8). In order to approach or exceed the lowest recorded participation rate for successful programmes, the pilot programme has a target participation rate of seventy-eight percent or higher in the first screening round. With the identification of ninety percent of eligible women using the electoral roll, the screening programme would have a coverage of seventy percent of the target population and thus seventy percent of the target population would be screened. Supplementing the electoral roll with general practice age-sex registers would be likely to improve identification of eligible women.

4.2.2 Surveys of the acceptability of screening

Several surveys to find out about the acceptability of screening were carried out. The first was a pre-screening survey which was conducted before the introduction of the screening programme, in order to identify characteristics of eligible women, and of a screening programme, that would be likely to affect participation. Surveys of screened women and women who attended the assessment clinics were also carried out. The purpose of these surveys was to find out about satisfaction with the service, to provide feedback for programme staff, to obtain an indication of future participation, and to collect information about costs to women which would be used in the economic analysis of the pilot programme. A survey of non-attenders was carried out to discover whether there were aspects of the screening programme that could be changed to encourage these women to attend, and improve participation. These surveys are described in detail in Chapter Six.

4.3 The economic evaluation of the pilot programme

The economic evaluation of the Otago and Southland pilot programme was coordinated by Dr Nancy Devlin, a health economist who acted as a consultant for the evaluation of the pilot programme (Devlin NJ et al 1993). The objective of the economic evaluation was to determine whether screening could be introduced in New Zealand “in a way that is effective from an economic viewpoint” (Skegg et al 1988). Policy makers could then determine whether the pilot
programme (and, by extrapolation, a national programme) would be a worthwhile use of health resources, and how it compared with alternative uses of those resources. Information was collected about the costs associated with the pilot programme, including the costs of running the programme, and costs to women taking part (collected in the surveys of screened women). With this information estimates were made of the cost per woman screened and the cost per cancer detected in the first round of screening.

These estimates were compared with those from existing breast cancer screening programmes to establish whether the pilot programme screened women and detected breast cancer for similar costs. After several years it will also be possible to estimate the cost per year of life saved, based on the number and stage distribution of cancers detected in the pilot programme, and on predictions of resulting life years gained. The predictions will be based on other programmes where it was possible to measure life years gained by comparing the group offered screening with a control group, with a follow up period of many years. The results from the economic evaluation of the pilot programme are presented as Appendix Two, and summarised and discussed in Chapter Seven.

4.4 Data collection for the evaluation

The collection of data for the evaluation is discussed in detail in Chapters Five and Six. A description of the data collection systems that were implemented at the start of the evaluation is given here.

Initially it was hoped that the pilot breast cancer screening programme would have a stand-alone computer system, adapted from the computer system of an established screening programme. This was not possible because of Dunedin Hospital's requirement for the pilot programme to use its existing radiology computer system. This computer system had been designed for booking appointments, generating letters, and recording all investigations and results. Some adaptations were made to the system to allow for instance, results from double reading of mammograms, and results from the assessment clinic to be entered.
Because of the number of women who would be taking part in the screening programme, it was decided that as much data as possible for the evaluation should come directly from the pilot programme's computer system, rather than having separate data collection systems for the evaluation. It was also thought that this would be more accurate, since all the information about screened women would be automatically entered into the pilot programme's computer system. Each woman who attended the screening programme was allocated a national master patient index (NMPI) number as a unique identifier, so that her records could be matched with Dunedin Hospital records. At the commencement of the pilot programme Southland Hospital intended to change to the NMPI system, but this did not eventuate. The Southland women were identified using Southland hospital identification numbers instead.

The evaluation team specified the routinely collected data they would require, and a programme was written by the computer consultants who were modifying the Dunedin Hospital radiology software, so that this information could be downloaded regularly to the evaluation database. The information that was requested by the evaluation team is given as Appendix Three, and included personal information such as a unique identifying number (the National Master Patient Index or NMPI number), the woman's date of birth, ethnic group, a domicile code, the method of identifying that woman (e.g., electoral roll, general practice register, self-referred), and her eligibility for screening. For each screening visit information about the method of invitation, the screening centre (Dunedin, Invercargill, or mobile), the time and date of the appointment, the number of films (and reason for any extra films), the result of the screen and the date the result was produced, results of assessment (if applicable), and the final result (whether or not the woman was diagnosed with breast cancer) was to be downloaded. The information for each screen was flagged with the screen number and the identification number so that a woman's entire screening history could be linked with her personal information into one file which was identified by the NMPI number.

This information was downloaded at regular intervals and then analysed using SPSS-X (Statistical Package for Social Sciences 1990). A separate programme was written to convert film reading results into a separate ASCII file. The information from these files was then used to evaluate the screening and assessment phases of the programme, including film reading, referral rate, biopsy rate, benign : malignant biopsy ratio, and breast cancer detection rate. Initially the download was designed so that the same information would be collected from the two pilot programmes, but
because the Waikato programme used a different computer system this did not eventuate, and so separate analyses were carried out for each pilot programme.

Quality control measures were undertaken, to ensure the validity of data that had been downloaded from the screening programme computer system for the evaluation of the Otago and Southland Pilot Breast Cancer Screening Programme:

At regular intervals Sheila Williams produced a printout of the National Master Patient Index (NMPI) numbers for those records where the dates on the download from the screening programme computer system were in question, or where there was missing data. Examples of inconsistent dates included where the woman’s date of birth was inconsistent with the age-range for the screening programme, or where dates for screening and assessment appointments, and results were out of sequence. Using the NMPI numbers from the printout to identify each record, Ann Richardson manually searched the records at the screening centres (this involved trips to Invercargill for Southland records), to extract the correct information. The correct data was then entered on the screening programme computer system. Sheila Williams then produced another printout from the next download, to check that the previous inconsistencies had been rectified.

Women who had biopsies were identified in the evaluation download. The screening programme also kept a manual record of women who had had biopsies (a file containing copies of the histology reports for screened women was kept at the screening centre). At regular intervals a printout from the evaluation download was compared with the manual biopsy records at the screening centre. Any discrepancies were followed up by obtaining the woman’s screening records. Where necessary, corrections were made to the data that had been entered on the screening programme computer system (and thus to the evaluation download). The need for such corrections was extremely rare.

Women diagnosed with breast cancer through the screening programme were identified from the evaluation download. This information was then checked against two separate sources. Firstly it was checked against the biopsy file held in the screening centre which contained copies of all histology reports for women who had been referred for a biopsy as a result of screening, to ensure that all women who had been diagnosed with breast cancer through the screening programme were
correctly identified on the evaluation download. Where there were inconsistencies, the NMPI number was used to find the woman’s records to check, and where necessary, correct the information on the computer system. Secondly, information on women diagnosed with breast cancer was checked against the pathology records from all the pathology laboratories in Otago and Southland. There were four such laboratories; two private and two hospital pathology laboratories in Dunedin and Invercargill. Obtaining this information was important, both to validate the data on the evaluation download, but also to identify women with interval cancers. In Dunedin information about breast histology in all women aged fifty and over was collected from the Dunedin Hospital Pathology Department computer system, and the NMPI number was used to determine whether the women were diagnosed as a result of screening or not. Information on breast histology for Southland women was produced from the Southland Hospital pathology computer system. The Southland Area Health Board did not adopt the NMPI system which meant that printouts with patient numbers could not be used to identify screened women. Whenever a woman over fifty was diagnosed with breast cancer through the Southland Hospital Pathology Department a copy of histology report was routinely sent to the evaluation coordinator, and for the private pathology laboratories (where smaller numbers were involved) the evaluation coordinator requested copies of the forms at regular intervals. In each case the histology form had the woman’s name blanked out. Date of birth was used to look for possible matches with the screening register, and where there were matches and the name was required, the woman’s surgeon asked her permission to release her name for the evaluation. This provided a method of identifying women with breast cancer diagnosed outside the programme, yet still maintaining confidentiality between the pathologist and the patient. The methods used to collect data from the pathology laboratories are also described in detail in Chapter Five, section 5.4.

Data that could not be collected routinely was collected manually or in the form of surveys. Before the screening programme began a survey of general practitioners in Otago and Southland was carried out to discover what information practitioners needed about the pilot programme, and to find out how many general practitioners had age-sex registers, and whether they would be prepared to help with identifying and inviting eligible women (Chapter Five, section 5.1.2). A pre-screening survey was carried out to find out about women’s attitudes towards breast cancer screening and to identify factors that could influence participation (Chapter Six, section 6.1). Surveys of women who attended the screening and assessment clinics were carried out to determine the acceptability
of the screening programme (Chapter Six, sections 6.3 and 6.4), and a survey of women who declined an invitation to screening was carried out to investigate reasons for non-participation (Chapter Six, section 6.5).

Some of the data that should have been available from the evaluation download were not available. The reasons for this failure of the download, the alternative methods used to collect the data, and the ramifications of problems with data collection are discussed in Chapters Five to Seven.

4.5 Summary
Comparing interim measures with the results of established programmes has now become an accepted way of evaluating screening programmes. Day et al (1989) suggested a method for evaluating screening programmes which compared early parameters of new programmes with those of established programmes that have been successful in the longer term. Their method deals with the projected effect of the programme on breast cancer mortality and focuses on participation and test sensitivity in particular. However evaluation must also include quality control. Verbeek et al (1991) described a method of short-term quality control that addresses some aspects of quality control in the evaluation of screening.

The evaluation method presented here uses short-term targets and addresses both quality control and the effect of the programme on breast cancer mortality. These targets were derived from the published results of randomised controlled trials of mammographic screening and together represent the minimum required for a pilot programme to be operating effectively. The results are fed back to the pilot programmes so that they can monitor their progress and so that corrective action can be taken if necessary (Table 4.4). Failure to reach one of the targets may be compensated for by far exceeding another relevant target. For example, a programme that is slightly below the target set for sensitivity but which attracts very high participation rates may still produce a worthwhile reduction in breast cancer morbidity and mortality in the eligible population.
### TABLE 4.4

**Action on failure to reach evaluation targets**

<table>
<thead>
<tr>
<th>Target:</th>
<th>Remedial action:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification</td>
<td>Supplement electoral roll with information from GP age-sex registers.</td>
</tr>
<tr>
<td>Participation</td>
<td>Survey non-participants to identify aspects of the programme that have deterred them from participating. Increase publicity and education about screening.</td>
</tr>
<tr>
<td>Film reading</td>
<td>Is technical quality adequate? Film reader training.</td>
</tr>
<tr>
<td>Recall rate</td>
<td>Reassess recall criteria.</td>
</tr>
<tr>
<td>Biopsy rate and benign:malignant ratio</td>
<td>Are sensitivity and specificity adequate? Assessment team training.</td>
</tr>
</tbody>
</table>
The results of the evaluation of the Otago and Southland pilot programme are described and discussed in Chapters Five and Six. The effectiveness of the pilot programme is covered in Chapter Five and the acceptability of the programme is discussed in Chapter Six. The implications of these results including the economic efficiency of the pilot programme, are discussed, and recommendations arising from these results are made in Chapter Seven.
CHAPTER FIVE

The Effectiveness of the Pilot Programme

Introduction

The methods and results of the evaluation of the effectiveness of the pilot breast cancer screening programme are presented and discussed in this chapter. The evaluation of the effectiveness of the pilot programme has been divided into three areas (the identification and invitation of eligible women, the screening test, and referral for further assessment), and the performance of the pilot programme has been compared against the targets for effectiveness that were described in Chapter Four. The identification and invitation of eligible women to take part in the pilot programme are described in section 5.1, including the establishment of the screening register (sections 5.1.1 and 5.1.2), and the investigation of invitation methods (section 5.1.3). The quality of mammography in the pilot programme is examined in section 5.2, including technical quality (section 5.2.1) and radiological interpretation (section 5.2.2). Referral for further assessment is covered in section 5.3 which includes information on the referral rate, the biopsy rate, and the benign : malignant biopsy ratio. Breast cancer diagnosis in the pilot programme is discussed in section 5.4 including breast cancer incidence in women aged fifty to sixty-four in Otago and Southland before the pilot programme (section 5.4.1), breast cancer detection by screening (section 5.4.2), interval cancers (section 5.4.3), the diagnosis of breast cancer outside the screening programme (section 5.4.4), the sensitivity and specificity of screening (section 5.4.5), the stage-distribution, size and grade of breast cancers detected in the screening programme (section 5.4.6), and the assessment of overdiagnosis in screening programmes (section 5.4.7). The results of the evaluation of the effectiveness of the pilot programme are summarised in section 5.5.
5.1 Identifying and inviting eligible women

The aim of a screening programme is to reduce the impact of breast cancer on the eligible population. To achieve this as many eligible women as possible must be identified and invited to take part in the programme. Several studies were carried out as part of the evaluation, to determine the best ways to identify eligible women (sections 5.1.1 and 5.1.2) and invite these women to take part in the pilot programme (section 5.1.3).

5.1.1 Establishing the screening register

It was important that all women aged fifty to sixty-four and living in Otago and Southland could be identified and invited to take part in the pilot programme so that the benefits from screening would be available to the entire eligible population (Chapter Four, section 4.2.1).

Methods

The names, addresses, and mesh block numbers of people aged fifty to sixty-four and registered as electors in the Otago, Dunedin North, Dunedin West, St Kilda, Wallace, Clutha, Awarua, and Invercargill electorates, and those on the Southern Maori roll corresponding to these electorates were obtained from the electoral roll centre (Chapter Four, section 4.1.1). Together these electorates matched the boundaries for the Otago and Southland Area Health Boards, which were the designated boundaries for the screening programme. These boundaries were also similar to the Otago and Southland statistical areas, used by the Department of Statistics for grouping census data by area.

The first names of the electors were used to select all females aged fifty to sixty-four. Mesh block numbers were used to group the women into Dunedin, Invercargill, and mobile screening areas. (A mesh block is the smallest geographical unit for which data is collected by the Department of Statistics. Mesh blocks are used to define electoral boundaries and they can be aggregated to form larger geographical units such as Statistical Area Units.) The number of fifty to sixty-four year old women identified through the electoral roll was compared with the census number of fifty to sixty-four year old women living in Otago and Southland at the time that the programme started to find out what percentage of eligible women had been identified.
Results

For Otago and Southland at the time that the screening programme started in 1991 there were 17,145 women aged fifty to sixty-four on the electoral roll. This compared with 18,555 women aged fifty to sixty-four and living in Otago and Southland at the 1991 census (ninety-two percent identification of fifty to sixty-four year old women). Therefore the target of identifying at least ninety percent of women aged fifty to sixty-four living in Otago and Southland was met by using the electoral roll alone. But the electoral roll rapidly becomes out of date (since people change address in between updates of the roll) so new information has to be obtained whenever the electoral roll is updated.

The electoral roll was one source of the names and addresses of women aged fifty to sixty-four and living in Otago and Southland. There was an eight percent discrepancy between the census and electoral roll figures. The reasons that some women might not appear on the electoral roll despite being included in census figures include not having registered as an elector, changing address or dying since the census, or being committed to a psychiatric institution or prison. As an alternative to a population register, the electoral roll, with ninety-two percent identification of women in the eligible age-range seems reasonable. In Edinburgh the general practice registers that were used to identify women for the screening programme correctly identified ninety percent of women in the eligible age-range (Roberts et al 1989).

The other source of names and addresses of eligible women was general practice age-sex registers. It was hoped that these registers could be used to supplement the information from the electoral roll. The advantages of using general practice age-sex registers to invite women for screening were that general practitioners could be involved with the pilot programme from the time the invitations were sent out, women who were not on the electoral roll could be identified, and ineligible women (for instance those being followed up after a diagnosis of breast cancer) would not be invited inappropriately.
5.1.2 Survey of general practitioners in Otago and Southland

A survey of all the general practitioners in Otago and Southland was carried out before the programme started, to discover what information practitioners needed about the pilot programme, and to find out how many general practitioners had age-sex registers, and whether they would be prepared to help with identifying and inviting eligible women.

Methods

The 159 general practitioners registered with the Medical Council as in practice in Otago and Southland were sent questionnaires (Appendix Four). Reminder letters and then telephone calls were used to follow up non-respondents. The general practitioners were asked whether they knew about the planned screening programme, what their policy for recommending mammographic screening was, whether they would encourage eligible women to take part in screening, and whether they needed any information before recommending screening to women in their practices. They were also asked whether their practices had age-sex registers, whether they could produce a list of women aged fifty to sixty-four, and whether they would be prepared to sign letters inviting eligible women in the practice to be screened.

Results

Of the 159 questionnaires sent out, five were returned because the practice no longer existed, three because the doctor was no longer in practice, and one because the doctor had died. Of the 150 remaining general practitioners 141 (ninety-four percent) took part in the survey, five refused to complete the questionnaire, three could not be contacted by telephone, and one was away on leave. Most general practitioners in Otago and Southland were aware of the planned pilot programme; 140 answered this question and 107 (seventy-six percent) already knew about the programme. Most practitioners had a policy of recommending mammography to certain groups of women in their practices (Table 5.1). Of the respondents 129 (ninety-one percent) said that they would encourage women to take part in the screening programme. A further six percent said they would do so provided general practitioners were involved in the programme. One hundred and one general practitioners (sixty-seven percent) were willing and able to produce lists of the women in their practices to supplement the screening register. Of 119 practitioners who answered the question about signing invitations to screening, 110 (ninety-two percent) said they would be willing to sign.
### TABLE 5.1

Recommendations for mammography in Otago and Southland general practices in 1991  
(before the start of the pilot programme)

<table>
<thead>
<tr>
<th>Recommendation</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>No policy</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Recommended only if risk factor or patient request</td>
<td>43</td>
<td>30</td>
</tr>
<tr>
<td>Recommended to women over 40</td>
<td>45</td>
<td>32</td>
</tr>
<tr>
<td>Recommended to women over 50</td>
<td>46</td>
<td>33</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>141</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>
Outcome

The general practitioners identified in the survey were approached by the screening programme manager. General practice age-sex registers were the initial source used when the invitations were sent out. Screening programme staff also helped to compile lists for practices which did not have age-sex registers. The response from the general practitioners was better than had been expected from the survey, with only two practices declining to provide lists of the names and addresses of the women aged fifty to sixty-four in their practices.

Each woman identified from a general practice list was checked off on the electoral roll, and women who had not been identified through the general practice registers were sent invitations later. Once the means of identifying and inviting eligible women had been developed, an investigation of the best method for inviting the women was carried out.

5.1.3 Determining the best invitation method

The method of inviting women to screening has been shown to affect participation. It had previously been shown that personal invitations result in higher participation in screening programmes than invitations which are not personalised, and that invitations with a specified appointment time result in higher participation than invitations which require women to make their own appointments (Hobbs 1986, Williams and Vessey 1990, Turnbull et al 1991). As part of the evaluation of the effectiveness of the pilot programme two randomised controlled trials were designed to find out the most effective ways of using personal invitations to invite women to take part in screening (Richardson et al 1994). The first randomised controlled trial investigated the effect of general practitioner endorsement of an invitation to screening, and the second investigated the effect of postal reminders compared with telephone reminders. Each invitation included a specified appointment time.

Trial to investigate the effect of general practitioner endorsement of screening invitations

A randomised controlled trial was carried out to investigate whether supporting letters from general practitioners accompanying the invitations from the screening centre affected participation in a population-based breast cancer screening programme.
Methods

All women aged fifty to sixty-four whose names were on the age-sex register of a large urban Health Centre with a practice population of 15,500 and who were eligible for screening (482 women) were randomly allocated either to receive a letter from their general practitioners with their screening invitations or not. The general practitioners at the health centre provided personally signed form letters on the practice letterhead. If there was no reply within two weeks of the invitation being sent, a postal reminder (not signed by the general practitioner) was sent from the screening centre. Once all the appointment dates had passed the two groups were compared to see whether they had differed in the need for reminders and in the final participation rate. The trial was designed to have eighty percent power to detect a difference in participation of fifteen percent at 0.05 significance, assuming a participation rate of fifty percent in the comparison group (Machin and Campbell 1987).

Results

Invitations were sent to 482 women, 248 with accompanying letters from their general practitioners and 234 without. Randomisation produced similar groups with respect to age. Also a similar percentage of each group was ineligible for screening (Table 5.2). A greater proportion of women who had not received letters replied specifically declining the invitation to be screened (fifteen percent compared with eight percent of those who had received letters). This difference was statistically significant (p = 0.03, difference 6.5 percent, 95% CI 0.8 to 12.2).

Excluding women who were ineligible or could not be contacted because they were no longer at that address left 203 women who had been sent letters with their invitations and 192 women who had not. Of these women fifty-six percent of those who were sent letters participated as a result of the first invitation compared with forty-three percent of those who were not sent letters. This difference was significant (p = 0.01, difference 13 percent, 95% CI 3.2 to 22.7).

Including those who had required reminders, participation among eligible women who had received letters was seventy-one percent compared with sixty-two percent among women who did not receive letters signed by their general practitioners (p = 0.059, difference 9 percent, 95% CI -0.3 to 18.2).
TABLE 5.2

The effect of letters from general practitioners

<table>
<thead>
<tr>
<th>Women thought to be eligible for screening:</th>
<th>482</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Excluded:</td>
<td></td>
</tr>
<tr>
<td>recent mammogram</td>
<td>26</td>
</tr>
<tr>
<td>mastectomy</td>
<td>1</td>
</tr>
<tr>
<td>out of country</td>
<td>0</td>
</tr>
<tr>
<td>no longer at address</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Remaining women:</td>
<td></td>
</tr>
<tr>
<td>no reply</td>
<td>26</td>
</tr>
<tr>
<td>Declined</td>
<td>21</td>
</tr>
<tr>
<td>Did not attend appointment</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Screened without reminder</td>
<td>113 (56%)</td>
</tr>
<tr>
<td>Screened after reminder</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Total screened</td>
<td>144 (71%)</td>
</tr>
</tbody>
</table>

randomised
Discussion

In this trial, as with programmes in other countries, personal endorsement of invitations by general practitioners increased participation (Irwig et al 1990, Williams and Vessey 1989, Dorsch et al 1991, Hurley et al 1992). Including those who had required reminders, participation among eligible women who had received letters was seventy-one percent compared with sixty-two percent among women who did not receive letters from their general practitioners (p = 0.059). This participation rate was slightly higher than in a similar study carried out in Australia, where 68.6 percent participation was achieved (Dorsch et al 1991). The women in the local trial had consulted their general practitioners within the previous two years and their names were therefore on the practice register. Participation may be different among women who do not regularly attend a general practitioner and are not on a general practice register.

Only forty-three percent of the women in the trial were screened as a result of an unaccompanied invitation without a reminder. This level of participation would be unsatisfactory for a population based screening programme as the benefits of screening would be available to fewer than half of the eligible women. The cost effectiveness of the programme would also be lowered, as the major costs in screening are salaries and equipment, while the cost of consumables (such as film and chemicals) is low. The cost per woman screened drops as the number of women participating increases (Sickles et al 1986, McLelland 1987, UK Trial of Early Detection of Breast Cancer Group 1988).

It is clear that reminders are very important in increasing participation. Apart from the effect on participation rates there are other advantages to general practitioner involvement in breast cancer screening programmes. Such advantages include avoidance of inappropriate invitations (for instance to women who are already being followed up after a diagnosis of breast cancer). Also, women can discuss the invitation and the screening programme with their general practitioners, who will know when women in their practices are being invited and can provide support for women who require further assessment and investigation as part of the programme. Although the system of invitations used in this trial was time-consuming for the general practitioners involved, it contributed to updating the practice register and was seen as a good exercise in quality assurance for the practice.
Trial to compare postal reminders with telephone reminders

This trial compared the effect of postal reminders with that of telephoned reminders for women who did not respond to an initial invitation to participate in a breast cancer screening programme.

Methods

The names of 641 consecutive women who had not replied within two weeks to an initial invitation were collected. Women with telephone numbers were randomly allocated to receive either a postal reminder or a telephoned reminder. The telephone calls were made up to three times at different times of the day. The 146 women without telephone numbers formed a third group and were sent a postal reminder. The trial was designed to have eighty percent power to detect a difference in participation of fifteen percent at 0.05 significance, assuming a participation rate of fifty percent in the comparison group (Machin and Campbell 1987).

Results

Of 641 women who had not replied to a postal invitation, 495 had telephone numbers and were randomly allocated to receive either a telephoned reminder or a postal reminder. There were 146 women without telephone numbers. These women formed a separate group and all were sent postal reminders.

Table 5.3 shows the results of telephoned reminders compared with postal reminders. There were 248 women randomised to receive telephoned reminders. Of this group 118 were screened (forty-eight percent). Thirteen women said that they would like later appointments but did not specify a time. None of these women had been screened by the end of the study period and they have been included with the non-participants.

Of the 247 women who were sent postal reminders 121 women were screened (forty-nine percent). There was no significant difference in participation between the intervention and control groups (p = 0.8, difference 1.4 percent, 95% CI -10.2 to 7.4). Of the 146 women who did not have telephone numbers and so could not be randomly allocated to either group only forty women (twenty-seven percent) were screened. Twenty percent of the invitations in this group were returned because the address was incorrect.
TABLE 5.3

Comparison of telephoned with postal reminders

Women who did not reply to an initial invitation

<table>
<thead>
<tr>
<th></th>
<th>Telephone reminder</th>
<th>Postal reminder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>248</td>
<td>247</td>
</tr>
<tr>
<td><strong>Not screened:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No reply</td>
<td>49</td>
<td>57</td>
</tr>
<tr>
<td>Declined</td>
<td>33</td>
<td>35</td>
</tr>
<tr>
<td>Ineligible</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>Did not attend appoint</td>
<td>33</td>
<td>17</td>
</tr>
<tr>
<td><strong>Screened:</strong></td>
<td>118 (48%)</td>
<td>121 (49%)</td>
</tr>
</tbody>
</table>
Discussion

In this trial a telephone reminder and a postal reminder were equally effective for women who did not respond to an initial invitation. Neither method resulted in a large number of missed appointments, but thirty-three (thirteen percent) of those who received telephone reminders missed their appointments compared with seventeen (seven percent) of those who received postal reminders ($\chi^2 = 5.6 \ p < 0.05$). There were thirteen women among those telephoned who said that they wished to be screened later but who did not specify a time. None of these women had been screened by the end of the study period. It is sometimes difficult to refuse a telephoned reminder and it may be that these women did not wish to be screened. If this is so, telephone reminders may be undesirable.

As the reminder methods are equally effective, whichever best suited the administrative arrangements of the screening programme could be used. An advantage of postal reminders is that some are returned marked “wrong address”. As there is then no possibility of the appointment being used it can be allocated to another woman straight away. Telephone reminders may be more time-consuming than postal reminders (especially if several calls have to be made in order to contact some women) and they have been found to cost more than postal reminders (Hurley et al 1992). But it is easier to organise an appointment at relatively short notice by telephone. This provides the programme with greater flexibility in scheduling appointments (for instance gaps in the screening schedule due to cancellations can be filled at short notice).

In Australia the effectiveness, costs, and cost-effectiveness of various strategies for recruiting women to be screened were considered (Hurley et al 1992). This was a non-randomised trial, where recruitment strategies were compared in an Australian pilot programme. The strategies included three public recruitment strategies (local newspaper articles, community promotion, and promotion to physicians), and five personal recruitment strategies (invitation letters with or without specified appointment times, either alone or with a follow-up letter, or telephone call to non-attenders). Personal recruitment strategies were more cost-effective than public strategies, at less than half the cost per woman screened of public recruitment strategies. Postal reminders were more cost-effective than telephone reminders (Hurley et al 1992).
As a result of the two trials the Otago and Southland pilot programme is now using general practitioner letters with all invitations where possible, and postal reminders are sent to women who do not reply to the initial invitation.

5.2 The screening test

There are two particularly important aspects to quality assurance of mammography; the technical quality of the mammograms (section 5.2.1), and the quality of their radiological interpretation (section 5.2.2). In the Otago and Southland programme technical quality was the responsibility of a medical physicist. However some information, such as the number of extra films taken and the percentage of screened women who had to be recalled because of technical problems, was collected as part of the evaluation, as an indicator of the technical quality of the mammograms taken in the pilot programme. Other factors which affect the technical quality of mammograms include the calibration and maintenance of the mammography machine, positioning and compression when taking the mammograms, and film processing. These factors were monitored separately by the pilot programme radiographers and the medical physicist.

5.2.1 Technical quality

Because the programme used two-view mammography, the expected number of films was four per woman (two views for each breast). Information about the number of films taken at each screen was collected. Any more than four films per screen were regarded as 'extra films'. The reasons for extra films were categorised as 'technical', 'large breasts', and 'other'. The number of women recalled for assessment because of technical problems with the screening films (technical recalls) was also recorded.

Over the first eighteen months of screening ten percent of women had more than four films taken (at the time of screening) to complete the mammogram adequately. Thirty percent of these women had extra films taken for other than technical reasons, the most common reason being that the woman had large breasts. For women screened at the fixed screening centres, extra films required for technical reasons could be taken at the same visit, since each woman’s films were checked before she left the screening centre. For women screened at the mobile unit, if extra films were
required for technical reasons, a second visit was required, since films taken at the mobile unit were not processed at the time, but were sent to Dunedin for batch processing at the end of each day.

In the first eighteen months for the Otago and Southland programme 1.2% of all the women screened were recalled because of technical problems. The 95% confidence interval was 0.97 to 1.48 percent (Rothman and Boice 1982). The technical recall rate compared reasonably well with published technical recall rates in other breast cancer screening programmes which range from 0.4% in the United States (Sickles et al 1987) to 3.8% in Australia (Rickard et al 1991).

5.2.2 Film reading
In the pilot programme each woman's films were independently read by two radiologists. The radiologists had only two possible decisions; either to refer the woman for further assessment, or to enter her name for routine rescreening in two years time if she was still under sixty-five. Where there was disagreement the radiologists met to discuss the films and made a joint decision.

Information about their film-reading was given to the radiologists at intervals after the pilot programme started. This information included two by two tables showing recall rates and agreement for each of the film readers. For women who were diagnosed with cancer through the programme we checked that both readers had considered the films to be abnormal. The radiologists were also given a printout with the NMPI numbers for all the women who had been referred for assessment and the final outcome after assessment, to allow later re-assessment of the films, in the knowledge of the final outcome for each woman. As part of the evaluation, the effect of double reading compared with single reading, and the agreement between readers were assessed. The assessment of agreement was designed as part of the evaluation of the effectiveness of the pilot programme but the analysis was carried out by Sheila Williams, the biostatistician who worked on the evaluation (Williams et al 1995). A paper reporting these results is included as Appendix One and the results are summarised below.

The extent of agreement between the two main radiologists in the pilot programme was kappa = 0.65 (95% CI 0.62 to 0.67). This was higher than in other major studies on agreement in breast cancer screening programmes. An interesting result is that double reading increased the breast cancer detection rate in this study. Had there been only one reader, three to five percent of women
with breast cancer would have been missed, depending on which radiologist did the reading. In Finland (Anttinen et al. 1993) single reading would have led to either six percent or ten percent of the women with breast cancer being missed. In a study from Scotland (Anderson et al. 1994) 10.9% of women with breast cancer would not have been detected without double reading.

5.3 Referral for assessment

Methods
Information from the assessment clinic was collected on the download from the screening programme computer system. Validity checks were carried out to ensure that the information obtained from the download was accurate (Chapter Four, section 4.4). In the Otago and Southland programme all women attending for screening were asked to complete a consent form and a questionnaire which included questions about breast symptoms. A copy of the consent form and symptom questionnaire is included as Appendix Five. If a woman had recent symptoms documented on the questionnaire she was automatically referred for assessment even if her mammogram was normal. This did not happen in most screening programmes. To allow for this difference, the recall rate for Otago and Southland has also been given excluding symptomatic women who had normal mammograms.

Women were recalled for assessment if their films were abnormal, if they had recent symptoms on the symptom questionnaire, if their films were technically inadequate, or for any combination of these reasons. The women were recalled to a special clinic staffed by a radiographer, radiologist, nurse, counsellor, and surgeon. Most women were seen at the assessment clinic within two weeks of receiving their screen results.

Some screening programmes have two phases leading to referral, where some women are recalled for more detailed mammograms, and only those with abnormalities on detailed views are referred. Programmes that have two-phase referral, such as the Swedish programmes, referred five percent of women for assessment in the first screening round (Gad et al. 1984, Tabar et al. 1992). The Otago and Southland pilot programme did not refer women for detailed mammograms as an intermediate step, and also had a policy of referring all women with breast symptoms, irrespective
of their mammogram results. These policies increased the referral rate. Initially the target referral rate for the New Zealand pilot programmes was set at five to ten percent (Chapter Four, section 4.1.3) to allow for the differences in referral policy between the pilot programme and some other programmes. The referral rate calculated in the Otago and Southland pilot programme included all women who were referred to the assessment clinic, even if it was only for more detailed mammography.

Results
Of the 7,182 women screened in the Otago and Southland programme in its first eighteen months, 832 women (11.6%) were referred for assessment. After excluding women with symptoms the referral rate was 10.4%. Most of the women referred required further mammography and/or physical examination alone but some women had several investigations. Altogether, of the 832 women referred to the assessment clinics, 735 had mammograms, 307 had physical examinations, 213 women had ultrasound examinations, 150 had fine needle aspiration cytology (FNA), and 141 (two percent) of the women had biopsies (Table 5.4). Seventy-three women were diagnosed with breast cancer, giving a detection rate of ten breast cancers per thousand women screened. The benign : malignant ratio was 0.9 : 1. Breast cancer detection is discussed in detail in section 5.4.

Symptomatic women
None of the women diagnosed with breast cancer were symptomatic at the time of screening (Table 5.5). This suggests that the symptom questionnaire was not a useful indicator of the likelihood of breast cancer, and consideration should be given to modifying it or eliminating it. Alternatively the questionnaire could be retained but symptomatic women be given the option of further assessment rather than an automatic referral, as some symptomatic women may be anxious if they are not referred.

Discussion
The referral rate of eleven percent (or 10.4% excluding women with symptoms) in the pilot programme was higher than that reported in many breast cancer screening trials for the first round of screening and exceeded the target of between five and ten percent. The referral rate was 1.5% in the first screening round of the Stockholm trial (Frisell et al 1986, Frisell et al 1991) and five percent in the first round of the Swedish Two-Counties trial (Tabar et al 1992). In the Health
TABLE 5.4

Assessment results for the first eighteen months of screening in the Otago and Southland pilot programme

<table>
<thead>
<tr>
<th>Women screened</th>
<th>7,182</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women referred</td>
<td>832</td>
</tr>
<tr>
<td>Referral rate</td>
<td>11.6%</td>
</tr>
<tr>
<td>Referral rate (excluding women with symptoms)</td>
<td>10.4%</td>
</tr>
<tr>
<td>Biopsies (excludes fine needle aspiration)</td>
<td>141</td>
</tr>
<tr>
<td>Biopsy rate</td>
<td>1.96%</td>
</tr>
<tr>
<td>Benign:Malignant ratio</td>
<td>0.9:1</td>
</tr>
<tr>
<td>Women diagnosed with breast cancer</td>
<td>73</td>
</tr>
<tr>
<td>Detection rate per 1000 women screened</td>
<td>10.2</td>
</tr>
</tbody>
</table>

(95% CI 7.8 to 12.5)
### TABLE 5.5

**Outcome of screening by reason for referral**

<table>
<thead>
<tr>
<th>Referral reason</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>abnormal films</td>
<td>72</td>
<td>588</td>
</tr>
<tr>
<td>symptoms, normal films</td>
<td>0</td>
<td>63</td>
</tr>
<tr>
<td>abnormal films &amp; symptoms</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>technical problem</td>
<td>0</td>
<td>79</td>
</tr>
<tr>
<td>abnormal films &amp; technical problem</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>symptoms &amp; technical problem</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>abnormal films, symptoms &amp; technical</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td>73</td>
<td>759</td>
</tr>
<tr>
<td></td>
<td>832</td>
<td></td>
</tr>
</tbody>
</table>
Insurance Plan (HIP) trial the referral rate was five percent but this was averaged over all screening
rounds (Shapiro 1977, Shapiro et al 1982) and is likely to have been higher in the first round
(Kopans and Swann 1988). But in the Canada II trial where the women were aged fifty to fifty-nine
(Miller et al 1992b) the referral rate in the first screening round was 11.0%. In the first
screening round of the Sydney breast x-ray programme for women aged 45 and over, the referral
rate was 16.5% (Rickard et al 1991). The Sydney programme, like the Otago and Southland
programme, had a policy of referring symptomatic women irrespective of their screen results.

Low referral rates may mean that the test is accurately distinguishing between women who require
further assessment and those who do not. However, it may mean that the sensitivity of the test is
low and women who require further assessment have not been referred. It is difficult to combine
high sensitivity with a low referral rate in the early stages of a new breast cancer screening
programme. The referral rate in the pilot programme dropped from twelve percent to ten percent
during the first eighteen months of screening (Table 5.6) but this decline was not significant ($\chi^2$ for
trend = 0.3, $p = 0.58$). It was encouraging to observe that the proportion of mammograms initially
regarded as abnormal by one reader and leading to referral declined significantly over the first
eighteen months of the pilot programme (Appendix One).

Biopsy rate and benign to malignant biopsy ratio

A high referral rate means that too many women experience anxiety related to screening.
Fortunately in the pilot programme, although the referral rate was high, the biopsy rate and the
benign to malignant ratio were low, so that few of these women underwent unnecessary biopsies.
The biopsy rate of 1.96 percent just met the target of less than two percent, and the benign to
malignant biopsy ratio of 0.9 : 1 was well within the target range (less than 3 : 1). The low benign
to malignant biopsy ratio partly compensated for a high referral rate in that very few women in the
pilot programme underwent unnecessary biopsies. In other programmes the benign to malignant
ratio has been considerably higher; four to one in the HIP trial (Shapiro et al 1982), and two to one
in the Swedish Two-Counties trial (Tabar et al 1985) and the Edinburgh trial (Roberts et al 1989).
The Malmö and Stockholm trials had benign to malignant ratios of 0.7 and 0.5 to one respectively
5.4 Breast cancer diagnosis in the pilot programme

Information from the New Zealand Cancer Register, the National Master Patient Index (NMPI), the histology departments at Dunedin and Kew hospitals, the private pathology laboratories in Dunedin and Invercargill, and the screening register was used to calculate the incidence of breast cancer in women aged fifty to sixty-four in Otago and Southland prior to the introduction of breast cancer screening (section 5.4.1), breast cancer detection by screening (section 5.4.2), breast cancers (interval cancers) diagnosed in women in the twelve months following a negative screen (section 5.4.3), and breast cancer diagnosis outside the screening programme (section 5.4.4). The information described in sections 5.4.2 to 5.4.4 was then used to calculate the sensitivity, specificity, and positive predictive value of the screening test (section 5.4.5) as part of the evaluation of the effectiveness of the pilot programme. Information about breast cancer in women aged fifty and over was collected from the Dunedin Hospital Pathology Department computer system, and the NMPI number was used to determine whether the women were diagnosed as a result of screening or not. Information on breast histology for Southland women was produced from the Southland Hospital pathology computer system. The Southland Area Health Board did not adopt the NMPI system which meant that printouts with patient numbers could not be used to identify screened women. When a woman over fifty was diagnosed with breast cancer through the Southland hospital pathology department or one of the private pathology laboratories a photocopy of the histology form was obtained. Each form had the woman's name blanked out. Date of birth was used to look for possible matches with the screening register, and where there was a match and the name was required the woman's surgeon (who could identify the woman from her laboratory number on the histology form) asked her for permission to release her name for the evaluation. This method was cumbersome but was the only way to identify women diagnosed with breast cancer outside the programme, while still maintaining confidentiality between the pathologist and the patient. No women refused permission. With the recent passage of the Cancer Registry Act 1993, which made it compulsory for all pathologists to notify diagnoses of cancer from July 1994 (Chapter Two, section 2.1) identification of women diagnosed with breast cancer outside the screening programme should become considerably more straightforward.

The stage-distribution of cancers detected by screening was recorded so that it could be compared with that found in the prevalence screen of the Swedish Two-Counties trial (section 5.4.6). Staging
# TABLE 5.6

Referral rates from September 1991 to March 1993 for the two main film-readers

<table>
<thead>
<tr>
<th>Time period</th>
<th>Total</th>
<th>Total referred</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>%</td>
</tr>
<tr>
<td>September - October 1991</td>
<td>407</td>
<td>48</td>
</tr>
<tr>
<td>November - December 1991</td>
<td>167</td>
<td>20</td>
</tr>
<tr>
<td>January - February 1992</td>
<td>417</td>
<td>35</td>
</tr>
<tr>
<td>March - April 1992</td>
<td>785</td>
<td>63</td>
</tr>
<tr>
<td>May - June 1992</td>
<td>709</td>
<td>65</td>
</tr>
<tr>
<td>July - August 1992</td>
<td>870</td>
<td>100</td>
</tr>
<tr>
<td>September - October 1992</td>
<td>667</td>
<td>83</td>
</tr>
<tr>
<td>November 1992 - January 1993 *</td>
<td>678</td>
<td>75</td>
</tr>
<tr>
<td>February - March 1993</td>
<td>959</td>
<td>93</td>
</tr>
</tbody>
</table>

$\chi^2$ for trend = 0.3, $p = 0.58$

* no screening in December 1992

was carried out by one of the surgeons in the pilot programme, the programme manager, and the author. The TNM stage classification was used (Beahrs et al 1988).

The histological grade (Bloom and Richardson 1957, Scarff and Torloni 1968) of the tumours was assessed by the pathologists who carried out the histology for the pilot programme. Tumour size, grade, and nodal status in the pilot programme were also compared with those in the Swedish Two-Counties trial (section 5.4.6).

5.4.1 Breast cancer incidence before the pilot programme

Before the introduction of screening the recorded incidence of breast cancer in women aged fifty to sixty-four in Otago and Southland was 2.2 per thousand per year. This incidence was calculated from ten years of breast cancer registrations at the national cancer registry from 1979 to 1988 (New Zealand Cancer Registry 1990), and 1981 and 1986 census data (Table 5.7). The incidence of breast cancer in women aged fifty to sixty-four in Otago and Southland from 1979 to 1988 was similar to that for the whole of New Zealand, which was 1.9 per 1000 per year, and the incidence did not change significantly over the ten years from 1979 to 1988 ($\chi^2$ test for trend = 2.537, $p = 0.11$) (Table 5.7).

5.4.2 Breast cancer detection by screening

This is the number of women per thousand screened in whom breast cancer was detected by screening. There were seventy-three women, of 7,182 women screened in the first eighteen months of the pilot programme, who had breast cancer diagnosed through the screening programme. This is a detection rate of ten per thousand, with a 95% confidence interval of 7.8 to 12.5 (Rothman and Boice 1982), which is significantly higher than the expected incidence of 2.2 per thousand per year in the absence of screening (Table 5.7) and exceeded the target of three times the expected incidence (Chapter Four, Table 4.1). This is further discussed in Chapter Seven, section 7.2.8.

5.4.3 Interval cancers

Interval cancers are cancers diagnosed in women who had a negative screen, in the interval between the negative screen and the time of their next screen being due. These cancers may be
TABLE 5.7

Breast cancer incidence in women aged 50-64 in Otago and Southland before the establishment of the screening programme (1979-1988)

<table>
<thead>
<tr>
<th>Year</th>
<th>Breast cancer Registrations</th>
<th>No. of women aged 50-64*</th>
<th>Breast cancer Incidence#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1979</td>
<td>59</td>
<td>19755</td>
<td>2.99</td>
</tr>
<tr>
<td>1980</td>
<td>46</td>
<td>19755</td>
<td>2.33</td>
</tr>
<tr>
<td>1981</td>
<td>46</td>
<td>19755</td>
<td>2.33</td>
</tr>
<tr>
<td>1982</td>
<td>43</td>
<td>19755</td>
<td>2.18</td>
</tr>
<tr>
<td>1983</td>
<td>37</td>
<td>19755</td>
<td>1.87</td>
</tr>
<tr>
<td>1984</td>
<td>46</td>
<td>18918</td>
<td>2.43</td>
</tr>
<tr>
<td>1985</td>
<td>31</td>
<td>18918</td>
<td>1.64</td>
</tr>
<tr>
<td>1986</td>
<td>42</td>
<td>18918</td>
<td>2.22</td>
</tr>
<tr>
<td>1987</td>
<td>42</td>
<td>18918</td>
<td>2.22</td>
</tr>
<tr>
<td>1988</td>
<td>41</td>
<td>18918</td>
<td>2.17</td>
</tr>
</tbody>
</table>

$\chi^2$ test for trend = 2.537, $p = 0.11$

Breast cancer incidence (1979-88) in women aged 50-64 in Otago and Southland = 2.2 per 1000 women per year.


* Denominator for 1979-83 taken from the 1981 census, and for 1984-88 from the 1986 census.

# Breast cancer incidence per 1000 women per year.

124
cancers which were not detected at the screening test, or they may be cancers which have entered the pre-symptomatic screen-detectable phase, and then the symptomatic phase, within the interval since the negative screening test. One way to measure interval cancers is to include all women diagnosed with breast cancer within twelve months of a negative screening result. By designating interval cancers as "false negatives" it is possible to estimate the sensitivity of screening (section 5.4.5). Of the 7,182 women who were screened in the first eighteen months, 6,350 had negative mammograms, and a further 759 had a negative result after assessment. Six of these 7,109 women were diagnosed with breast cancer in the twelve months following a negative screening result. Thus the number of interval cancers was six, with a 95% confidence interval of 2.5 to 12.7 (Rothman and Boice 1982).

The difference between the number of interval cancers and the number of cancers that would otherwise have been expected in the year following screening gives an indication of the number of cancers that were detected early by screening (Day 1991, Tabar et al 1992). This comparison is best made by comparing the number of interval cancers with the number of cancers detected in the control group in a randomised controlled trial. In the absence of a control group historical data can be used, but this is not ideal as, unlike the two arms of a randomised controlled trial, the women in the two groups being compared may differ in their underlying risk of breast cancer. Because of the lack of a control group and the delay in registrations at the Cancer Registry, the historical incidence of breast cancer in women aged fifty to sixty-four in Otago and Southland was used as an estimate of the expected incidence among screened women, in the absence of screening. Each of the 7,109 women was followed for a year after her screening. In the pilot programme the expected incidence in the absence of screening was 2.2 per thousand per year. The expected incidence in 7,109 women in a year would have been sixteen. Only six women were diagnosed with breast cancer during the twelve month follow up period, so the incidence of breast cancer was 37.5 percent of that expected in a year.

This result is similar to the recent results from the National Health Service breast screening programme in Britain, where the incidence was thirty-one percent of that expected in the first year (Woodman et al 1995). However a better result was seen in the Nijmegen programme, where the proportional incidence of interval cancers was twenty-five percent (Woodman et al 1995), and the Swedish Two-Counties trial where the incidence of breast cancer in the first year following
screening for women aged fifty to fifty-nine was ten percent of that expected, and for women aged sixty to sixty-nine was seventeen percent of that expected (Tabar et al 1992). These results should be compared cautiously, since some (the Swedish results) are derived from comparisons made within randomised controlled trials while others (the Dutch and British results) are not.

5.4.4 Breast cancer diagnosis outside the screening programme

Of the 11,373 women aged fifty to sixty-four in Otago and Southland who had not been screened in the first eighteen months of the pilot programme, seventy-one were diagnosed with breast cancer within eighteen months of the start of the programme (about 48 women per year, or four per thousand per year). The number of women diagnosed with breast cancer outside the screening programme was higher than the number of women previously diagnosed each year in the whole of Otago and Southland in the years 1979 to 1988 (Table 5.7).

This incidence is clearly much higher than expected from the previous Cancer Registry figures. Three factors could have contributed to this. Firstly, the unscreened women may have had a higher underlying risk of breast cancer. However this seems unlikely since the detection rate among screened women in the pilot programme was high compared with other screening programmes. Secondly, some of the women who were not screened in the pilot programme may have had mammography outside the programme, as there was a private screening service available to Dunedin women from about the time that the pilot programme started. If this was so it would explain a higher than expected incidence of breast cancer among women outside the pilot programme. But this also seems unlikely since a survey of women who did not attend the pilot programme found that less than five percent gave their reason for non-attendance as having already had a recent mammogram (Chapter Six, Table 6.21). The third possibility is that registration of women with breast cancer has been incomplete in the past, so that the true incidence of breast cancer in Otago and Southland was higher than the published data suggested. It was thought that cancer registrations for the period 1980 to 1991 represented at least ninety-five percent of all tumours (National Health Statistics Centre 1989, New Zealand Health Information Service 1994).

However, the reviews of the Cancer Registry that were carried out in 1988 and 1991 (Cancer Registration Working Group 1988, Cancer Registry Working Group 1991) found that there had
been incomplete registration of cancers, and two studies of melanoma in New Zealand, based on pathology reports, found that there was considerable under-reporting to the Cancer Registry (National Health Statistics Centre 1989, Brown and Palmer 1991, Cooke et al 1992). It is also possible that there had been under-reporting of breast cancer in Otago and Southland. Once the Cancer Registry data for the first screening round are available it will be possible to compare the number of registrations with the number of women identified through the evaluation of the screening programme to find out if under-registration occurred during the first screening round. Another indication of past under-reporting to the Cancer Registry will be if there is an increase in registrations across all cancer sites from July 1994 when compulsory reporting from pathology laboratories was introduced.

5.4.5 Sensitivity and Specificity
Sensitivity is usually calculated as true positive screen results divided by the sum of true positives and false negatives. If interval cancers are regarded as “false negatives” it is possible to estimate the sensitivity of screening by dividing screen-detected cancers by the sum of screen-detected cancers plus interval cancers in the year following screening. This method of estimating sensitivity was described as Method One in Chapter Four, section 4.1.2. There were 7,182 women screened in the first eighteen months of the pilot programme, seventy-three of whom had breast cancer detected by screening, and six women with interval cancers diagnosed in the year following a negative screen. So the sensitivity of the screening process in the pilot programme calculated using Method One was seventy-three divided by seventy-nine, or ninety-two percent, with a 95% confidence interval of 84.4 to 96.5% (Rothman and Boice 1982). This was a very good result compared with other screening programmes, where sensitivity in the first screening round, calculated using the same method, ranged from seventy-eight percent to ninety-six percent (Chapter Four, Table 4.1). The target set for the pilot programme was eighty-five percent sensitivity.

One of the women who had an interval cancer had actually had a positive mammography result, but subsequently nothing was found at assessment. It is difficult to be certain even in retrospect, whether the initial mammogram was correct and the tumour was missed at assessment, or the assessment was correct, and a tumour only became detectable subsequently. If the initial mammogram correctly identified the tumour which was later diagnosed as an interval cancer, the
screen result was a true positive result and there were really only five interval cancers rather than six. In this scenario the sensitivity of the screening test was ninety-four percent. However, this woman did have a cancer that was diagnosed after she had been cleared by the programme, so the sensitivity of the screening process was ninety-two percent. There is no discussion of this type of problem, nor the most appropriate response to it, in the literature, and it appears that sensitivity is usually calculated with reference to the screening process, while specificity is calculated with reference to the screening test. In this situation where there is uncertainty, rather than overestimate sensitivity, and because this better represents the actual outcome of screening, the “process” sensitivity of ninety-two percent has been used, and the woman’s result has been categorised as a false negative (Table 5.8).

The proportional incidence of interval cancers can also be used to estimate the sensitivity of the screening test (Method Two, described in Chapter Four, section 4.1.2), using the formula 1 - PI
\[1\] where PI is the proportional incidence of interval cancers (Day 1985). Using this formula, the sensitivity of the test is only 62.5%, since the proportional incidence of interval cancers was calculated as 37.5% (section 5.4.3). This method of estimating sensitivity consistently produces lower estimates of sensitivity than Method One (Chapter Four, Table 4.2). Moreover, it is likely that there was under-registration of breast cancer at the Cancer Registry in the past (section 5.4.4.). Since historical data were used to estimate the expected incidence of breast cancer among screened women in the absence of screening, the expected incidence may have been underestimated. This would have caused the proportional incidence of interval cancers to be overestimated, thus underestimating sensitivity in the pilot programme. Because of this, sensitivity calculated by Method One rather than Method Two (Chapter Four, section 4.1.2) was used for the evaluation of the pilot programme.

The specificity of the pilot programme in its first eighteen months was estimated using the method described in Chapter Four (section 4.1.2) where specificity is estimated before the second screening round by dividing all negative tests (including false negatives) by the sum of all negatives and false positives (Miller 1985, Morrison 1985). This is an adequate estimate of specificity (although false negatives have been included in the numerator and the denominator) because the number of false negatives is very small in relation to the number of true negatives. In the first eighteen months of screening, 6,350 women had negative results (in other words they were not referred to the
assessment clinic). Of the 832 women who were referred for assessment, seventy-three were diagnosed with breast cancer, leaving 759 women as "false positives". So the specificity of screening in the pilot programme was 6,350 divided by 7,109 or eighty-nine percent, with a 95% confidence interval of 88.6 to 90.0% (Rothman and Boice 1982). This is lower than the specificity in the first round of screening in other programmes, which ranged from ninety-five to ninety-seven percent (Chapter Four, Table 4.1). The low specificity is partly related to the high referral rate in the pilot programme. The target set for the pilot programme was ninety-five percent specificity.

The positive predictive value, or the predictive value of a positive screening test (PPV) measures the likelihood that a woman with a positive screening test actually has breast cancer. The PPV is calculated by dividing the number of true positive tests by all positive tests, and therefore gives the proportion of positive tests that are true positives. In the pilot programme the predictive value of a positive test in the first eighteen months was nine percent, with a 95% confidence interval of 7.0 to 10.9% (Rothman and Boice 1982). Unlike sensitivity and specificity, which measure the validity of the screening test, and are independent of the prevalence of the disease, the PPV is affected by disease prevalence. As the prevalence of the disease that is being screened for increases, the PPV will increase. This should be kept in mind when comparisons of PPV are made between programmes where the prevalence of breast cancer in the screened population may differ.

The PPV is a useful measure for the clinicians working in the assessment clinic who want to provide information for women who have had positive tests. The PPV is affected by the referral rate, in that a high referral rate will tend to result in a low PPV, since there will usually be a high number of false positive tests if the referral rate is high. In other breast cancer screening programmes the PPV in the first screening round ranged from 1.5 percent to over thirty-eight percent (Hurley and Kaldor 1992). This range is determined partly by the underlying prevalence of breast cancer in the population being screened, and partly by the referral rate. Sensitivity, specificity and PPV results for the first eighteen months of the Otago and Southland pilot programme are shown in Table 5.8. The sensitivity of screening, combined with the coverage of the eligible population gives the maximum breast cancer detection that can be achieved by the screening programme. The lowest reported breast cancer detection in a successful screening programme was sixty-four percent (Chapter Four, section 4.1.2). In the Otago and Southland pilot programme the maximum breast cancer detection was sixty-eight percent (based on the
TABLE 5.8

<table>
<thead>
<tr>
<th>Breast Cancer</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>73</td>
<td>759</td>
<td>832</td>
</tr>
<tr>
<td>Negative</td>
<td>6</td>
<td>6,344</td>
<td>6,350</td>
</tr>
<tr>
<td>Total</td>
<td>79</td>
<td>7,103</td>
<td>7,182</td>
</tr>
</tbody>
</table>

Sensitivity 92%  
(95% CI = 84.4 to 96.5)  
Specificity 89%  
(95% CI = 88.6 to 90.0)  
PPV* 9%  
(95% CI = 7.0 to 10.9)

The data in this table are derived from the results of the first eighteen months of screening in the pilot programme, with twelve months follow-up for all the women screened in the first eighteen months. Sensitivity and specificity are calculated with respect to the outcome of the initial screening test. A positive test result is defined as referral to the assessment clinic. One of the women who had an interval cancer had previously had a positive mammography result, but subsequently nothing was found at assessment. To avoid an overestimate of sensitivity her result has been categorised as a false negative, with respect to the cancer which was subsequently diagnosed (section 5.4.5).

* PPV is the predictive value of a positive test.
seventy-four percent coverage and ninety-two percent programme sensitivity achieved during the first eighteen months of screening.

5.4.6 Characteristics of cancers detected at screening

If screening detects breast cancer early then there should be a higher proportion of early stage disease among screen-detected cancers than among non-screen-detected cancers. It is interesting to compare the stage distribution of cancers detected in the pilot programme with those detected in existing programmes, but it is uncertain whether this can be used as an indicator of effectiveness since the stage distribution is likely to be affected by length, selection, and over-diagnosis biases, and the lead-time resulting from screening could cause a shift in stage distribution even if there were no effect on breast cancer mortality. The stage-distribution of screen-detected cancers has been suggested as a surrogate endpoint for reduced mortality (Bull and Mountney 1991), on the basis that in order to be effective, screening programmes must detect breast cancers earlier than they would otherwise have been detected. It is interesting to compare the stage distribution of tumours detected in the pilot programme with the Swedish Two-Counties trial, where there was a proven mortality reduction. But stage distribution is vulnerable to the biases described above. This means it can only be used as a rough indicator of the likely effect of the programme on breast cancer mortality, and it should be interpreted cautiously. The stage-distribution of cancers detected in the first eighteen months of the Otago and Southland pilot programme is shown in Table 5.9.

In the Swedish Two-Counties programme 33.3 percent of tumours detected at the prevalence screen in women aged 50-59 were stage two or higher, and 34.3 percent in women 60-69 were stage two or higher (Day et al 1989). In the Otago and Southland programme three tumours could not be staged. Depending on the status of these three tumours, between thirty-eight (95% CI 28.1% to 49.8%) and forty-two percent (95% CI 31.8% to 53.9%) of the tumours detected in the first eighteen months were stage two or higher (Table 5.9). The diagnosis of ductal carcinoma in situ (DCIS) is made only rarely in the absence of mammographic screening (de Koning et al 1990) but twelve percent of tumours diagnosed in the first eighteen months of screening in the pilot programme were DCIS. Diagnosis of these very early tumours, where the prognosis is uncertain since they were previously rare, has led to concern about over-diagnosis in screening programmes. The issue of over-diagnosis in the pilot programme is discussed in section 5.4.7.
TABLE 5.9

Stage distribution of tumours detected in the Otago and Southland programme

<table>
<thead>
<tr>
<th>Stage</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ductal carcinoma in situ</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Stage I</td>
<td>34</td>
<td>49</td>
</tr>
<tr>
<td>Stage II</td>
<td>24</td>
<td>34</td>
</tr>
<tr>
<td>Stage III</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Stage IV</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Unknown</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>100</td>
</tr>
</tbody>
</table>
TABLE 5.10

Breast cancers detected in the Otago and Southland prevalence screen compared with the Swedish Two-Counties prevalence screen:

<table>
<thead>
<tr>
<th>Size (mm)</th>
<th>O&amp;S (°/o)</th>
<th>S2C (°/o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-9</td>
<td>19</td>
<td>74</td>
</tr>
<tr>
<td>10-14</td>
<td>15</td>
<td>84</td>
</tr>
<tr>
<td>15-19</td>
<td>12</td>
<td>56</td>
</tr>
<tr>
<td>20-29</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>30-49</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>50+</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>not known</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

($\chi^2 = 1.9, df = 5, p = 0.86$)

<table>
<thead>
<tr>
<th>Node status</th>
<th>O&amp;S (°/o)</th>
<th>S2C (°/o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive</td>
<td>21</td>
<td>56</td>
</tr>
<tr>
<td>negative</td>
<td>36</td>
<td>213</td>
</tr>
<tr>
<td>not known</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

($\chi^2 = 6.5, df = 1, p = 0.01$)

<table>
<thead>
<tr>
<th>Histological grade</th>
<th>O&amp;S (°/o)</th>
<th>S2C (°/o)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>not known</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

(excludes 8 DCIS)
The distribution of tumour size for the first eighteen months of the pilot programme was similar ($\chi^2 = 1.9$, df = 5, $p = 0.86$) to that for the prevalence screen of the Swedish Two-Counties trial (Table 5.10). The programmes differed for nodal status, with a higher proportion of node-negative tumours detected in the Swedish Two-Counties trial ($\chi^2 = 6.5$, df = 1, $p = 0.01$). No comparison was made for grade, partly because some of the tumours detected in the operation of the pilot programme had not been graded, but also because in the Swedish Two-Counties trial the distribution of grades differed significantly between Östergötland and Kopparberg, the two counties in the trial. It was suggested that this was due to "the subjective nature of histological grading" (Tabar et al 1992). This is of considerable concern because if grading is this subjective its use as a prognostic indicator must be in question.

It has been suggested that tumour grade, size and nodal status could be used to predict the effect of screening on breast cancer mortality (Duffy et al 1991, Tabar et al 1992). Duffy et al collected information on women diagnosed with breast cancer in the Swedish Two-Counties trial. Data were collected from the prevalence screen on women diagnosed with breast cancer in the group offered screening (screen-detected cancers, interval cancers, and cancers diagnosed in women who refused screening) and in the control group. Incidence screen data were collected for the period from immediately after a screening test to immediately after the subsequent screening test. Because the screening test was of high sensitivity, cancers diagnosed in this period would not be greatly affected by length bias. These cancers plus breast cancers diagnosed in women who refused the offer of screening were then compared with breast cancer diagnosed in the control group over the same period. Cancers in the two groups were equivalent except that those in the group offered screening were diagnosed earlier overall. The effect of earlier diagnosis on tumour size, stage, and grade could then be assessed without being distorted by length bias. The survival of women with screen-detected breast cancer was better than that of women with breast cancer in the control group or women with interval cancers, and the survival of women with breast cancer who had refused screening was worst. Size, nodal status, and grade explained much of the difference in survival between women with incidence screen-detected, interval and control group breast cancers. These three factors explained much less of the difference between screen-detected cancers, interval cancers and control group cancers in the prevalence screen where the effect of length bias is greatest (Duffy et al 1991). The screen-detected tumours were smaller and of a more favourable
grade than those diagnosed in the control group and Duffy et al hypothesised that tumour grade worsens as the size of the tumour increases.

These results led to the development of targets for the size, nodal status and grade of tumours detected in breast screening programmes (Tabar et al 1992). The suggested targets are listed in Table 5.11. The results from the prevalence screen of the Otago and Southland programme are given for each of the targets, with ninety-five percent confidence intervals where appropriate.

Comparisons such as that made in Table 5.11 should be interpreted with caution because a similar study was carried out using data from the United Kingdom trial of early detection of breast cancer (Moss et al 1994). In this study only about a third of the improved prognosis in women who were diagnosed by screening could be explained by changes in size and nodal status. It was suggested that the difference between the British and Swedish results could be partly because the British analysis could not adjust for tumour grade. Another possible reason for the discrepancy between the two studies is that participation in screening was lower in the United Kingdom and there was self-selection of women with a poor prognosis into the non-participant group (Moss et al 1994).

5.4.7 Over-diagnosis

The amount of over-diagnosis in a screening programme can be estimated by examining the deficit in the number of cancers diagnosed in the interval after screening (compared with that expected in the absence of screening), and comparing this to the total number of cancers diagnosed at screening. The most accurate way of determining the deficit in interval cancers is where breast cancer incidence following a negative test in screened women can be compared with the incidence in a comparison group (ideally in a randomised controlled trial, but in the absence of a comparison group a historical comparison may be used). The extent to which the screening prevalence exceeds the interval deficit indicates the extent to which cancers diagnosed at screening would not have surfaced in the time interval considered (Day 1991).

Since the lead time gained by screening in women over fifty has been estimated at about two years (Day and Chamberlain 1988), it is not appropriate to use this method to estimate over-diagnosis in the Otago and Southland pilot programme until the second screening round (when data on two years follow up after screening will be available).
**TABLE 5.11**

**Suggested targets for size, histological grade, and nodal status for tumours detected in breast cancer screening programmes**

<table>
<thead>
<tr>
<th>Suggested Target</th>
<th>Otago &amp; Southland first screen result</th>
<th>95% confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>At least 50% of screen-detected tumours should be less than 15mm diameter</td>
<td>59%</td>
<td>46% to 72%</td>
</tr>
<tr>
<td>At least 30% of grade 3 tumours should be less than 15mm.</td>
<td>Not assessed</td>
<td>(only 4 detected)</td>
</tr>
<tr>
<td>At least 70% should be node negative</td>
<td>63%</td>
<td>50% to 76%</td>
</tr>
<tr>
<td>These targets should be met with a recall rate of less than 9%</td>
<td>10.4% recall rate</td>
<td></td>
</tr>
</tbody>
</table>

5.5 Summary

The results of the evaluation of the effectiveness of the pilot programme are summarised briefly here. Detailed comparisons with other screening programmes, implications, and recommendations for the pilot programme and for a national breast cancer screening programme in New Zealand are discussed in Chapter Seven. The pilot programme met the targets that had been set for identification of eligible women and coverage of the target population. An appropriate invitation method was developed, which resulted in a seventy-one percent response rate to an invitation. The technical quality of the mammography in the pilot programme was high, and double-reading of films increased sensitivity. The referral rate of eleven percent (or 10.4 percent excluding women with symptoms) in the pilot programme was higher than that reported in many breast cancer screening trials for the first round of screening and exceeded the target of between five and ten percent. A high referral rate means that too many women experience unnecessary anxiety related to screening. Fortunately in the pilot programme, although the referral rate was high, the biopsy rate and the benign : malignant biopsy ratio were low. The biopsy rate of 1.96 percent just met the target of less than two percent, and the benign : malignant ratio of 0.9 : 1 was well within the target range (less than 3 : 1). The low benign : malignant ratio partly compensated for a high referral rate in that very few women in the pilot programme underwent unnecessary biopsies. The pilot programme easily met the target set for sensitivity, but did not meet the target set for specificity (due in part to the high referral rate in the first eighteen months of screening). The implications of this, and recommendations for the future are discussed in Chapter Seven.

The stage-distribution of cancers detected in the Otago and Southland programme was recorded (Table 5.9). It has been suggested that, in the prevalence screen of breast cancer screening programmes, not more than forty percent of cancers detected should be stage two or higher (Day et al 1989). In the Otago and Southland programme between thirty-eight (95% CI 28.1% to 49.8%) and forty-two percent (95% CI 31.8% to 53.9%) of the tumours detected in the first eighteen months were stage two or higher (section 5.4.6). The suitability of the stage distribution of screen-detected cancers as an indicator of the effectiveness of screening is discussed further in Chapter Seven.
CHAPTER SIX

The Acceptability of the Pilot Programme

Introduction

An important factor affecting the success of any screening programme is the level of participation among the target population. When participation is high the benefits from screening are available to a greater proportion of eligible women. The efficiency and cost effectiveness of screening are also improved by high participation levels (Sickles et al 1986, McLelland 1987, UK Trial of Early Detection of Breast Cancer Group 1988). Participation depends on several factors, including awareness of the service, perception of its relevance, and its acceptability (Strax 1980). Several studies have investigated factors which affect participation in mammographic screening programmes (French et al 1982, Calnan 1984a, Calnan 1984b, Roberts et al 1986, Hobbs et al 1980, Fink et al 1968, Leather and Roberts 1985, MacLean et al 1984). Participation in screening programmes has been associated with younger age, higher socioeconomic group, higher educational level, greater knowledge and beliefs about breast cancer, and previous health-related behaviour (such as attending for cervical smears), but results vary according to region.

The acceptability of the programme could affect not only the continued participation of the women who had already taken part, but also that of women who had not yet attended but heard about the programme “by word of mouth”. The acceptability of the pilot programme was one of the three key areas (acceptability, effectiveness, and economic efficiency) that were recommended for evaluation by the Skegg committee (Skegg et al 1988). Several surveys were undertaken to find out about the acceptability of the pilot programme. These surveys are described in this chapter.
6.1 Pre-screening survey: factors likely to affect participation in mammographic screening.

Before the pilot programme began, a random sample of 290 Otago and Southland women was surveyed to find out about aspects of a programme which would make it more acceptable and thereby enhance participation (Richardson 1990). The aim was to use the findings from this survey to help design a pilot programme that would be acceptable and would thereby encourage participation. This survey was carried out before the pilot programme (and its evaluation) started, but the results were used in planning the pilot programme and the evaluation, so it is included here to provide background information.

Methods

The survey population was women aged between forty and seventy who lived in Otago and Southland. This age range was chosen because the recommendations for screening in New Zealand (Skegg et al 1988) had not been released when the study was designed, but it was recognised that eligibility was likely to be somewhere within that range.

A random sample of a thousand women whose names appeared on the electoral roll was contacted by mail. Each woman was asked to indicate whether she was in the required age range for the study, since the study was carried out before the electoral roll centre was able to provide lists of voters by age-group for health-related purposes (Chapter Four, section 4.1.1). Each woman was asked whether she would be prepared to answer a postal questionnaire. Follow up phone calls were used to contact women who did not reply, or who were within the required age range but were not prepared to answer a postal questionnaire. (The second group was asked if they would be prepared to answer the questionnaire by telephone.)

Replies were received to sixty-five percent of the initial letters. Fifty-five percent (192) of those who did not reply were contacted by telephone. Of all women contacted forty-two percent (353) were in the eligible age range, and of these eighty-seven percent (307) agreed to complete a postal questionnaire. Of forty-six women who were not prepared to complete a postal questionnaire, sixteen agreed to complete the questionnaire by telephone. The results from the 290 questionnaires which were returned were analysed using SPSS-X (Statistical Package for Social Sciences, 1990).
Results

The 290 respondents were representative of women in the forty to seventy age range in Otago and Southland according to the 1986 census, in terms of age, ethnic group, domicile, and marital status (Table 6.1). However women who had attended university were over-represented; fourteen percent of the respondents had attended university at some time compared with six percent of Otago and Southland women in the age range.

Eighty-four percent (244) of the women surveyed intended to participate in screening. Eight percent stated that they would participate according to certain conditions; the most common were distance to the screening centre, appointment time, and finding time to attend. Eight percent did not intend to participate. Of those not intending to participate, half were over sixty, although women in this age group made up only a third of the women surveyed.

Knowledge about breast cancer and mammographic screening was scored according to the women's responses to five true/false statements about breast cancer and mammographic screening (the expected response is given in parentheses after each statement):

1. most lumps in the breast are harmless (true)
2. the risk of breast cancer increases with age (true)
3. breast cancer is common in young women (false)
4. early diagnosis of breast cancer increases the chance of cure (true)
5. mammographic screening can detect breast cancer early (true)

The more women knew about breast cancer and mammographic screening the more likely they were to intend to participate in a screening programme (Table 6.2).

Health related behaviour (such as attending a dentist regularly, or having a cervical smear within the last three years) was associated with an intention to participate in mammographic screening (Table 6.2). Women who knew someone who had been diagnosed with breast cancer were more likely to intend to be screened. A greater proportion of those intending to be screened had a family history of breast cancer. In contrast to the previous factors, this last finding was not significant.
### TABLE 6.1

Demographic characteristics of eligible respondents and women aged forty to seventy in Otago and Southland at the 1986 census

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Women surveyed</th>
<th>O &amp; S women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. (%)</td>
<td>(%)</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-44</td>
<td>64 (22)</td>
<td>19</td>
</tr>
<tr>
<td>45-49</td>
<td>58 (20)</td>
<td>17</td>
</tr>
<tr>
<td>50-54</td>
<td>41 (14)</td>
<td>15</td>
</tr>
<tr>
<td>55-59</td>
<td>43 (15)</td>
<td>16</td>
</tr>
<tr>
<td>60-64</td>
<td>46 (16)</td>
<td>16</td>
</tr>
<tr>
<td>65-69</td>
<td>38 (13)</td>
<td>17</td>
</tr>
<tr>
<td><strong>Ethnic group</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maori</td>
<td>3 (1)</td>
<td>2</td>
</tr>
<tr>
<td>Non Maori</td>
<td>287 (99)</td>
<td>98</td>
</tr>
<tr>
<td><strong>Domicile</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dunedin</td>
<td>118 (41)</td>
<td>38</td>
</tr>
<tr>
<td>Other urban</td>
<td>113 (39)</td>
<td>41</td>
</tr>
<tr>
<td>Rural</td>
<td>59 (20)</td>
<td>21</td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never married</td>
<td>7 (2)</td>
<td>5</td>
</tr>
<tr>
<td>Married</td>
<td>230 (79)</td>
<td>74</td>
</tr>
<tr>
<td>Separated</td>
<td>10 (3)</td>
<td>3</td>
</tr>
<tr>
<td>Widowed</td>
<td>29 (10)</td>
<td>13</td>
</tr>
<tr>
<td>Divorced</td>
<td>14 (5)</td>
<td>4</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University attendance</td>
<td>40 (14)</td>
<td>6</td>
</tr>
<tr>
<td>Trade qualification</td>
<td>24 (8)</td>
<td></td>
</tr>
<tr>
<td>Secondary school</td>
<td>199 (69)</td>
<td>} 94</td>
</tr>
<tr>
<td>Primary school</td>
<td>26 (9)</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 6.2

Factors affecting intention to participate in mammographic screening

<table>
<thead>
<tr>
<th>Intended Participation</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40-49</td>
<td>107</td>
<td>88</td>
</tr>
<tr>
<td>50-59</td>
<td>74</td>
<td>88</td>
</tr>
<tr>
<td>60-69</td>
<td>59</td>
<td>74</td>
</tr>
<tr>
<td>Knowledge¹</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>118</td>
<td>91</td>
</tr>
<tr>
<td>&lt;5</td>
<td>123</td>
<td>79</td>
</tr>
<tr>
<td>Cervical smear request²</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>137</td>
<td>90</td>
</tr>
<tr>
<td>No</td>
<td>104</td>
<td>78</td>
</tr>
<tr>
<td>Friend/acquaintance with breast cancer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>154</td>
<td>89</td>
</tr>
<tr>
<td>No</td>
<td>84</td>
<td>77</td>
</tr>
</tbody>
</table>

¹ Knowledge was scored according to the women’s responses to five true/false statements about breast cancer and mammographic screening

² Women who had requested a cervical smear within the previous three years

142
TABLE 6.3

Preferred location for a screening centre (women who intend to be screened)

<table>
<thead>
<tr>
<th>Location</th>
<th>Dunedin women (%)</th>
<th>Other women (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile unit</td>
<td>11</td>
<td>79</td>
</tr>
<tr>
<td>Hospital</td>
<td>47</td>
<td>34</td>
</tr>
<tr>
<td>Screening centre</td>
<td>43</td>
<td>23</td>
</tr>
</tbody>
</table>

143
**TABLE 6.4**

Principal reason suggested by respondents that women may not participate in screening

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fear</td>
<td>83</td>
<td>42</td>
</tr>
<tr>
<td>Embarrassment</td>
<td>60</td>
<td>31</td>
</tr>
<tr>
<td>Lack of knowledge</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Felt healthy</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Other</td>
<td>37</td>
<td>19</td>
</tr>
</tbody>
</table>
Seventeen percent of the women said they would prefer to attend outside normal working hours (this was particularly the case for city women, where twenty percent said they would prefer appointments outside working hours). Only twenty-eight percent of respondents felt it would be reasonable to expect women to travel more than thirty minutes to get to a screening centre. Women who lived outside Dunedin city and who intended to participate said they would prefer to be screened at a mobile unit (Table 6.3).

Thirty-five percent of the women said they would prefer to be screened by female staff while 65% did not think the gender of staff was important. Among women who were not definitely intending to be screened, fifty-eight percent preferred female staff. Intention to participate was not associated with state of health (women who rated their health as good or excellent were as likely to intend to participate in screening as women who rated their health as fair or poor), occupation, educational level, or socioeconomic group. Neither were these factors associated with having had a cervical smear within three years. Seventy-one percent of respondents thought that some women would decline an invitation to mammographic screening. Respondents were asked to suggest a reason for this. One hundred and ninety-six women (sixty-eight percent) suggested possible reasons. The most common reasons given were fear (of the procedure or the outcome) or embarrassment (Table 6.4).

Discussion
The pre-screening survey identified some features of a screening service and of eligible women that were likely to affect participation. Intention to participate varied according to response group. Women who replied to the initial letter and were prepared to complete the questionnaire had the highest level of intended participation (eighty-nine percent). Seventy-two percent of women who responded after a telephone reminder intended to participate. Only forty percent of women who had not wished to complete a postal questionnaire but completed the telephone questionnaire intended to participate.
The survey results indicated that the screening centre should be within thirty minutes travel by car. For planning purposes this meant that mobile units would be required to screen women outside main centres. The staff at the screening centre would preferably be female. This was especially important for women who were undecided or thought that they might not participate in a screening programme. The results suggested that some appointment times outside normal working hours would be needed. In Dunedin approximately a fifth of women who answered the questionnaire said they would prefer appointments outside normal working hours.

It was planned that publicity and information programmes about screening would provide information so that women knew what to expect when they attended for screening. Fear (of embarrassment associated with screening, or of the outcome of screening) was suggested as the main reason for nonparticipation; publicity about the screening programme therefore endeavoured to (a) reduce fear of the outcome by emphasising that screening would be negative for the majority of women screened, and that early detection and treatment increase the chance of cure, and (b) reduce women’s expectation of embarrassment by explaining what the procedure involved and by assuring privacy.

Intention to participate and knowledge about breast cancer and mammographic screening decreased with age. It was kept in mind that it might be necessary to specifically target women over sixty. Women outside Dunedin were more likely to state that their participation would be influenced by the location of a screening centre. A mobile screening centre was included in the pilot programme so that screening would be more accessible to women outside the main centres. Women whose knowledge about breast cancer and mammographic screening was good, were more likely to intend to participate. This suggested that education programmes could have a significant effect in increasing participation.

The pilot programme was designed to incorporate the features that were identified in this survey as likely to affect participation, with a mobile unit so that rural women did not have to travel too far for screening, some evening sessions, female radiographers and clerical staff, and publicity and educational leaflets about the programme. Publicity was targeted to specific areas such as rural towns in the month before the invitations to attend the mobile unit were sent out.
6.2 Participation in screening

One of the most important determinants of the effectiveness and efficiency of a screening programme is the involvement of the target group. When participation is high, the benefits of screening are available to more of the eligible women. In the evaluation of the pilot programme the participation rate was defined as the percentage of invited women who were screened as a result of an invitation. Coverage was defined as the percentage of the target population screened. The two measures were required since some women were screened without having received an invitation. In order to approach or exceed the lowest recorded coverage for successful programmes, the pilot programme had a target participation rate of seventy-eight percent or higher in the first screening round (Chapter Four, Table 4.1). With over ninety percent identification of eligible women, so that ninety percent of the eligible population could be invited, a participation rate of seventy-eight percent would result in seventy percent coverage (seventy percent of women in the target population being screened).

In screening programmes which had access to population registers, and where invitations were sent out to all women on the register, participation has been defined as the percentage of women on the register who attended for screening (Breast Cancer Screening Working Group 1987). This is not exactly the same as the definition used in the evaluation of the Otago and Southland pilot programme. In many published reports from breast cancer screening programmes it is not clear whether the reported participation rates include self-referred women, although in the United Kingdom national breast cancer screening programme the percentage of eligible women screened, which included self-referred women was called “acceptance” rather than participation (Chamberlain et al 1993). Discrepancies in the way participation rates are calculated mean that using participation as the only measure of the response to screening in the pilot programme could be misleading. By calculating two measures, participation and coverage, both the response to an invitation and the percentage of eligible women who were screened in the Otago and Southland pilot programme were examined.

Methods

Each month information was collected on the number of women invited, the number responding, the number accepting, the number declining, the number not responding, the number of invited
women screened, the number of booked women who did not attend, and the total number of women screened. This information was collected manually, from the appointment books at the screening centres, because although it had been specified as part of the evaluation download, the data was not available from the computer system. The manual system that was devised was very time-consuming, but was the only means to collect information on participation, given that the computer system could not provide the required information. On every invitation that was sent out there was a suggested appointment time, and the woman’s name was written in the appointment book for this time, with the initial “I” against her name to identify her as a woman who had been sent an invitation. If a woman phoned to change her appointment her name was removed from the original appointment and inserted at the new appointment time and also identified with an “I”. If a woman had not replied to her invitation, a self-referred woman was sometimes given that appointment time, but this was added at the bottom of each page, so that the information about the number of invited women who declined, cancelled, or did not attend could still be collected. Women who self-referred were given an appointment time, and their names were flagged with the initials “SR” in the appointment book. Similarly women who had been referred by general practitioners were identified with “GP” in the appointment book. Women who attended were checked off in the book, and women who did not attend were identified with a cross against their names. Each month a member of the evaluation team would extract the data about the number of women who attended or did not attend in each of the three categories (invited, self-referred, and GP-referred) from the appointment books.

Although there were limitations to this manual system, which are discussed below with the results, it was the only way that it was possible to estimate how many women were screened as a result of an invitation, were self-referred, or were referred by their general practitioners each month. It was also the only way to identify women who had declined an invitation, so that they could be included in the survey of non-attenders (section 6.5). This survey of non-attenders was a very important part of the evaluation, to discover whether there were aspects of the screening programme that could be changed to encourage such women to attend.

The coverage of the programme (percentage of eligible women screened) could not be calculated until the end of the first screening round since the total number of women screened in the first round was not available until then. Coverage was calculated by dividing the number of women
screened in the whole of the first screening round by the number of women who were eligible for screening in that round according to the census.

The participation rate was the number of invited women screened divided by the number of women invited, and this measured the response to the invitations. When the evaluation targets were set, it was thought that a very high participation (seventy-eight percent) would be required to compensate for the fact that not all women would be identified and invited, since some eligible women would be on neither the electoral roll nor on general practice registers. It had been expected that nearly all the women who attended for screening would do so in response to an invitation. But in the Otago and Southland programme there was a constant level of self-referral (women who requested screening although they had not received an invitation). As well as self-referred women there were also women who were referred by their general practitioners or through the Radiology Department of Dunedin Hospital. This meant that the number of women who were screened as a result of an invitation was constantly supplemented by women who had self-referred or had been referred. It also meant that by the time some women received their invitations, they had already been screened.

Results
The participation rate (response to an invitation) was fifty-eight percent (Table 6.5). This was the percentage of women who, having been sent an invitation in the first eighteen months of screening, had attended screening (women who were invited during the eighteenth month of screening had to have attended within the following month to be included). It does not include women who attended before being invited; these women were designated “self-referred” or “general practitioner-referred”. The names of women who had self-referred or been referred by a general practitioner were crossed off the electoral roll so that they would not be invited again in the first screening round. General practitioners also removed the names of women who had already attended the programme from the screening lists derived from their age-sex registers. It is possible nonetheless, that some invitations could have been sent out inappropriately (for instance if a self-referred woman had changed her address since the electoral roll was updated). This would cause the true participation rate to be underestimated. Also any woman who declined an invitation but later changed her mind and made an appointment would be categorised as “self-referred” for that appointment. This would cause the number of self-referrals to be overestimated. A third possible problem is that women who were very enthusiastic about screening may have self-referred early in
the programme, while the women who waited to receive an invitation may have been less 
enthusiastic about screening. The response of these women to an invitation could therefore 
underestimate the true response for all eligible women. The target was that seventy-eight percent 
of women who received an invitation would be screened. This was on the assumption that about 
ninety percent of women would be accurately identified, so with a participation rate of about 
seventy-eight percent, seventy percent of the women living in Otago and Southland aged fifty to 
sixty-four would be screened. In fact, the participation rate was considerably below the target that 
had been set (Chapter Four, Table 4.1).

At the end of the first screening round 14,129 women had been screened, out of a total of 18,555 
women aged fifty to sixty-four and living in Otago and Southland at the 1991 census. Using the 
1991 census figure as the denominator gave the programme seventy-six percent coverage of the 
eligible women in Otago and Southland. However, some women aged into the eligible age-range 
during the first screening round, thus increasing the number of eligible women, and during the same 
period some women who were aged sixty-three and sixty-four in 1991 aged out of the eligible age­
range, thus reducing the number of eligible women. In 1991 there were 2,886 women in Otago 
and Southland aged forty-eight and forty-nine. These women would have entered the eligible age­
range over the following two years. However, there were 2,382 women aged sixty-three and sixty­
four at the 1991 census, who would have become ineligible over the following two years. The 
result is a 2.7 percent increase (504 women) over two years in the number of women eligible for 
screening. If this increase in the denominator is taken into account the coverage is seventy-four 
percent.

Discussion

The target coverage of seventy percent of fifty to sixty-four year old women living in Otago and 
Southland screened had been met by the end of the first screening round, but this was despite a 
participation rate that was well below the target set. By sending personal invitations signed by
TABLE 6.5

Participation rate* in response to an invitation in the first eighteen months of screening.

<table>
<thead>
<tr>
<th>Screening centre</th>
<th>No. invited</th>
<th>No. of invited women screened</th>
<th>Participation rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dunedin</td>
<td>4,596</td>
<td>2,385</td>
<td>51.9%</td>
</tr>
<tr>
<td>Invercargill</td>
<td>3,258</td>
<td>2,003</td>
<td>61.5%</td>
</tr>
<tr>
<td>Mobile #</td>
<td>1,803</td>
<td>1,188</td>
<td>65.9%</td>
</tr>
<tr>
<td>TOTAL</td>
<td>9,657</td>
<td>5,576</td>
<td>57.7%</td>
</tr>
</tbody>
</table>

* the participation rate is calculated as the percentage of invited women who attended as the result of an invitation issued in the first eighteen months of screening in the Otago and Southland pilot programme.

# the mobile unit had only been operating for six months during this period.
general practitioners it was found that a seventy-one percent participation rate could be achieved (Chapter Five, section 5.1.3). However, because of the consistently high level of self-referral throughout the first screening round, a target for participation in the Otago and Southland programme became almost irrelevant once the programme started. When the pilot programme was being planned the very high level of self-referral and referral by general practitioners that occurred throughout the first screening round was not anticipated. This meant that many women were screened despite never receiving an invitation, and some women would already have been screened by the time they received an invitation. The implications of this are discussed further in Chapter Seven.

Initially the programme functioned under its expected capacity, because of problems recruiting staff, and also because there were unavoidable delays in commissioning the mobile unit, which did not start screening until a year after the programme started. Despite these difficulties, the programme still managed to exceed the target of screening seventy percent of the eligible population in the first screening round.

6.3 Acceptability of Screening

A survey was carried out to measure satisfaction with the screening service, to provide feedback for programme staff, and to obtain an indication of future participation in the programme. Information about the costs to participating women was also collected so that it could be used later in the economic analysis of the pilot programme.

Methods

The questionnaire was finalised after a draft questionnaire had been piloted on the first 140 women to participate in the screening programme. Screening at the fixed units in Dunedin and Invercargill started in September 1991 and November 1991 respectively, and screening on the mobile unit started later, in September 1992. A random sample of ten in every fifty women screened at the fixed units was selected by computer-generated random numbers based on the order in which the women were screened during February to April 1992. This “urban” sample was sent a self-administered questionnaire (Appendix Six) with a reply-paid envelope. The proportions of
screened women sampled were the same for each of the two fixed units. Women who did not respond within a month were sent a follow-up letter and a second copy of the questionnaire. Of 168 women sent questionnaires, 156 (ninety-three percent) replied. The results from this sample can be interpreted allowing a margin of error of up to ten percent with 95% confidence. The initial survey was carried out early in the course of the programme so that rapid action could be taken if any major problems were identified. The survey was designed to be repeated during the course of the pilot programme.

The same questionnaire was subsequently posted to all the women who attended the mobile screening unit in its first three months of operation (September to November 1992) to give the 'rural' sample. The same follow up reminders were used. Questionnaires were sent to 306 women and 286 (ninety-three percent) replied. Results were analysed by cross tabulation and $\chi^2$ statistics.

Results
Replies were received from 156 'urban' and 286 'rural' women, the response rates being the same (ninety-three percent) in each of these groups.

Demographic characteristics:
Ninety percent of the respondents answered all of the demographic questions (Table 6.6). The urban and rural samples of women were similar in age and ethnic group. They differed in education, with fewer rural women having attended university, while more had trade or vocational qualifications. Their occupations also differed; fewer rural women were in paid employment but more were self-employed. The women were asked about their yearly household income before tax. Three categories which corresponded to the categories used by the government to assess eligibility for community support were used in the questionnaire. The rural women had significantly lower household incomes than the urban women (although it is difficult to assess farm incomes in this way).

Participation in the screening programme
Twenty percent of the women were self-referred (i.e. attended screening without having received an invitation) or had been referred by their doctors without an invitation. The other eighty percent
## TABLE 6.6

Demographic characteristics of urban and rural participants in the Otago and Southland breast cancer screening pilot programme

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Urban Women</th>
<th></th>
<th>Rural Women</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number</td>
<td>Percent</td>
<td>Number</td>
<td>Percent</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>156</td>
<td>100</td>
<td>286</td>
<td>100</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-54</td>
<td>54</td>
<td>35</td>
<td>118</td>
<td>41</td>
</tr>
<tr>
<td>55-59</td>
<td>47</td>
<td>30</td>
<td>91</td>
<td>32</td>
</tr>
<tr>
<td>60-64</td>
<td>55</td>
<td>35</td>
<td>77</td>
<td>27</td>
</tr>
<tr>
<td>((x^2 = 3.6, \ \text{df} = 2, \ p = 0.17))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ethnic group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maori</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Non Maori</td>
<td>148</td>
<td>98</td>
<td>272</td>
<td>98</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>13</td>
<td>9</td>
<td>21</td>
<td>8</td>
</tr>
<tr>
<td>Secondary school</td>
<td>102</td>
<td>71</td>
<td>208</td>
<td>75</td>
</tr>
<tr>
<td>University</td>
<td>16</td>
<td>11</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Trade qualification</td>
<td>12</td>
<td>8</td>
<td>41</td>
<td>15</td>
</tr>
<tr>
<td>((x^2 = 18.2, \ \text{df} = 3, \ p &lt; 0.001))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Household income</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(before tax)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>less than $15,000</td>
<td>45</td>
<td>33</td>
<td>118</td>
<td>46</td>
</tr>
<tr>
<td>$15,000 - $30,000</td>
<td>56</td>
<td>41</td>
<td>99</td>
<td>38</td>
</tr>
<tr>
<td>more than $30,000</td>
<td>36</td>
<td>26</td>
<td>41</td>
<td>16</td>
</tr>
<tr>
<td>((x^2 = 8.7, \ \text{df} = 2, \ p = 0.01))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paid employment</td>
<td>52</td>
<td>35</td>
<td>48</td>
<td>17</td>
</tr>
<tr>
<td>Self-employed</td>
<td>9</td>
<td>6</td>
<td>51</td>
<td>18</td>
</tr>
<tr>
<td>No paid employment</td>
<td>32</td>
<td>21</td>
<td>77</td>
<td>28</td>
</tr>
<tr>
<td>Retired</td>
<td>57</td>
<td>38</td>
<td>103</td>
<td>37</td>
</tr>
<tr>
<td>((x^2 = 24.8, \ \text{df} = 3, \ p = 0.001))</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
were screened because they had received an invitation from the programme. Half of the women said that they first found out about the programme when they received the invitation to take part. The women were asked whether they would have arranged to have a mammogram if they had not been invited to the screening programme; fewer than half (forty-seven percent) said that they would have, and only a third (thirty-two percent) said they would have had mammograms if they had had to pay for them. There was no difference between urban and rural women for any of these.

Anxiety related to screening

The women were asked how they felt when they received the invitation to be screened (Table 6.7a). While most women were not worried or only slightly worried by the invitation, significantly more of the urban women (thirteen percent) were ‘very worried’ by the invitation compared with less than one percent of rural women (Table 6.7a). For both rural and urban women there was no difference in the anxiety experienced during screening or while awaiting the results of screening so the results from the two groups were combined (Table 6.7b). Anxiety was not related to a history of breast cancer in family or close friends ($p = 0.06$, Table 6.8). There was no significant difference in anxiety related to screening for women who had had family or close friends with breast cancer compared with women who had not. Although there was a downward trend in the percent not at all worried with time until results were received, there was no significant relationship between anxiety and the time that it took for the women to receive their results ($p = 0.7$, Table 6.9). The women who worried most while waiting for their results were the women who had also worried most about the screening invitation ($p < 0.001$, Table 6.10).

Experience of the screening test

Most women found the mammogram uncomfortable rather than painful (Table 6.11). There was no difference between screening centres nor between urban and rural women in their experience of the screening test. Eleven women (three percent) had severe bruising after the mammogram. Women with bruising were significantly more likely to have found the mammogram painful ($p <0.001$). Results for urban women are shown in Tables 6.12 and 6.13. The level of anxiety before the screening test was not related to the amount of discomfort felt ($p = 0.12$, Table 6.12), but anxiety while awaiting the results of the screening test was positively related to the amount of discomfort that was experienced during the test ($p = 0.04$, Table 6.13).
TABLE 6.7

Anxiety related to screening

(a) Anxiety on receiving invitation to be screened

<table>
<thead>
<tr>
<th></th>
<th>Urban Number</th>
<th>(%)</th>
<th>Rural Number</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all worried</td>
<td>81</td>
<td>(55)</td>
<td>139</td>
<td>(63)</td>
</tr>
<tr>
<td>A little bit worried</td>
<td>43</td>
<td>(29)</td>
<td>75</td>
<td>(34)</td>
</tr>
<tr>
<td>Quite worried</td>
<td>4</td>
<td>(3)</td>
<td>4</td>
<td>(2)</td>
</tr>
<tr>
<td>Very worried</td>
<td>19</td>
<td>(13)</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

($\chi^2 = 27$, df = 3, $p < 0.001$)

(b) Anxiety due to screening and while awaiting the result of screening

<table>
<thead>
<tr>
<th></th>
<th>Screening Number</th>
<th>(%)</th>
<th>Results Number</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all worried</td>
<td>224</td>
<td>(51)</td>
<td>221</td>
<td>(50)</td>
</tr>
<tr>
<td>A little bit worried</td>
<td>192</td>
<td>(44)</td>
<td>192</td>
<td>(44)</td>
</tr>
<tr>
<td>Quite worried</td>
<td>15</td>
<td>(3)</td>
<td>21</td>
<td>(5)</td>
</tr>
<tr>
<td>Very worried</td>
<td>4</td>
<td>(1)</td>
<td>5</td>
<td>(1)</td>
</tr>
</tbody>
</table>

($\chi^2 = 1.1$, df = 3, $p < 0.77$)

(Not all women answered every question so the totals in the tables may differ)
TABLE 6.8

Anxiety related to screening by history of breast cancer in family or close friend

| Anxiety related to screening | Yes | No  |  |  |
|-----------------------------|-----|-----|  |  |
|                             | No. (%) | No. (%) |  |  |
| Not at all worried          | 17 (71) | 58 (45) |  |  |
| A little bit worried        | 6 (23)  | 64 (50) |  |  |
| Quite to very worried       | 1 (4)   | 7 (7)   |  |  |

\( \chi^2 = 5.5, \text{df} = 2, \ p = 0.06 \)

(Not all women answered every question so the totals in the tables may differ)
TABLE 6.9

Anxiety while awaiting results and time until results received

<table>
<thead>
<tr>
<th>Time until results received (weeks)</th>
<th>&lt; 1</th>
<th>1-2</th>
<th>2+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety while awaiting results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all worried</td>
<td>21</td>
<td>39</td>
<td>11</td>
</tr>
<tr>
<td>A little bit worried</td>
<td>12</td>
<td>36</td>
<td>13</td>
</tr>
<tr>
<td>Quite worried</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Very worried</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>82</td>
<td>26</td>
</tr>
</tbody>
</table>

\( \chi^2 = 6, \text{ df} = 6, p = 0.7 \)

(Not all women answered every question so the totals in the tables may differ)
### TABLE 6.10

**Anxiety on receiving invitation by anxiety while awaiting results**

<table>
<thead>
<tr>
<th>Anxiety while awaiting results</th>
<th>Not at all worried</th>
<th>A little bit worried</th>
<th>Quite worried</th>
<th>Very worried</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety on receiving screening invitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all worried</td>
<td>177</td>
<td>51</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>A little bit worried</td>
<td>43</td>
<td>136</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>Quite worried</td>
<td>3</td>
<td>4</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Very worried</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

\( \chi^2 = 282, \text{ df } = 9, p < 0.0001 \)

(Not all women answered every question so the totals in the tables may differ)
TABLE 6.11

Experience of the screening test

<table>
<thead>
<tr>
<th>Experience</th>
<th>Number</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A little uncomfortable</td>
<td>194</td>
<td>(45)</td>
</tr>
<tr>
<td>Very uncomfortable</td>
<td>102</td>
<td>(24)</td>
</tr>
<tr>
<td>Painful</td>
<td>82</td>
<td>(19)</td>
</tr>
<tr>
<td>Very painful</td>
<td>53</td>
<td>(12)</td>
</tr>
<tr>
<td>TOTAL</td>
<td>431</td>
<td>(100)</td>
</tr>
</tbody>
</table>

(Not all women answered every question so the totals in the tables may differ)
TABLE 6.12

Anxiety before screening and experience of screening

<table>
<thead>
<tr>
<th>Experience of screening</th>
<th>A little uncomfortable (No.)</th>
<th>Very uncomfortable (No.)</th>
<th>Painful (No.)</th>
<th>Very Painful (No.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all worried</td>
<td>42</td>
<td>13</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>A little bit worried</td>
<td>28</td>
<td>20</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>Quite worried</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Very worried</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>152</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( \chi^2 = 14, \text{ df} = 9, p = 0.12 \)

(Not all women answered every question so the totals in the tables may differ; 97% of the respondents answered this question)
TABLE 6.13

Anxiety while awaiting results and experience of screening

<table>
<thead>
<tr>
<th>Anxiety while awaiting results</th>
<th>A little uncomfortable</th>
<th>Very uncomfortable</th>
<th>Painful</th>
<th>Very Painful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>156</td>
<td>19</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

(\chi^2 = 17, df = 9, p = 0.04)

(Not all women answered every question so the totals in the tables may differ; 100% of respondents answered this question)
Experience of the screening unit

Information was collected about each screening centre. This included waiting times, time spent at the centre, and women's opinions about the waiting area, changing rooms, and the attitude of staff at the screening centre. At the fixed units eighty percent of women were screened within fifteen minutes of their appointment time. At the mobile unit ninety percent of women were screened within fifteen minutes of the time that they had been given. Waiting times are slightly shorter at the mobile screening unit because film processing is not carried out there; at the fixed units the women wait until their films have been processed in case any extra views need to be taken; if extra views are taken this can sometimes increase the waiting time for the next woman to be screened. When asked what they thought was an acceptable time to have to wait for a mammogram forty-eight percent thought less than fifteen minutes, forty-nine percent thought up to thirty minutes, and three percent thought up to an hour.

Over ninety-five percent of the women were happy with the size of the waiting areas and the temperature and privacy of the changing rooms. There was no difference between screening centres. Ninety-five percent of women were happy with the way the staff had treated them. Once again there was no significant difference between screening centres.

Costs to women

When the women were asked their reasons for participating in the pilot programme 154 women (thirty-five percent) gave the fact that it was a free service as one of their reasons. When they were asked how much they would be prepared to pay if the screening programme could not continue as a free service (Table 6.14) twenty percent of women said that they would not take part in the programme if there were a charge. Forty-two percent said they would pay up to twenty dollars, thirty-two percent said they would pay between twenty and fifty dollars, and six percent said they would be prepared to pay more than fifty dollars.

Although the two groups differ in household income there was no difference between urban and rural women for these questions. There was no relationship between household income and the amount women would be prepared to pay if the screening programme were not free. Most women travelled to the screening centre by car (eighty-one percent of urban women and ninety-two percent
TABLE 6.14

Amount women would be prepared to pay if the screening programme could not continue as a free service

<table>
<thead>
<tr>
<th>Amount prepared to pay</th>
<th>Number</th>
<th>(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Would not take part if charged</td>
<td>86</td>
<td>(20)</td>
</tr>
<tr>
<td>Under $20</td>
<td>180</td>
<td>(42)</td>
</tr>
<tr>
<td>$20 to $50</td>
<td>135</td>
<td>(32)</td>
</tr>
<tr>
<td>Over $50</td>
<td>25</td>
<td>(6)</td>
</tr>
<tr>
<td>Total</td>
<td>426</td>
<td>(100)</td>
</tr>
</tbody>
</table>
of rural women). Several rural women said that they would have used car pools if they had known when other women would be attending.

Future participation in screening
When asked whether they would return for another screen in two years if they received an invitation, almost all the women (ninety-four percent) said they would. This percentage was lower (seventy-five percent) in those who had found the mammogram 'very painful' \( (p < 0.01) \). The number of women in this category was very small, so no further analysis of this subgroup was undertaken.

Discussion
The response rate to this survey was very high; ninety-three percent after one reminder. This is probably partly because accurate addresses were available for all the women from the screening programme records. More importantly these women had already accepted an invitation to participate in the programme, and as part of giving their informed consent they were informed that they might be asked to take part in surveys about the pilot programme. The women were aware that the information they provided would be used to help with decisions about a national breast cancer screening programme. Some demographic characteristics of urban and rural women differed but their opinions about the screening programme were similar.

Although the pilot programme had received a lot of publicity, about half the women said that they found out about the programme only when they received an invitation to take part. About six weeks before invitations were sent out in a particular area the programme manager spoke to women's groups in the area and distributed posters and pamphlets locally. Most local newspapers gave the programme considerable coverage, and radio interviews and talkbacks were also used. Television campaigns were not used as they could not be directed specifically at small local areas.

Most women found the mammogram uncomfortable rather than painful. This is consistent with other investigations \( \text{(Stomper et al 1988, Jackson et al 1988, Wolosin 1989, Rutter et al 1992)} \). The suggestion that discomfort is related to anxiety before the test \( \text{(Rutter et al 1992)} \) is not supported by the results of this study, although the women who experienced the most discomfort were more anxious while awaiting the results of the mammogram. Those women who experienced
the most pain also had more bruising after the mammogram. The small number of women who do not intend to participate in the programme again were more likely to have found the mammogram very painful.

As in other reports (Dean et al 1986, Ellman et al 1989, Baines et al 1990) most women were either not worried or only slightly worried by taking part in the screening programme. The women who worried most while waiting for their results were the women who had also worried most about the screening appointment. Anxiety was not related to a history of breast cancer in family or close friends, nor to the time that the women had to wait to get their results.

A third of the women in this survey said that one of the reasons they took part in the screening programme was because it was free, and a third said that they would not have had a mammogram in the absence of the programme if they had had to pay for it. When asked how much they would be prepared to pay if the programme could not continue as a free service, twenty percent said they would not take part if there was a charge and only twenty-five women (six percent) said they would be prepared to pay more than fifty dollars. It is interesting to note that in the private sector in New Zealand it is not possible to have a mammogram for less than fifty dollars (Department of Health 1992).

It is very important that a screening service is acceptable to eligible women. If the service is unsatisfactory participation at re-screening will drop, and other women may decline to take part because of adverse comments from participants. In this survey overall satisfaction with the service was extremely high, with over ninety percent of respondents satisfied with the screening units and staff, and ninety-four percent planning to have another mammogram in two years. The fixed and mobile units were equally acceptable.

The survey results provided valuable management information for the programme, for example the survey showed that some women did not receive their results within the ten day maximum period initially set for the programme, and so an additional clerical person was appointed. The method of processing results was also been changed as a result of the survey. Because films have to be forwarded by courier for reporting, it is not possible to provide results as rapidly from the mobile screening unit. The ten day target was extended for the mobile unit.
Several rural women said that they would have used car pools if they had known when other women were being screened, but the need for confidentiality prevents the distribution of appointment lists. However there is always a public meeting before the mobile unit goes to a rural town and some women have organised car pools at these meetings, voluntarily giving their names to the programme staff.

Although the demographic characteristics of urban and rural women differed, their responses to the screening programme were similar. Satisfaction with the service was extremely high, with 94% of respondents planning to participate again in two years. Women who had found the mammogram very painful were less likely to intend to participate again. The radiographers are aware of this and are conscious of the need to maintain a balance between achieving adequate compression for a good image, and not causing pain to the women. The results of this survey, with a high level of satisfaction and a high level of intention to participate in future, suggest that the pilot programme provides a service which is acceptable to eligible women in Otago and Southland.

6.4 Acceptability of assessment

It is accepted that screening reduces breast cancer mortality among women aged fifty to sixty-four. However as well as this benefit there can be problems associated with screening. In order to detect and treat breast cancer, women with abnormal mammograms are referred for further investigations. Some of the women referred are found not to have breast cancer and have then needlessly suffered the anxiety associated with a positive screen result and the subsequent investigations. In order to minimise this problem it is important that the screening test has high specificity (and therefore few false positives) but also that the process of assessment for women with abnormal mammograms is acceptable.

A survey was carried out to measure satisfaction with the assessment process, to provide information for the assessment clinic staff, and to collect information about costs to women which would be used in the economic analyses of the pilot programme.
Methods
A self-administered questionnaire (Appendix Seven) was posted to all women who had attended the Invercargill and Dunedin assessment clinics during November 1991 to April 1992. Women who did not respond were sent a reminder letter and another copy of the questionnaire. The questionnaires were analysed using SPSS-X (Statistical Package for Social Sciences 1990). Comparisons have been made between this survey and the earlier survey of women after screening which was described in section 6.3.

Results
The questionnaire was posted to 140 women and 125 (eighty-nine percent) responded. The women in this survey in general had higher household incomes \( (\chi^2 = 9.24, \text{df} = 2, p < 0.01) \) than those in the screening survey (Tables 6.6 and 6.15), but did not differ from the women in the earlier screening survey in age distribution, ethnic group, education, and occupation.

Anxiety was measured using a four point Likert scale, as in the screening survey (section 6.3). In the assessment survey, the women were asked about their level of anxiety upon receiving the assessment clinic appointment and also about their level of anxiety on receiving an invitation to the screening clinic. Because this comparison was made within the same group of women, the sign test rather than chi-squared test was used to test for significance. The women surveyed had been more anxious on receiving the assessment appointment than they had been on receiving their screening appointment (Sign test \( (37+, 78=, 8-) \) \( \chi^2 = 17.6, \text{df} = 1, p < 0.005 \)). This increase in anxiety affected thirty percent of the women; it is important to note that for most women (seventy-eight of the 123 women who answered both questions) their level of anxiety did not change. Where the level of anxiety changed, it was significantly more likely to increase than to decrease (Table 6.16).

The level of anxiety experienced by the assessment survey respondents on attending the assessment clinic was also compared with that of the screening survey respondents (a different group of women) on attending the screening clinic. Comparing the responses in the assessment survey with those in the screening survey (Table 6.17) showed that attending the assessment clinic caused more anxiety \( (\chi^2 = 38.6, \text{df} = 2, p < 0.001) \) than attending the screening clinic. Most women (sixty-seven percent) received the results of their assessment within a week of the appointment, with
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-54</td>
<td>57</td>
<td>46</td>
</tr>
<tr>
<td>55-59</td>
<td>34</td>
<td>27</td>
</tr>
<tr>
<td>60-64</td>
<td>34</td>
<td>27</td>
</tr>
<tr>
<td>Ethnic group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maori</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Non-Maori</td>
<td>122</td>
<td>98</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary school</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Secondary school</td>
<td>80</td>
<td>68</td>
</tr>
<tr>
<td>University</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Trade qualification</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paid employment</td>
<td>45</td>
<td>37</td>
</tr>
<tr>
<td>Self-employed</td>
<td>10</td>
<td>8</td>
</tr>
<tr>
<td>No paid employment</td>
<td>33</td>
<td>27</td>
</tr>
<tr>
<td>Retired</td>
<td>33</td>
<td>27</td>
</tr>
<tr>
<td>Household income (before tax)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>less than $15,000</td>
<td>31</td>
<td>27</td>
</tr>
<tr>
<td>$15,000 - $30,000</td>
<td>49</td>
<td>43</td>
</tr>
<tr>
<td>more than $30,000</td>
<td>34</td>
<td>30</td>
</tr>
</tbody>
</table>
TABLE 6.16

Anxiety related to assessment appointment compared with screening invitation
in 125 women who attended the assessment clinic

<table>
<thead>
<tr>
<th>Anxiety</th>
<th>Screening Invitation</th>
<th>Assessment Appointment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Percent</td>
</tr>
<tr>
<td>Not at all worried</td>
<td>61</td>
<td>49</td>
</tr>
<tr>
<td>A little bit worried</td>
<td>51</td>
<td>41</td>
</tr>
<tr>
<td>Quite to very worried</td>
<td>13</td>
<td>10</td>
</tr>
</tbody>
</table>

Sign test (37+, 78=, 8-) $\chi^2 = 17.6$, df = 1, p < 0.005.
### TABLE 6.17

Anxiety related to being screened and being assessed

<table>
<thead>
<tr>
<th>Anxiety</th>
<th>Screening</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No.</td>
<td>Percent</td>
</tr>
<tr>
<td>Not at all worried</td>
<td>224</td>
<td>51</td>
</tr>
<tr>
<td>A little bit worried</td>
<td>192</td>
<td>44</td>
</tr>
<tr>
<td>Quite to very worried</td>
<td>19</td>
<td>4</td>
</tr>
</tbody>
</table>

\( \chi^2 = 38.6, \text{ df} = 2, \ p < 0.001 \)

(In this table the responses of two separate groups of women, in two separate surveys, have been compared; the responses given in the assessment survey have been compared with the responses given in the screening survey)
thirty-nine percent of the women being given the results while they were still at the assessment clinic. Twelve percent of women waited for more than two weeks to receive their results (Table 6.18). The level of anxiety while waiting for the results from the assessment clinic was not related to the length of time that the women had to wait for their results (Table 6.19, $\chi^2 = 13.3$, df = 9, $p = 0.15$).

While there were three screening centres (the two fixed screening centres in Dunedin and Invercargill, and a mobile unit for rural women), there were only two assessment clinics, one in Dunedin and the other in Invercargill.

Most women (eighty-six percent) travelled to the assessment clinic by car. Only two women had to travel for more than an hour to reach the assessment clinic. Most women (seventy-seven percent) attended the clinic on their own but eleven women came with an accompanying person to provide support.

Over half the women (fifty-two percent) spent less than thirty minutes at the clinic, and only thirteen percent were there for more than an hour. Over ninety-five percent of the women found the staff friendly and the clinic surroundings satisfactory.

Of the women who responded ninety-eight percent said they would take part in the screening programme again in two years if they received another invitation (Table 6.20). This is higher than the percentage (ninety-four percent) of women in the screening survey who said they would take part in the screening programme in future ($\chi^2 = 4.08$, $p = 0.04$).

Discussion
The results from this survey are similar to assessment clinic surveys carried out elsewhere, where over ninety-five percent of women who attended assessment clinics planned to attend for rescreening (Gram et al 1990, Smith et al 1991). Orton et al (1991) also found that there was no
<table>
<thead>
<tr>
<th>Time</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immediately</td>
<td>47</td>
<td>39</td>
</tr>
<tr>
<td>Within 1 week</td>
<td>33</td>
<td>28</td>
</tr>
<tr>
<td>1 - 2 weeks</td>
<td>26</td>
<td>22</td>
</tr>
<tr>
<td>More than 2 weeks</td>
<td>14</td>
<td>12</td>
</tr>
</tbody>
</table>
TABLE 6.19

Anxiety while awaiting results of assessment by time until results received

<table>
<thead>
<tr>
<th>Time until results received</th>
<th>Immediately (at clinic)</th>
<th>&lt; 1 week</th>
<th>1 - 2 weeks</th>
<th>2 weeks +</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not at all worried</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>A little bit worried</td>
<td>20</td>
<td>17</td>
<td>14</td>
<td>7</td>
</tr>
<tr>
<td>Quite worried</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Very worried</td>
<td>11</td>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

($\chi^2 = 13.3$, df= 9, p = 0.15)
<table>
<thead>
<tr>
<th>Intention to participate in future</th>
<th>Number</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>116</td>
<td>98</td>
</tr>
<tr>
<td>No</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>
difference between women who had false positive results at a previous screen compared with women who had negative screen results, in their subsequent participation in screening.

It has been shown that women who have had routine breast cancer screening examinations do not experience increased psychiatric morbidity (Dean et al 1986, Ellman et al 1989). But, as in the Otago and Southland assessment survey, women requiring further assessment have significantly higher anxiety scores than those attending for routine screening (Lerman et al 1991, Ellman et al 1989 Gram et al 1990). It is important to minimise the anxiety related to assessment; ensuring relaxed surroundings, providing adequate information, and minimising waiting times for results are important (Marteau 1990, Smith et al 1991). Two studies of the effects of attending an assessment clinic have found that there are few long-term consequences of assessment, even for women with false positive results (Ellman et al 1989, Gram et al 1990) although one study found that women who have experienced false positive tests remain more anxious about breast cancer afterwards (Bull and Campbell 1991) This study concluded that “the psychological effects of breast cancer screening are small in the general population, but of note in women who undergo open biopsy of a benign lesion”.

The Otago and Southland assessment survey results are consistent with those from other similar surveys. The results also show that the Dunedin and Invercargill assessment clinics are providing an acceptable service for women who take part in the pilot programme, at least in the shorter term. Determination of the longer-term consequences of assessment in the pilot programme will require longer follow-up.

6.5 Survey of women who did not attend after an invitation to screening

A survey was carried out to identify factors that influenced attendance or non-attendance at the screening programme. The aim of this study was to identify any characteristics of eligible women, and of the screening programme, that had deterred women from participating. If any of these characteristics were modifiable it could be possible to increase participation, for instance by changing features of the programme, or by specifically targeting certain groups of women. The author planned this survey and designed the questionnaires as part of the evaluation of the
acceptability of the pilot programme, but the survey was still ongoing at the time the author left New Zealand. The data management and analysis of the results were carried out by Bronwen McNoe and Mark Elwood (McNoe et al, in press). A description of the survey, and its results follows.

Methods
A random sample of ten in every hundred women who attended the screening programme from January to July 1993 after being sent an invitation, and who had a normal screening result was taken. Another random sample of ten in every thirty women who had been invited but did not attend for screening from January 1993 to March 1994 was taken. The sampling methods and time frames differed for the two groups of women because there were fewer non-attenders than attenders, and non-attenders were considerably more difficult to contact.

There were 272 attenders identified, and of these one had changed address, four had no known telephone numbers (although telephone numbers were requested at the time of screening), and forty-four women could not be contacted after three attempts. Of the 223 women contacted, twenty-nine had self-referred and so were ineligible. Of the 194 eligible attenders, one had hearing problems and two declined to participate in the study. The remaining 191 attenders were interviewed; this is ninety-eight percent of those contacted and seventy-nine percent of the estimated total sampling frame (272 minus twenty-nine). Copies of the questionnaires are included as Appendix Eight.

Four hundred and ninety-seven non-attenders were identified, but five had changed address, 156 had no telephone numbers, and 115 could not be contacted after three attempts. There were seventeen women among the non-attenders who said that they had not received an invitation although the screening programme records showed that an invitation had been sent. There were 204 women who had definitely received an invitation but declined screening, of these three could not be interviewed because of language problems or severe illness, and twenty-seven declined to be interviewed. This left 174 (eighty-six percent) of those who could be contacted and had definitely declined an invitation, but this was only thirty-five percent of the 497 women who had been sent an invitation but had not attended during the study period.
Standardised telephone interviews were carried out by an experienced interviewer from the evaluation team. The interview included both closed and open questions. Where possible the interviews were identical for attenders and non-attenders, but they differed in terms of questions about reasons for attendance or non-attendance. The interviewer could not be blinded to the attendance status of the interviewee as different introductory statements were required (Appendix Eight). Confidentiality was guaranteed.

Results
There were 191 attenders and 174 non-attenders interviewed in the survey. There were no differences in age, education, income, socioeconomic status, or current employment between the two groups. Several closed questions and an open question, were asked to determine why the women had attended (or not attended) for screening (Table 6.21).

About two thirds of the attenders selected reassurance (sixty-nine percent) and early detection of breast cancer (sixty-five percent) as reasons for attendance. Forty-two percent selected the free service as one of the reasons that they had attended, thirty-four percent because it was recommended by their general practitioner, and thirty-one percent because it was a pilot programme. Seventeen percent said they had been influenced by positive reports from other women, ten percent because of a family history of breast cancer, and four percent because of a current breast symptom. Thirty-four women (eighteen percent) answered the open question and most (seventeen) gave a perception that they were at increased risk (because of a personal or family history of breast cancer, or because they were on hormone replacement therapy) as reasons for attending.

Most of the non-attenders (seventy-four percent) answered the open question on their reasons for non-attendance. Twenty-three percent of them said they had not attended because they could not easily attend (because of health problems or other activities). Twenty-one percent were not convinced that mammography was valuable. Responses to the closed questions are shown in Table 6.21. Fifteen percent of the non-attenders said that negative reports from other women had influenced them. Seven percent thought they would be ineligible (having had a recent mammogram or under treatment for breast cancer). Twenty percent cited practical difficulties that led them to
### TABLE 6.21

**Reasons for attendance or non-attendance at screening**

**Attenders (n = 191) Reasons for attendance**

(multiple responses could be given)  

<table>
<thead>
<tr>
<th>Reason</th>
<th>n</th>
<th>% of all attenders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reassurance</td>
<td>131</td>
<td>68.6</td>
</tr>
<tr>
<td>To detect breast cancer early</td>
<td>125</td>
<td>65.4</td>
</tr>
<tr>
<td>Because it is a free service</td>
<td>81</td>
<td>42.4</td>
</tr>
<tr>
<td>Recommended by my family doctor</td>
<td>65</td>
<td>34.0</td>
</tr>
<tr>
<td>Because it is a pilot programme</td>
<td>60</td>
<td>31.4</td>
</tr>
<tr>
<td>Positive reports from other women</td>
<td>32</td>
<td>16.8</td>
</tr>
<tr>
<td>Family history of breast cancer</td>
<td>19</td>
<td>9.9</td>
</tr>
<tr>
<td>Breast symptom</td>
<td>8</td>
<td>4.2</td>
</tr>
</tbody>
</table>

**Non-attenders (n = 174) Reasons for non-attendance**

(within major categories, multiple responses could be given)  

<table>
<thead>
<tr>
<th>Reason</th>
<th>n</th>
<th>% of all non-attenders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likely to be ineligible</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recent mammogram</td>
<td>8</td>
<td>6.9</td>
</tr>
<tr>
<td>Being treated for breast problem</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Practical difficulties</td>
<td>35</td>
<td>20.1</td>
</tr>
<tr>
<td>Inconvenient appointment time</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>No transport</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Too far from screening centre</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Negative opinion of screening</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fear of procedure</td>
<td>22</td>
<td>19.5</td>
</tr>
<tr>
<td>Negative reports from other women</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Fear of possible outcome</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>No reason stated</td>
<td>93</td>
<td>53.4</td>
</tr>
</tbody>
</table>

decide against attending, and another twenty percent cited fear of the procedure. Three percent of the non-attenders objected to the invitation procedure, either because an appointment had been made in advance, or because they were not happy that the programme had access to their names and addresses. Both groups of women were asked about practical aspects of attending for screening (Table 6.22), including the time it would take to travel there, their mode of transport, what they would otherwise have been doing at the time specified for the screen, and whether it would be easy for them to fit such an appointment into their usual routine. There was no difference in mode of transport, but the attenders assessed the time for them to travel to the screening centre as significantly less than non-attenders (p = 0.02). The groups also differed in what they would otherwise have been doing at the time specified for the screen (p < 0.01), with attenders less likely to be in paid employment or caring for children. Nine percent of the non-attenders were ill. Ninety-five percent of attenders found it very easy or quite easy to fit a screening appointment into their usual routine, compared with only forty-nine percent of non-attenders (p < 0.01).

The women were also asked, using a four-point Likert scale, how anxious the invitation to screening had made them feel (Table 6.23), how they perceived their risk of breast cancer, and whether they had any family members or close friends who had breast cancer. There was no difference between the groups in their level of anxiety on receiving the invitation. Attenders were more likely to regard themselves at higher risk of breast cancer compared with other women their age (p = 0.02).

Both groups were asked whether they would attend for screening if they received an invitation in two years (Table 6.24). Ninety percent of the attenders said they would attend, and forty-three percent of the non-attenders said they would attend the next screening round ($\chi^2 = 90$, p < 0.0001). Intention to participate varied among the non-attenders according to their reason for non-attendance. In the non-attenders who cited practical difficulties as their main reason for not attending, intended participation at the next screening round was eighty percent, which is not significantly different from the attenders. The rest of the non-attenders had intended participation rates of forty-two percent or less (Table 6.24).
TABLE 6.22

Practicalities of screening for women invited to screening

<table>
<thead>
<tr>
<th></th>
<th>% attenders n = 191</th>
<th>% non-attenders n = 174</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mode of transport</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walked</td>
<td>7.4</td>
<td>5.3</td>
</tr>
<tr>
<td>Car/bicycle</td>
<td>84.2</td>
<td>81.1</td>
</tr>
<tr>
<td>Taxi</td>
<td>8.4</td>
<td>13.6</td>
</tr>
<tr>
<td>missing</td>
<td>(1)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>( \chi^2 = 2.9 ) (2 df) ( p = 0.23 )</td>
<td></td>
</tr>
<tr>
<td>Time to travel to screening centre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 15 minutes</td>
<td>61.8</td>
<td>52.6</td>
</tr>
<tr>
<td>15 to 30 minutes</td>
<td>31.9</td>
<td>32.2</td>
</tr>
<tr>
<td>30 minutes or more</td>
<td>6.3</td>
<td>15.2</td>
</tr>
<tr>
<td>missing</td>
<td>(0)</td>
<td>(3)</td>
</tr>
<tr>
<td></td>
<td>( \chi^2 = 8.2 ) (2 df) ( p = 0.02 )</td>
<td></td>
</tr>
<tr>
<td>Usual activity at time of screening appointment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working (paid &amp; voluntary)</td>
<td>21.9</td>
<td>24.4</td>
</tr>
<tr>
<td>Leisure activity</td>
<td>32.6</td>
<td>34.1</td>
</tr>
<tr>
<td>Caring for child or relative</td>
<td>4.3</td>
<td>7.3</td>
</tr>
<tr>
<td>Housework</td>
<td>41.2</td>
<td>25.6</td>
</tr>
<tr>
<td>Ill</td>
<td>0</td>
<td>8.5</td>
</tr>
<tr>
<td>missing</td>
<td>(4)</td>
<td>(10)</td>
</tr>
<tr>
<td></td>
<td>( \chi^2 = 23.9 ) (4 df) ( p &lt; 0.01 )</td>
<td></td>
</tr>
<tr>
<td>How easy was it to fit in a screening visit?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very easy</td>
<td>44.5</td>
<td>11.8</td>
</tr>
<tr>
<td>Quite easy</td>
<td>50.3</td>
<td>36.7</td>
</tr>
<tr>
<td>A little difficult</td>
<td>4.7</td>
<td>12.4</td>
</tr>
<tr>
<td>Quite/very difficult</td>
<td>0.5</td>
<td>39.1</td>
</tr>
<tr>
<td>missing</td>
<td>(0)</td>
<td>(5)</td>
</tr>
<tr>
<td></td>
<td>( \chi^2 = 116.6 ) (3 df) ( p &lt; 0.01 )</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 6.23

Anxiety on being invited to screening, and perceived risk of breast cancer

<table>
<thead>
<tr>
<th></th>
<th>% attenders n = 191</th>
<th>% non-attenders n = 174</th>
</tr>
</thead>
<tbody>
<tr>
<td>How worried did the invitation make you?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not worried</td>
<td>75.0</td>
<td>78.9</td>
</tr>
<tr>
<td>A little worried</td>
<td>19.7</td>
<td>14.5</td>
</tr>
<tr>
<td>Quite/very worried</td>
<td>5.3</td>
<td>6.6</td>
</tr>
<tr>
<td>missing</td>
<td>(3)</td>
<td>(8)</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 2.0 \ (2 \text{ df}) \ p = 0.4 \]

<table>
<thead>
<tr>
<th>Women’s self-perceived breast cancer risk</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Less at risk</td>
<td>21.6</td>
<td>19.5</td>
</tr>
<tr>
<td>Same risk</td>
<td>33.2</td>
<td>28.4</td>
</tr>
<tr>
<td>More at risk</td>
<td>11.6</td>
<td>4.7</td>
</tr>
<tr>
<td>Don’t know</td>
<td>33.7</td>
<td>47.3</td>
</tr>
<tr>
<td>missing</td>
<td>(1)</td>
<td>(5)</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 10.0 \ (3 \text{ df}) \ p = 0.02 \]

<table>
<thead>
<tr>
<th>Family member or close friend with breast cancer</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>78.3</td>
<td>82.2</td>
</tr>
<tr>
<td>Yes</td>
<td>21.7</td>
<td>17.8</td>
</tr>
<tr>
<td>missing</td>
<td>(2)</td>
<td>(5)</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 0.6 \ (1 \text{ df}) \ p = 0.42 \]

### TABLE 6.24

Intention to participate in future screening

<table>
<thead>
<tr>
<th>Total number</th>
<th>Yes</th>
<th>Percent</th>
<th>No/uncertain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Attenders (all)</strong></td>
<td>191</td>
<td>172</td>
<td>90.1</td>
</tr>
<tr>
<td><strong>Non-attenders</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>practical difficulties</td>
<td>35</td>
<td>28</td>
<td>80.0</td>
</tr>
<tr>
<td>thought ineligible</td>
<td>12</td>
<td>5</td>
<td>41.7</td>
</tr>
<tr>
<td>no reason given</td>
<td>93</td>
<td>33</td>
<td>35.5</td>
</tr>
<tr>
<td>negative opinion</td>
<td>34</td>
<td>9</td>
<td>26.5</td>
</tr>
<tr>
<td><strong>all non-attenders</strong></td>
<td>174</td>
<td>75</td>
<td>43.1</td>
</tr>
</tbody>
</table>

Intention to participate in attenders compared with non-attenders $\chi^2 = 90$, $p < 0.001$.

Discussion

This study showed that the major difference between attenders and non-attenders was that women who did not attend rated breast cancer screening as less important in comparison to the time and effort needed to attend. Although non-attenders assessed the time for them to travel to the screening centre as significantly more than attenders, this perception was not supported by the actual distances involved. The actual distance from each woman's residence to the screening centre was measured on a map, and there was no significant difference between the attenders (mean distance 9.2 km) and non-attenders (mean distance 10.7 km, \( p = 0.43 \)). A study carried out in London found that the actual distance from a woman's home to the screening centre was not related to attendance (Sutton et al. 1994). These findings are in accord with the Health Belief Model, which proposes that decisions are related to beliefs about the seriousness of the condition, personal susceptibility, and effectiveness of the measures suggested, compared to the costs described as inconvenience, discomfort, or risk (Calnan 1984).

The results of this study suggest that participation could be increased by reducing practical barriers to screening (for instance greater availability of appointments outside normal working hours) and by increasing women's knowledge about breast screening. But it is important to respect the right of women not to take part in screening if they do not wish to. Women who have been provided with appropriate information about breast cancer screening should not feel pressured to participate if they do not wish to. An important result is that forty-four percent of the women who did not attend in the first screening round said that they would attend if they received an invitation to the second round. This means that it is reasonable to continue to invite women who have declined an invitation to a previous screening round. However this survey was restricted to those women who could be contacted and who had definitely declined an invitation, so this result may not apply to all women who are sent an invitation but do not attend.

6.6 Summary

This chapter has examined the acceptability of the Otago and Southland pilot programme to the women who are eligible for screening. The pilot programme was designed to incorporate the features identified in the pre-screening survey as important to the acceptability of breast cancer screening.
screening for women in Otago and Southland. The pilot programme included a mobile unit so that rural women would not have to travel too far for screening, some evening screening sessions, female radiographers and clerical staff, and publicity about the programme.

Intended participation had been high in the pre-screening survey, and many women referred themselves for screening rather than waiting for an invitation. But this level of self-referral may not happen in other places, and means that one cannot predict what will happen in other parts of New Zealand. It seems likely however, that many women will self-refer or be referred by their general practitioners before receiving an invitation. In the Waikato pilot programme sixteen percent of the women screened in 1994 were self-referred or had been referred by their general practitioners before receiving an invitation, and in 1995 fifteen percent of the women screened were self-referred or referred by their general practitioners (McKinnon, 1996 personal communication). It is very helpful to know however, that seventy-one percent participation can be achieved with an appropriate invitation signed by a general practitioner (Chapter Five, section 5.1.3), although this may not apply to all general practices. It will be important to monitor participation at re-screening in the second round of the pilot programme, as continuing high participation will be necessary to maintain over seventy percent coverage of the women in the target population. This is discussed further in Chapter Seven.

There was a very high level of satisfaction with all aspects of the screening service, with no difference between centres. Ninety-four percent of the women would return for another screen in two years. The survey showed that some women did not receive their results within the maximum period initially set for the programme. An additional clerical person was appointed as a result. The method of processing results has also been changed as a result of the survey. There was a high level of satisfaction with the assessment clinics, although understandably assessment caused women more anxiety than screening. Women who could be contacted and who had definitely declined an invitation to take part in screening did not differ in demographic characteristics from the women who took part, but they assessed the value of being screened as less than the practical problems of attending. The acceptability of screening in the Otago and Southland pilot programme and the implications for the acceptability of screening in a national breast cancer screening programme are discussed in Chapter Seven.
CHAPTER SEVEN

Summary and Discussion

Introduction

This chapter summarises the results of the evaluation of the Otago and Southland pilot programme. The strengths and weaknesses of the evaluation method are examined in section 7.1, and the principal findings of the evaluation are discussed in sections 7.2 to 7.4, under the three areas specified for the evaluation (Skegg et al. 1988); effectiveness (section 7.2), acceptability (section 7.3), and economic efficiency (section 7.4). The implications of these results are discussed in section 7.5, some recommendations are made in section 7.6, and conclusions are drawn in section 7.7.

7.1 Method strengths and limitations

The evaluation method used here required information from randomised controlled trials for the development of targets. The method was appropriate for the evaluation of the pilot programme because it had already been shown in randomised controlled trials that mammographic screening can reduce breast cancer mortality. Randomised controlled trials provide the best evidence of the efficacy of a health intervention, so a particular strength of this evaluation method was that the pilot programme could be compared against programmes which had proven efficacy. The evaluation targets were derived from reviewing the relevant literature (Chapter One), and methods were developed for measuring these targets in the pilot programme (Chapters Three and Four). The targets provided a means to ensure that the pilot programme was similar to programmes which had been shown to reduce breast cancer mortality. It was assumed that if the pilot programme was similar in its early performance to successful programmes, it too would reduce breast cancer mortality in the longer term. This is an accepted approach to the evaluation of screening.
programmes, where interim measures have been used to predict the likely outcome of screening (Hakama et al 1985, Day et al 1989, Paci et al 1990, Verbeek et al, 1991, Morrison 1991). In this evaluation method suitable interim measures, or targets, were selected after reviewing the results from randomised controlled trials, in order to develop a standard against which the pilot programme could be evaluated. Although this approach has the advantages described above, the use of randomised controlled trials is also a limitation of this method. The method is useful in evaluating interventions or programmes where there is evidence of efficacy from randomised controlled trials. In many cases such evidence is not available, so this evaluation method would not be applicable.

Another advantage of the evaluation method used here is its flexibility; targets can be chosen that relate to the specific objectives of the evaluation, and this information can easily be supplemented with other information that is not part of the goal-attainment evaluation. The goal-attainment method was used to evaluate the effectiveness and one aspect of the acceptability of the pilot programme (participation), and supplemented with information from acceptability surveys and an economic analysis. The objectives of the evaluation of the Otago and Southland pilot programme were to ensure that the programme was well-organised, therefore keeping the benefit : risk ratio high, and to measure the acceptability, effectiveness, and economic efficiency of population-based mammographic screening in New Zealand. The evaluation required the collection of diverse data, from several sources. This is a strength of the evaluation method, in that more than one aspect of the pilot programme could be examined. However the collection of diverse data from several sources is also a potential weakness of this evaluation method, since it is time-consuming and complex to organise such data collection. The collection of data for the evaluation is summarised in section 7.1.1 below.

7.1.1 Data collection

In order to assess the acceptability, effectiveness, and efficiency of the pilot programme, a considerable amount of information had to be collected for the evaluation. Because of the diversity of data that had to be collected, several sources were used (Chapter Four, section 4.4). These sources included the pilot programme computer system and records, the pathology departments at Dunedin and Invercargill Hospitals, private pathology laboratories in Dunedin and Invercargill,
general practitioners in Otago and Southland, women in the eligible population, the Resource Utilisation System at Dunedin Hospital, and the National Health Information service.

Collecting the information for the evaluation of the Otago and Southland pilot programme required help from all the people involved in the programme, and in many cases, generated extra work for them. The evaluation could not have been carried out without this help, and this should be kept in mind when plans for a national screening programme are made (section 7.6). Some information had to be specially generated for the evaluation of the pilot programme. For instance, information about double reading was collected in a separate database so that analysis of inter-rater reliability could be carried out for the evaluation. Before the introduction of the screening programme grading of breast tumours had routinely been done at neither Dunedin Hospital nor Invercargill Hospital. The introduction of histological grading increased the workload of the pathologists who read the slides for the pilot programme. Staging was also carried out specifically for the programme, although it would normally have been done anyway, in the department of surgery, to help determine the treatment offered to women undergoing surgery for breast cancer. Staging involved regular meetings of the specialist breast surgeon, screening programme manager, and evaluation coordinator.

Collecting much of the information for the evaluation relied on the goodwill of people who took part in surveys related to the pilot programme. Many surveys were carried out during the evaluation, including surveys of women who were eligible for screening, women who had attended screening, women who had attended the assessment clinic, women who had declined an offer of screening, and general practitioners in Otago and Southland. The response rates were high (over eighty percent) in all these surveys. The survey of general practitioners included all general practitioners in Otago and Southland. Because random samples were used in the surveys of women, and response rates were high, the results can be generalised to the wider population of eligible women in Otago and Southland.

There were two areas in which there was particular difficulty in collecting data for the evaluation; the pilot programme computer system, and obtaining information about women diagnosed with breast cancer outside the pilot programme:
The pilot programme computer system

The first problem with collecting data from the pilot programme computer system was the requirement by Dunedin Hospital that the screening programme computer system should be part of the Dunedin Hospital information system, rather than a stand-alone system. This meant that the pilot programme was restricted to a computer system modified from the pre-existing Radiology system. Because of the limitations of this system some information had to be collected manually, for instance invitation and participation data, (Chapter Four, section 4.4, Chapter Six, section 6.2). Also, the screening programme computer system was not user-friendly, and was not sophisticated enough to prevent incorrect data entry; for instance incorrect dates could be entered with impunity. This meant that a considerable effort had to be made to ensure that the data for the evaluation of the screening programme was of high quality. Regular validity checks were carried out on the data that was obtained from the screening programme computer system, and corrections were made when necessary (Chapter Four, section 4.4). Validity checks are very important in clinical databases as published error rates in clinical databases range from two percent to thirty-eight percent (Reynolds-Haertle and McBride 1992, Hohnloser et al 1994, Horbar and Leahy 1995).

Even more importantly, the Dunedin Hospital information system was modified in mid-1994, during the second screening round, and this affected the screening programme computer system with the result that the screening programme data was no longer available in the format required for the evaluation download. From mid-1994 the evaluation download was unusable. This was completely unsatisfactory, and after many months spent trying to solve the problem, a data manager was appointed by the pilot programme to work with the evaluation team to ensure that the required data could be downloaded properly.

Identifying women diagnosed outside the screening programme

A further problem was that of identifying women who were diagnosed with breast cancer outside the screening programme. At the time the evaluation began, there was no legislation to make cancer registration compulsory. There was concern among pathologists that releasing information about women who were diagnosed outside the screening programme would breach confidentiality. It was essential to identify women with interval cancers so that the sensitivity of screening could be estimated, as this was one of the key targets for the evaluation of the pilot programme. In Dunedin, by using National Master Patient Index (NMPI) numbers, hospital histology records could be
matched with the screening records. Names were not required, thus maintaining confidentiality.
NMPI numbers were not used at Invercargill Hospital nor in the private pathology laboratories so
another method was devised to identify women who were diagnosed with interval cancers outside
Dunedin hospital. It was possible to identify all woman over fifty who were diagnosed with breast
cancer either through the Southland hospital pathology department or one of the private pathology
laboratories, using the laboratory computer records. For Southland hospital copies of histology
reports for women over fifty were routinely sent to the evaluation coordinator, and for the private
pathology laboratories (where smaller numbers were involved) the evaluation coordinator
requested copies of the forms at regular intervals. In each case the histology form had the woman's
name blanked out. Date of birth was used to look for possible matches with the screening register,
and where there was a match and the name was required the woman's surgeon (who could identify
the woman from her laboratory number on the histology form) asked her permission to release her
name for the evaluation. This method was cumbersome but was the only way to identify women
diagnosed with breast cancer outside the programme while still maintaining confidentiality between
the pathologist and the patient. An audit of the accuracy of this method is planned; the names of
women diagnosed with interval cancers in the evaluation will be compared with those registered at
the Cancer Registry (this check could not be done straight away because of the time-lag between
cancer diagnosis and subsequent registration at the Cancer Registry).

With the recent passage of the Cancer Registry Act 1993, which made it compulsory from July
1994 for all pathologists to notify diagnoses of cancer (Chapter Two, section 2.1) identification of
women diagnosed with breast cancer outside the screening programme should become
considerably easier. It is to be hoped that the pathologists could send copies of the histology forms
(with names), directly to the evaluation team, as well as to the Cancer Registry, so that the
identification of women with interval cancers can be carried out simply and rapidly.

The advantages and limitations of the method used in the evaluation of the Otago and Southland
pilot programme have been outlined, with particular reference to the collection of data. The
advantages of the method are its use of evidence from randomised controlled trials as a standard, its
flexibility, and its ability to evaluate several aspects of a programme concurrently. The method
would not be appropriate for evaluating interventions where evidence of efficacy from randomised
controlled trials is lacking. However, it may be possible to adapt this method to situations where
there is adequate information from observational studies to allow the development of suitable targets. The timely collection of accurate and appropriate data is essential to this method, as it is to any evaluation method. The implications and relevance of this evaluation are discussed in sections 7.5 to 7.7, but first the results of the evaluation of the Otago and Southland pilot programme are summarised in sections 7.2 to 7.4, under effectiveness (section 7.2), acceptability (section 7.3), and economic efficiency (section 7.4).

7.2 The effectiveness of the pilot programme

This section and sections 7.3 and 7.4 give the principal findings of the evaluation of the Otago and Southland pilot programme. The performance of the pilot programme, measured against the targets that were set as part of the evaluation, is shown in Table 7.1.

7.2.1 Identification of eligible women

The target for the programme was to identify ninety percent of eligible women. Actually ninety-two percent of women aged fifty to sixty-four and living in Otago and Southland were identified using the electoral roll alone (Chapter Five, section 5.1.1). General practice age-sex registers were also used. Each woman identified from a general practice list was checked off on the electoral roll, and the remaining women who had not been identified through the general practice registers were sent invitations later. Because general practice age-sex registers were used as well as the electoral roll, identification of eligible women was well over the target of ninety percent (Chapter Five, sections 5.1.1 and 5.1.2). It was very important to identify a high proportion of eligible women so that as many women as possible could be invited for screening.

7.2.2 Invitation method

Two studies were carried out to assess different methods of inviting women to take part in the screening programme (Chapter Five, section 5.1.3). In the first study the effect of general practitioners' letters endorsing screening was assessed in 482 women who were on the register of a large health centre. Each woman was randomly allocated to receive a letter from her general practitioner with her invitation or to receive only an invitation. Including those who had
TABLE 7.1

Characteristics of the first screening round compared with the targets for the pilot programme

<table>
<thead>
<tr>
<th>Characteristic evaluated:</th>
<th>Target:</th>
<th>Actual:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of eligible women</td>
<td>90%</td>
<td>92%</td>
</tr>
<tr>
<td>Coverage *</td>
<td>70%</td>
<td>74%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>85%</td>
<td>92%</td>
</tr>
<tr>
<td>Specificity</td>
<td>95%</td>
<td>89%</td>
</tr>
<tr>
<td>Percent of screened women referred</td>
<td>5-10%</td>
<td>11.6%</td>
</tr>
<tr>
<td>Biopsy rate (% of screened women)</td>
<td>&lt;2%</td>
<td>1.96%</td>
</tr>
<tr>
<td>Benign:malignant ratio of biopsies</td>
<td>&lt;3:1</td>
<td>0.9:1</td>
</tr>
<tr>
<td>screen prevalence expected incidence</td>
<td>3+</td>
<td>4.5</td>
</tr>
</tbody>
</table>

* Coverage was defined as the proportion of eligible women in Otago and Southland who were screened in the first screening round.
required reminders, participation among eligible women who had received letters was seventy-one percent compared with sixty-two percent among women who did not receive letters from their general practitioners ($p = 0.059$). Participation resulting from an invitation alone, with no reminder, was only forty-three percent. Forty-three percent is an unsatisfactory participation rate for a population-based screening programme.

The second study was carried out to see whether telephone reminders were as effective as postal reminders for women who did not reply to an invitation to take part in the screening programme. 495 women who had not replied to an invitation and who had telephone numbers were randomly allocated to receive either a telephone or a postal reminder. There was no difference in participation between the group receiving telephone reminders (forty-eight percent) and the group receiving postal reminders (forty-nine percent). As might be expected, participation was lower among these women, who had not replied to an earlier invitation.

In the pilot programme, as in breast cancer screening programmes overseas, general practitioner endorsement of invitations increased participation in breast cancer screening. Postal reminders were as effective as telephone reminders in encouraging women who did not respond to an initial invitation, to participate in screening.

7.2.3 The screening test

In the first eighteen months of the Otago and Southland pilot programme 1.2 percent of all the women screened were recalled because of technical problems (Chapter Five, section 5.2.1). The 95% confidence interval was from 0.97 to 1.48 percent. This result compared reasonably well with the range published for technical recall rates in other breast cancer screening programmes which was from 0.4 to 3.8 percent (Sickles et al 1987, Rickard et al 1991).

Film reading was carried out to a high standard in the pilot programme (Chapter Five, section 5.2.2). Inter-rater reliability was as high as that reported from any other screening programmes (Boyd et al 1982, Boyd et al 1986, Baines et al 1986, Ciccone et al 1992). All the radiologists involved in reading had attended special courses in screening mammography before the programme began. Originally it was planned to use “gold standard films” from an established screening centre to ensure that the standard of film reading was acceptable before the programme started.
Arrangements were made for the gold standard films to be sent to New Zealand from Britain, but unfortunately their production was delayed. It was accepted instead that the courses attended by the radiologists had provided suitable training for them to achieve an adequate standard of film reading. This proved to be so for the Otago and Southland programme, in terms of sensitivity, but not specificity. Ideally new programmes should not begin until it has been demonstrated that the level of film reading is adequate. Possibly a set of gold-standard films, to be used for reviewing films and training new radiologists, could be developed by the radiologists in the two pilot programmes, for use both in the pilot programmes and in other parts of New Zealand. This may help to reduce the recall rate and improve specificity in the pilot programme, and also help to ensure appropriate sensitivity and specificity of mammography in other screening centres.

The extent of agreement between the two radiologists who did most of the film-reading in the pilot programme was higher than in similar studies which have been published (Appendix One). An interesting result is that the double reading process did not necessarily increase the number of women referred for assessment. Had the initial radiological decision from a single radiologist been used, without double reading, the referral rate would have ranged from 9.4 to 12.3 percent depending on the radiologist. The breast cancer detection rate in the pilot programme would have been lower without double reading of the mammography films. If there had been only one reader three to five percent of women with breast cancer would have been missed, depending on which radiologist did the reading. These results are consistent with findings from other screening programmes. In Finland (Anttinen et al 1993) single reading would have led to either six percent or ten percent of the women with breast cancer being missed. In a study from Scotland (Anderson et al 1994) 10.9 percent of breast cancers would not have been detected without double reading.

Double-view mammography was used in the pilot programme. A randomised controlled trial comparing single-view with double-view mammography in England (Wald et al 1995) found that double-view mammography reduced the proportion of screened women recalled from 8.16% to 6.97%, a fifteen percent reduction (95% CI 6% to 23%). Also, two-view mammography detected twenty-four percent more women with breast cancer (95% CI 16% to 31%) than one-view mammography (Wald et al 1995). The study by Wald et al also found that although the cost of two-view screening was higher (£26.46 compared with £22.00 per examination), the average cost
per cancer detected was similar (£5,330 compared with £5,310) and the marginal cost per extra cancer detected with two views was similar to the average cost (£5,400).

7.2.4 Recall for assessment

The target recall rate in the pilot programme was between five and ten percent. In the first eighteen months of screening the Otago and Southland programme recalled 11.6 percent of women for further assessment. With symptomatic women excluded the recall rate was 10.4 percent (Chapter Five, section 5.3). This was higher than expected and it did not decline during the first year of screening. The referral rate was higher than that reported in many breast cancer screening trials for the first round of screening. The referral rate was 1.5 percent in the first screening round of the Stockholm trial (Frisell et al 1986, Frisell et al 1991) and five percent in the first round of the Swedish Two-Counties trial (Tabar et al 1992), but these programmes had an intermediate recall phase for more detailed mammography before women were referred for assessment. In the Health Insurance Plan (HIP) trial the referral rate was five percent but this was averaged over all screening rounds (Shapiro 1977, Shapiro et al 1982) and is likely to have been higher in the first round (Kopans and Swann 1988). In the Canada II trial in women aged fifty to fifty-nine (Miller et al 1992b) the referral rate in the first screening round was eleven percent. In the first screening round of the Sydney Breast X-Ray programme for women aged forty-five and over (Rickard et al 1991), which also had a policy of referring symptomatic women irrespective of their screen results, the referral rate was 16.5 percent. It is possible to reduce referral rates fairly quickly; in the British national screening programme there was a significant decline (p < 0.0001) in referral rates from 6.2 percent in 1991-2 to 5.4 percent in 1992-3 (National Health Service Breast Screening Programme 1994). However, part of this decline was probably due to a lower prevalence of pre-symptomatic, screen-detectable breast cancer in the screened population in the latter period, since some of the women screened in 1992-3 were undergoing incidence (second) screens. The breast cancer detection rate also declined over the same time from 6.2 to 5.7 per 100 women screened (National Health Service Breast Screening Programme 1994). Also, the availability of previous films for comparison tends to lead to lower referral rates in the incidence screening round (Chamberlain et al 1993).

A high referral rate means that too many women experience unnecessary anxiety related to screening. But it is very difficult to achieve a balance where both sensitivity and specificity are high.
To improve specificity the level of suspicion at film reading must be altered, so that fewer women are referred, but this carries the risk of reducing sensitivity. Fortunately in the pilot programme, although the referral rate was high the biopsy rate and the benign to malignant ratio were low, so that few of the referred women underwent unnecessary biopsies. The radiologists are aware of the relatively high recall rate and are endeavouring to reduce it while still maintaining high sensitivity. This is discussed further in section 7.2.7.

7.2.5 Benign : malignant biopsy ratio
The biopsy rate of 1.96 percent just met the target biopsy rate of less than two percent, and the benign : malignant ratio of 0.9 : 1 was well within the target range of less than 3 : 1 (Chapter Five, section 5.3). The low benign : malignant ratio partly compensated for a high referral rate, in that very few women in the pilot programme underwent unnecessary biopsies. In other programmes the benign : malignant ratio was considerably higher; 4 : 1 in the HIP trial (Shapiro et al 1982), and 2 : 1 in the Swedish Two-Counties trial (Tabar et al 1985) and the Edinburgh trial (Roberts et al 1989). However, the Malmö and Stockholm trials had benign to malignant ratios of 0.7 : 1 and 0.5 : 1 respectively (Andersson et al 1988, Frisell et al 1986, Frisell et al 1991).

7.2.6 Sensitivity
Sensitivity in the pilot programme was estimated at ninety-two percent, with a 95% confidence interval of 84.4 to 96.5 percent (Chapter Five, section 5.4.5). This was a very good result compared with other breast cancer screening programmes, where sensitivity in the first screening round, calculated using the same method, ranged from seventy-eight to ninety-six percent (Chapter Four, Table 4.1). It means that very few women with breast cancer were missed in the pilot programme compared with other programmes. The target set for the pilot programme was eighty-five percent sensitivity. The estimate of sensitivity in the pilot programme depended on identifying all women with interval cancers. Women with interval cancers were identified from copies of histology forms from pathology laboratories in Otago and Southland for all women aged fifty or over who were diagnosed with breast cancer (Chapter Four, section 4.4). This method should have identified all women with interval cancers diagnosed in Otago and Southland. However, if a woman had a negative screen in the pilot programme and then moved to a different region and was subsequently diagnosed with breast cancer outside Otago and Southland, she would not have been identified as having an interval cancer. Six of the 7,109 women who had negative screen results
were identified as having developed interval cancers after screening. The 95% confidence interval was 2.5 to 12.7 (Rothman and Boice 1982). Because the number of interval cancers was small, missing even one interval cancer could mean that sensitivity was overestimated. However, several interval cancers would have to have been missed for sensitivity to have been below the target that was set for the pilot programme. If the actual number of interval cancers had been seven rather than six, sensitivity would have been ninety-one percent rather than ninety-two percent, and only if the actual number of interval cancers had been more than double the number identified, would sensitivity have been below the target of eighty-five percent.

7.2.7 Specificity
Specificity measures the ability of the screening test to correctly identify women who do not have breast cancer. It provides an indication of the number of screened women who were referred unnecessarily for assessment, and thus is a useful quality control measure. The specificity in the first eighteen months of the pilot programme was eighty-nine percent, with a 95% confidence interval of 88.6 to 90.0 percent (Chapter Five, section 5.4.5). This was lower than the specificity in the first round of screening in other programmes, which ranged from ninety-five to ninety-seven percent. The low specificity was related to the high referral rate in the pilot programme. The target set for the pilot programme was ninety-five percent specificity.

One way to increase specificity is to reduce the number of women with false positive results, and this means reducing the referral rate. To do this the radiologists will need to reassess their criteria for referral, so that fewer women are referred for assessment. This is a difficult task because of the possibility that reducing the referral rate in order to increase specificity could reduce sensitivity. But the radiologists now have the advantage of being able to review films for which the outcome is known, and this should help in reviewing the criteria for referral. Such films have been used in the United States and in England, for standardising mammographic interpretation (Breast Cancer Screening Working Group 1987, Elmore et al 1994, Warren and Duffy 1995). Referral rates can be expected to drop in the incidence screening rounds, because for most of the women, the radiologists have access to previous films for comparison (Chamberlain et al 1993). This is further discussed in section 7.5.1.
The targets for sensitivity and coverage for the pilot programme were eighty-five percent sensitivity and seventy percent coverage of the target population. Meeting these targets would have given the programme a maximum breast cancer detection rate of sixty percent, which is close to the lowest detection rate for a successful programme (Chapter Four, section 4.1.2). In fact the Otago and Southland programme had sixty-eight percent maximum detection of breast cancers, based on the seventy-four percent coverage and ninety-two percent sensitivity achieved during the first eighteen months of the programme (Chapter Five, section 5.4.5). This compares very well with other breast cancer screening programmes (Chapter Four, Table 4.3).

7.2.8 Breast cancer diagnosis in the pilot programme

The first round detection rate including ductal carcinoma in situ (DCIS) was 10.2 women with breast cancer detected per thousand women screened with a 95% confidence interval of 7.8 to 12.5 (Chapter Five, section 5.4.2). This detection rate was higher than that in the latest results from the national breast screening programme in Britain where the first round detection rate including ductal carcinoma in situ (DCIS) was 6.3 per thousand (Chamberlain et al 1993). The detection rate for invasive tumours less than or equal to ten millimetres in the British national screening programme was 1.4 per thousand. The target for the British national screening programme had been at least 1.5 per thousand. The Otago and Southland result was 2.9 per thousand (95% confidence interval 1.7 per thousand to 4.1 per thousand).

The detection rate of ten per thousand was 4.5 times the expected breast cancer incidence in the absence of screening and thus easily met the target of more than three times the expected incidence in the absence of screening (Table 7.1). The expected incidence of breast cancer in women aged fifty to sixty-four in Otago and Southland was 2.2 per thousand (Chapter Five, section 5.4.1). This expected incidence was calculated using the most recent ten years of Cancer Registry data available. Surprisingly, the number of women diagnosed with cancer outside the screening programme was much higher than expected based on Cancer Registry data, and this raises the possibility of under-reporting to the Cancer Registry in the past (Chapter Five, section 5.4.4).

As part of the evaluation of the pilot programme, information about all diagnoses of breast cancer in women over fifty years in Otago and Southland has been collected. Once the Cancer Registry releases its registration data for the period corresponding to the first eighteen months of the
screening programme it will be possible to compare the registrations with the information collected directly from the pathology labs in Otago and Southland. This will be a useful quality control check both for the evaluation and for the Cancer Registry. If there had previously been under-reporting of breast cancer to the Cancer Registry, the expected incidence in the absence of screening would have been higher than 2.2 per thousand and the ratio of screen prevalence to expected incidence would be lower than 4.5. However, the ratio would only fall below the target of 3 : 1 if a third of all diagnoses of breast cancer in Otago and Southland had previously not been reported to the Cancer Registry, or in other words, if the true incidence of breast cancer had been over 3.3 per thousand, rather than 2.2 per thousand.

7.2.9 Tumour characteristics
The stage distribution of cancers detected by screening will be affected by lead time. A shift in stage distribution is therefore “necessary but not sufficient” (Day et al 1989) for an effective screening programme. Stage distribution alone cannot be used to determine the efficacy of screening in reducing breast cancer mortality, since it would be possible for a shift in stage distribution to occur even without any reduction in mortality. Size, grade, and nodal status have been used to predict the reduction in breast cancer mortality that can be achieved by screening (Duffy et al 1991, Tabar et al 1992). Duffy et al collected information on women diagnosed with breast cancer in the Swedish Two-Counties trial. In the Swedish Two-Counties trial, differences in size, nodal status, and grade explained the difference in survival between screened women and controls diagnosed during incidence screening rounds. But a recent analysis of survival of patients with breast cancer diagnosed in the United Kingdom Trial of Early Detection of Breast Cancer (Moss et al 1994) found that only about a third of the improved prognosis in women who had been diagnosed by screening could be explained by differences in size and nodal status. It was suggested that the disparity between the British and Swedish results could be partly because the British analysis could not adjust for tumour grade. However, in the Swedish Two-Counties trial the distribution of grades differed significantly between Östergötland and Kopparberg, the two counties in the trial. It was suggested that this was due to “the subjective nature of histological grading” (Tabar et al 1992). If the subjectivity of grading allows such differences, then the utility of grade as a prognostic indicator must be in question. Another possible reason for the discrepancy between the British and Swedish results is that participation was lower in the United Kingdom, and there was self-selection of women with a poor prognosis into the non-participant group.
7.3 The acceptability of the pilot programme

7.3.1 Pre-screening survey
Before the pilot programme began, a random sample of 290 Otago and Southland women was surveyed to establish which aspects of a programme would increase acceptability, thereby enhancing participation (Chapter Six, section 6.1). Aspects of the service that were likely to improve participation included the presence of a screening centre within thirty minutes travel by car (only twenty-eight percent of women surveyed thought more than thirty minutes travel time was reasonable), appointment times outside normal working hours, female staff, and publicity before invitations are sent out.

The pilot programme was designed to incorporate these features, with a mobile unit so that rural women did not have to travel excessive distances for screening, some evening sessions, female radiographers and clerical staff, and publicity about the programme. Publicity was targeted to specific areas such as rural towns in the month before invitations to attend the mobile unit were sent out.

7.3.2 Participation and coverage
The randomised controlled trials of invitation methods (Chapter Five, section 5.1.3), showed that seventy-one percent of women who received general practitioner letters (with reminders for those who did not reply) took part in screening. It was worth sending reminders to women who did not respond initially, as nearly half of these women took part after receiving a second letter. However, participation in response to an invitation only partly accounted for the coverage achieved by the Otago and Southland pilot programme.

Coverage is the proportion of eligible women who were screened in the first screening round of the Otago and Southland pilot programme. A target of seventy percent had been set (Chapter Four, section 4.2.1). At the end of the first screening round 14,129 women had been screened, out of a total of 18,555 women aged fifty to sixty-four and living in Otago and Southland at the 1991 census, with an additional 504 women becoming eligible during the first screening round. This gave the programme seventy-four percent coverage of the eligible women in Otago and Southland.
This compared extremely favourably with other programmes (Chapter Four, Table 4.1, section 4.2.1) where coverage ranged from fifty to seventy-eight percent.

Participation was also calculated, by dividing the number of invited women screened as a result of the invitation, by the number of women invited. This measured the response to the invitations, and was one measure of acceptability (Chapter Six, section 6.2). When the evaluation targets were set, it had been thought that a very high participation rate would be needed, and a target of seventy-eight percent participation was set (Chapter Four, section 4.2.1). The target was set at this high level to compensate for the fact that not all women could be identified and invited for screening, since some eligible women would be on neither the electoral roll nor on general practice registers. Also, it had been expected that almost all the women who attended for screening would do so in response to an invitation. However, in the Otago and Southland programme there was a constant demand from women who had self-referred or had been referred by their general practitioners or through the Radiology Department of Dunedin Hospital. About twenty percent of women had self-referred or been referred. Intended participation had been high in the pre-screening survey, and it may be that some women were very eager to take part in screening. The high level of self-referral and general practice referral meant that the number of women who were screened as a result of an invitation was constantly supplemented by women who had self-referred or had been referred. It also meant that by the time some women received their invitations, they had already been screened. This was unexpected, and although it reflected the positive response to the pilot programme in Otago and Southland, it meant that the participation rate after an invitation was lower than expected, at fifty-eight percent (Chapter Six, section 6.2).

The survey of screened women (Chapter Six, section 6.3) illustrates the relationship between participation and coverage; twenty percent of the respondents had self-referred or been referred by their general practitioners, and eighty percent had attended as the result of an invitation. These results suggested that the ratio of referred to invited women was 1:4. Extrapolating from these results suggests that for every hundred eligible women, fifty-eight percent would respond to an invitation, and fourteen would be self-referred or referred by general practitioners, giving a total coverage of seventy-two percent. This is close to the seventy-four percent coverage measured in the pilot programme.
Whether there will be equally high self-referral and general practitioner-referral rates in other parts of New Zealand is unknown. However, the Waikato pilot programme had a similarly high level of self-referral and general practitioner-referral (Chapter Six, section 6.6). If the pilot programme had had a policy of screening women only after an invitation had been sent (by reassuring women who self-referred that they would definitely be invited, and asking them to wait until they received an invitation), a better indication of the true response to an invitation would have been obtained. But this policy may have dissuaded some women from attending at all. It is helpful to know that seventy-one percent participation can be achieved, with an appropriate invitation signed by a general practitioner. However a limitation of this result is that it was obtained in one large group-practice in Dunedin, which may not be representative of all general practices in New Zealand. It will be important to monitor participation at re-screening in the second round of the pilot programme, to investigate whether women who have already been screened once, maintain similar levels of acceptance of screening. It is encouraging that the acceptability of the Otago and Southland programme is high, with ninety-four percent of screened women intending to return for a second screen (Chapter Six, section 6.3).

Because the target for participation was later found to be inappropriate, given the unexpectedly high numbers of women who were screened without receiving an invitation, coverage is a better measure of the acceptability of the pilot programme than participation, which is why only coverage has been reported in the summary of results in Table 7.1. A crucial issue for the effectiveness of the programme will be the coverage in the second screening round. Most programmes experience a decline in second and subsequent screens (Breast Cancer Screening Working Group 1987).

7.3.3 Post-screening survey
A post-screening survey was carried out to investigate the acceptability of breast cancer screening (Chapter Six, section 6.3). A questionnaire was sent to 474 women, and the response rate was very high, with ninety-three percent of women returning the questionnaires. There was a very high level of satisfaction with all aspects of the screening service, with no difference between screening centres. Ninety-four percent of the women said they would return for another screen in two years. The survey showed that some women did not receive their results within the maximum period initially set for the programme. An additional clerical person was appointed as a result of this
finding. The method of processing results was also changed in response to the results of the survey.

7.3.4 Assessment clinic survey
Questionnaires were sent to 140 women who had been to the assessment clinics (Chapter Six, section 6.4). There was an eighty-nine percent response rate. There was a high level of satisfaction with the assessment clinics, with ninety-five percent of women being satisfied, and ninety-eight percent intending to be screened again. Assessment caused women more anxiety than screening, which underlines the importance of reducing the referral rate in the pilot programme as far as possible without decreasing sensitivity.

7.3.5 Survey of non-attenders
A telephone survey was carried out to compare women who attended for screening with women who declined an invitation to be screened (Chapter Six, section 6.5). Women who did not attend perceived that the worth of screening was outweighed by the practical difficulties of attending. The results suggested that participation in breast cancer screening could be increased by reducing practical barriers to screening, for instance by improving the availability of appointments outside normal working hours, and by increasing women's knowledge about breast screening. However, women should be provided with information about screening with the understanding that, having received the information, it is their right to decide not to take part in screening if they do not wish to. Forty-four percent of the women who did not attend in the first screening round said that they would attend if they received an invitation to the second round. This means that it would be appropriate to invite women who have declined an invitation to screening, to the next screening round.

7.4 The economic effectiveness of the pilot programme
Economic analyses of the pilot programme were planned as part of the evaluation, but because the work required expertise in health economics, the economic analysis was carried out by a consultant health economist and a Masters in Public Health student. A paper presenting the results of an economic analysis of the pilot programme is included as Appendix Two (Menon et al 1994) to
provide background information and so that the economic aspects of breast cancer screening in New Zealand can be commented on here.

One of the three areas which were to be evaluated (Skegg et al 1988) was the economic efficiency of the pilot programme. Two issues were particularly important. First, it was important to discover whether the pilot programme could provide screening in a way that would not be "wasteful of scarce resources" (Skegg et al 1988), and secondly it was important to be able to extrapolate from the pilot programme so that an estimate of the cost of a national breast cancer screening programme could be made. When resources are finite, any decision to implement a particular health intervention means that fewer resources are available for other health interventions. It has been argued therefore that it is unethical not to consider economic factors in allocating health care resources (Williams 1992, Warnock 1994). Economic analysis is used in the health service to help make decisions when choices have to be made between several courses of action, so it is important to find out how breast cancer screening compares with other competing potential uses of health resources. Economic "efficiency" occurs when choices in health care are made that derive the maximum benefit from the resources available (Drummond 1990).

There are three main types of economic analysis; cost-benefit, cost-effectiveness, and cost-utility analyses (Drummond et al 1993, Robinson 1993a). In economic terms cost refers to benefits that are forgone elsewhere, or opportunities that are forsaken by deciding to take a certain course of action (Drummond 1990).

A cost-effectiveness analysis is conducted to find out how much a programme costs to produce a certain effect. This is then expressed in terms of cost per unit of effectiveness, such as the cost per year of life gained, the cost per life saved, or the cost per pain-free day (Robinson 1993b). It is possible to compare different programmes using this approach, provided the outcomes of the programmes can be expressed in the same units of effectiveness.

A cost-utility analysis compares the costs of different procedures, with their outcomes measured in units that are adjusted for a person's well-being, for instance per quality-adjusted life year, or QALY (Robinson 1993c). Because ranking quality of life is a subjective assessment, and varies according to who is doing the ranking, and also with a person's age, state of health or prognosis,
there has been some criticism of the use of QALYs (Smith 1990, Petrou and Renton 1993, Nord 1994). When information on the QALYs obtainable for different procedures is available it is possible to rank the procedures in terms of their costs per QALY. Such a ranking is called a QALY table, and these tables can be used by policy makers to help decide where health resources should best be spent. The major problem with QALY tables is that the methodologies used in determining the costs and the QALYs for various cost-utility analyses may differ. This means that it may be inappropriate to compare certain cost-utility analyses with other analyses. A second problem is that the results of cost-utility analyses may be specific to a certain health system or country, so that international comparisons across countries with widely differing systems of health care may be questionable (Robinson 1993c).

A cost-benefit analysis requires that all the costs of an undertaking, such as a screening programme, (for example staff and equipment for the programme, and travel, time away from work, and anxiety for women) and all the benefits (for instance reduced breast cancer mortality and earlier treatment of breast cancer) are valued in monetary terms. This can be problematic, as many benefits (such as reassurance or extra years of life) and costs (such as anxiety) are difficult to measure in monetary terms (Drummond 1990).

7.4.1 The economic evaluation of the pilot programme

The economic part of the evaluation of the Otago and Southland pilot programme was designed to find out whether the pilot programme was efficient compared with existing programmes, and by extrapolating from the cost of the pilot programme, to provide information on the likely cost of a national breast cancer screening programme.

It was decided that two methods of economic analysis would be used in the evaluation of the pilot programme. First, it would be possible to calculate two costs, the cost per woman screened, and the cost per cancer detected, relatively early in the operation of the programme. These measures had been calculated for other screening programmes, so the costs per woman screened and per cancer detected in the pilot programme could be compared with the equivalent costs in other breast cancer screening programmes. Later, provided the programme reached the targets set for effectiveness, it would be possible to estimate, by extrapolation from randomised controlled trials of breast cancer screening, the number of years of life gained by screening. It would then be possible
to calculate the cost per year of life gained for the pilot programme, and this could be compared with other uses of health resources in New Zealand where similar cost-effectiveness analyses had been done.

The second analysis that was planned as part of the economic evaluation of the pilot programme was a cost-benefit analysis. This would allow policy makers to decide whether the social benefits generated by the programme would outweigh its costs. The costs of the programme would include health service costs such as the costs of staff, equipment, publicity, and hospital-based procedures, and costs to the women including travel costs, the opportunity cost of the time required to participate in the programme, anxiety, any unnecessary investigations (for women with false positive tests), and delay resulting from false negative tests. The benefits of the programme would include savings resulting from fewer women requiring treatment for advanced breast cancer, less invasive treatment, and decreased breast cancer mortality. The preliminary results from the cost-effectiveness analysis are presented as Appendix Two, and are briefly summarised below. The cost-benefit analysis is still under way, as it requires information from the second screening round.

7.4.2 Results from the cost-effectiveness analysis

The cost per woman screened in the first year of screening was $178, which dropped to $117 in the second year, and is expected to be $113 in each year after that, assuming that eight thousand women are screened in the programme each year from year three onwards. On the basis of the cost per woman screened in the pilot programmes in year three, and assuming an eighty percent participation rate, it is estimated that a national screening programme would add between $9.3 and $9.9 million to health service costs each year in New Zealand, in 1991 dollars (Appendix Two).

7.4.3 Is breast cancer screening an efficient use of health resources?

Based on the results from the pilot programmes (Devlin et al. 1993, Menon et al. 1994), the cost to the health service when a national screening programme is introduced in New Zealand will be between 9.3 and 9.9 million dollars per year. The costs will be higher in the early years when the programme is first being set up. It is important to know whether spending this amount of money on breast cancer screening is an efficient use of resources. This can be investigated by comparing the cost-effectiveness of breast cancer screening with the cost-effectiveness of other health care interventions. When the cost per year of life gained and the cost per QALY for the pilot
programme can be estimated, later in the evaluation, it will be possible to compare these with similar measures from other health care programmes in New Zealand. In the absence of New Zealand data at present, (the cost per year of life saved for the New Zealand pilot programmes will not be estimated until there are some results from the second screening round), some indication can be obtained from published reports from other countries.

A cost-effectiveness analysis of breast cancer screening in the Netherlands (de Koning et al 1991) found that the cost per year of life gained from screening women aged fifty to seventy, assuming seventy percent participation, was between US$3,000 and $5,000, and the cost per QALY was between US$3,200 and $5,300. The authors reported that to their knowledge, no superior cost-utility ratios had been reported for other screening programmes or cancer programmes. A second cost-effectiveness analysis, also in the Netherlands found that two-yearly breast cancer screening in women aged fifty to seventy would reduce the total breast cancer mortality in the Netherlands by twelve percent. The costs per death prevented and per life year saved were lower than those in other Dutch health care programmes for which cost-effectiveness ratios had been calculated, including screening for cervical cancer, the treatment of end-stage liver disease, and liver transplantation (van der Maas et al 1989).

A cost-effectiveness analysis of breast cancer screening in Australia (Carter et al 1993) found that, for two-yearly screening in women aged fifty to sixty-nine, the cost per life year gained was less than that in some other Australian studies where cost-effectiveness ratios had been calculated, including screening for cervical cancer, hospital haemodialysis, and the treatment of AIDS with zidovudine.

Table 7.2 is a QALY table which includes the cost per QALY for breast cancer screening (Mason et al 1993). The cost of breast cancer screening in this table is less than the cost of some other procedures (for instance heart transplantation, and haemodialysis). However comparisons such as those in Table 7.2 should be made with caution. Mason et al report that although all the data in this QALY table came from British cost-utility studies, the methods used in the studies differed. In particular, the study investigating neurosurgical interventions discounted neither costs nor benefits. In programmes where costs and/or benefits occur over many years this could make a considerable difference to the cost per QALY. Also, it is sometimes difficult to know whether savings relating
TABLE 7.2

Cost per quality adjusted life year (QALY) of competing treatments

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Cost per QALY (£)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cholesterol testing and diet therapy only (all adults aged 40-69)</td>
<td>220</td>
</tr>
<tr>
<td>Neurosurgical intervention for head injury</td>
<td>240</td>
</tr>
<tr>
<td>Advice to stop smoking from general practitioner</td>
<td>270</td>
</tr>
<tr>
<td>Neurosurgical intervention for subarachnoid haemorrhage</td>
<td>490</td>
</tr>
<tr>
<td>Antihypertensive treatment to prevent stroke (ages 45-64)</td>
<td>940</td>
</tr>
<tr>
<td>Pacemaker implantation</td>
<td>1,100</td>
</tr>
<tr>
<td>Valve replacement for aortic stenosis</td>
<td>1,140</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>1,180</td>
</tr>
<tr>
<td>Cholesterol testing and treatment</td>
<td>1,480</td>
</tr>
<tr>
<td>Coronary artery bypass graft (left main vessel, severe angina)</td>
<td>2,090</td>
</tr>
<tr>
<td>Kidney transplant</td>
<td>4,710</td>
</tr>
<tr>
<td>Breast cancer screening</td>
<td>5,780</td>
</tr>
<tr>
<td>Heart transplantation</td>
<td>7,840</td>
</tr>
<tr>
<td>Home haemodialysis</td>
<td>17,260</td>
</tr>
<tr>
<td>Coronary artery bypass graft (one vessel disease, moderate angina)</td>
<td>18,830</td>
</tr>
<tr>
<td>Continuous ambulatory peritoneal dialysis</td>
<td>19,870</td>
</tr>
<tr>
<td>Hospital haemodialysis</td>
<td>21,970</td>
</tr>
<tr>
<td>Neurosurgical intervention for malignant intracranial tumours</td>
<td>107,780</td>
</tr>
</tbody>
</table>

to procedures have been included as benefits, or subtracted from the costs in cost-utility analyses. QALY tables should ideally include only studies that were undertaken in similar health care systems, and using similar methodology, including consistent use of discounting. They provide a guide to policy makers, but because of the potential problems outlined above they should be interpreted with caution.

7.4.4 Savings resulting from screening

The estimates in the cost-effectiveness analysis of the pilot programme (Appendix Two) did not take into account any savings related to the earlier detection of breast cancer. Breast screening results in fewer women requiring treatment for advanced disease, with associated savings to the health service. In terms of cost to the health service alone, it has already been shown that it is initially more expensive to screen than not to screen, but in Sweden it was found that five years after the introduction of screening in Kopparberg county, the cost of continued screening was less than the saving in treatment costs that had resulted from fewer women needing treatment for advanced breast cancer (Tabar 1987). In other words, after five years of screening it would have been more expensive to discontinue the breast cancer screening programme than to continue it. In the Netherlands it was found that a third of the cost of breast cancer screening was offset by savings in treatment costs (van der Maas et al 1989), and in the longer term it is estimated that almost half of the annual cost of breast cancer screening in the Netherlands will be offset by savings in the cost of treatment for advanced disease (de Koning et al 1992). Two analyses in Australia estimated that savings resulting from reduced treatment costs would offset the cost of screening by about eighteen percent (Hall et al 1992, Salkeld and Gerard 1994). The difference between these estimates and those for the Netherlands was attributed to “differences in treatment patterns” there being no generally accepted protocol for the treatment of breast cancer in Australia (Salkeld and Gerard 1994). In addition, both Australian analyses excluded nursing home and hospice costs associated with advanced breast cancer, while the Dutch study by de Koning et al included these costs.

If there were similar savings in treatment costs in New Zealand, the estimated annual cost of a national breast cancer screening programme (between $9.3 and $9.9 million) could drop to between $7.6 and $8.1 million, based on the Australian reduction of eighteen percent (Hall et al 1992, Salkeld and Gerard 1994) or even to between $4.7 and $5 million, based on the long-term
Dutch estimate (de Koning et al 1992). This projected decrease in costs should be taken into account when the national breast cancer screening programme in New Zealand is considered, especially over the longer term.

7.5 Implications

It is important to examine the results from the pilot programme with respect to the future of the Otago and Southland pilot programme, but also with respect to a national breast cancer screening programme in New Zealand.

7.5.1 What do these results mean for the Otago and Southland pilot programme?

The pilot programme met the targets that had been set for identification of eligible women, coverage, sensitivity, the biopsy rate, the benign to malignant biopsy ratio, and the screen prevalence. The acceptability of the programme was high, and its cost was similar to that of other breast cancer screening programmes. However, the pilot programme failed to reach the targets set for participation, referral rate, and specificity.

In terms of its likely effect on breast cancer mortality, the programme matched all the important predictors of success, including coverage, sensitivity, and screen prevalence. The high sensitivity in the pilot programme meant that very few women with breast cancer were missed, and as a result, the screen-detected prevalence was more than three times higher than the incidence expected in the absence of screening. The stage and size distributions in the pilot programme were similar to those in the prevalence screen of the Swedish Two-Counties study, but a slightly higher proportion than expected were node-positive. It is suggested that at least seventy percent of tumours detected by screening should be node-negative, whereas sixty-three percent of those detected in the pilot programme were node-negative. The proportion of node-negative tumours is likely to increase in the second and subsequent screening rounds (Tabar et al 1992).

The results for sensitivity, tumour size and stage suggest that the main benefit of screening, a reduction in breast cancer mortality, is likely to be achieved in the longer term. But the referral rate in the pilot programme was higher than expected, and accordingly, specificity was low. As a result
of this too many women underwent unnecessary investigations in the pilot programme, and this is a negative aspect of screening which every programme aims to minimise. Fortunately very few of these women had unnecessary biopsies, and the biopsy rate and benign to malignant ratio met the targets that had been set. In Chapter Four (Table 4.4) it was suggested that failure to reach this target should result in a reassessment of the criteria used for referral. In the pilot programme the focus is now on reducing the referral rate. This should continue to be evaluated, against second round screening targets (Table 7.3), in the second screening round.

The results from the surveys of women showed that the programme is acceptable to most women, and there is a very high rate of intended participation in the next screening round. But intended participation does not always translate into actual coverage (although the intended participation rate in women aged fifty to sixty-nine in the pre-screening survey was seventy-nine percent, and seventy-four percent of the eligible population were actually screened in the first screening round). Because of the high level of self-referral in the first screening round, a target for participation became almost irrelevant (section 7.3.2). In terms of evaluating the proportion of eligible women screened, coverage is the more important target. It will be important to measure coverage in the second screening round.

Information collected during the first round of screening showed that the acceptable results for effectiveness and acceptability described above were accomplished at a similar cost to other breast cancer screening programmes. A cost-benefit analysis is being undertaken to determine whether in New Zealand the benefits of screening outweigh the costs.

7.5.2 Targets for the incidence screening rounds

The results from the pilot programme suggest that the programme is acceptable, and is likely to be effective and economically efficient, according to the standards set for the evaluation. However, the programme is still in its early stages, and continued evaluation is required, especially since many of the targets set for the first round of screening do not apply to subsequent rounds. Targets for the incidence screening rounds should be developed. Suggested targets are given in Table 7.3. An explanation for the choice of these targets follows.
TABLE 7.3

Suggested targets for the incidence screens of the Otago and Southland pilot programme

<table>
<thead>
<tr>
<th>Characteristic to be evaluated</th>
<th>Target:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of eligible women</td>
<td>90%</td>
</tr>
<tr>
<td>Coverage</td>
<td>65%</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>85%</td>
</tr>
<tr>
<td>Specificity</td>
<td>95%</td>
</tr>
<tr>
<td>Percent of screened women referred</td>
<td>5%</td>
</tr>
<tr>
<td>Biopsy rate (% of screened women)</td>
<td>&lt;2%</td>
</tr>
<tr>
<td>Benign:malignant ratio of biopsies</td>
<td>&lt;1:1</td>
</tr>
</tbody>
</table>
The target for identification of eligible women remains unchanged as this is unaffected by which screening round is taking place. It is clearly possible to identify ninety percent of eligible women using the electoral roll and general practice registers, but because of changes of address, the electoral roll rapidly becomes out of date. If it is used as the basis for a screening register regular updates of the roll have to be purchased from the electoral roll centre. A target for participation has not been included, since it was not particularly relevant due to the high rate of self-referral in the first round of the pilot programme. It would be interesting to measure participation in response to an invitation and self-referral rates in the second screening and subsequent rounds however, to see whether these change compared to the first round. The more important measure is the coverage (or what proportion of the eligible women are screened), and this target has been reduced slightly, in recognition of the results from other screening programmes, where the percentage of eligible women screened invariably declined in the second round compared with the first screening round, ranging from fifty-five to seventy percent (Breast Cancer Screening Working Group 1987).

The targets for sensitivity and specificity remain unchanged, but the target for the referral rate has been reduced because the underlying prevalence of breast cancer is lower in incidence screens, and since, in many cases, the radiologists will have the benefit of previous films for comparison. Both of these factors tend to lower the referral rate in incidence screening rounds compared with the prevalence screening round (Chamberlain et al 1993, National Health Service Breast Screening Programme 1994). If the programme meets the referral rate target in the second screening round it should also meet the target for specificity.

The target biopsy rate has been set at less than two percent. In the prevalence screen the biopsy rate was 1.96 percent, which, combined with a low benign to malignant ratio (0.9 : 1) is a good result since it means that not too many women underwent unnecessary biopsies. In the second screening round it should be possible to reduce the biopsy rate further, since the underlying prevalence of breast cancer will be less in the second round, with most of the women having been screened two years previously.

The target benign : malignant biopsy rate has been reduced for the incidence screening rounds, so that it is in line with ratios reported in European screening programmes, and because the ratio usually drops after the first screening round (Kopans and Swann 1988, Stockdale et al 1988). The
pilot programme should meet this target easily, since the benign : malignant biopsy ratio in the prevalence screen was already below 0.9 : 1.

7.5.3 What do these results mean for a national breast cancer screening programme?
The purpose of the evaluation was to investigate the acceptability, effectiveness, and economic efficiency of population based mammographic screening in New Zealand. The results of the evaluation of the Otago and Southland pilot programme showed that it is possible for an acceptable and effective breast cancer screening programme to be carried out in New Zealand. The preliminary results of the economic analysis showed that breast cancer screening can be accomplished in New Zealand at a cost similar to that of breast cancer screening programmes in other countries. Using cost data from the pilot programmes, an estimate was made of the likely cost of a national breast cancer screening programme in New Zealand.

The results from the evaluation of the Otago and Southland pilot programme suggest that, if the national breast cancer screening programme can be carried out to a similar standard, it would start to reduce breast cancer mortality after about five years (Chapter One, section 1.2), among women aged fifty to sixty-four. Each year as a result of screening, about fifty-eight fewer women would die from breast cancer in New Zealand (Chapter Two, section 2.2). The cost of a national breast cancer screening programme in New Zealand is estimated to be between $9.3 and $9.9 million per year initially. Later, this cost will be partly offset by reduced treatment costs (section 7.4.4), and could drop to between $7.6 and $8.1 million, based on Australian estimates (Hall et al 1992, Salkeld and Gerard 1994) or even to between $4.7 and $5 million, based on Dutch estimates (de Koning et al 1992).

7.6 Recommendations for the national breast cancer screening programme

The New Zealand Ministry of Health received interim reports from the Otago and Southland, and Waikato pilot programmes in September 1993 (Richardson et al 1993, Health Waikato Ltd 1993). In June 1995, the Minister of Health announced plans for the introduction of a national breast cancer screening programme in New Zealand. The national breast cancer screening programme is
to be fully implemented by the end of 1998 (Ministry of Health, 1995a). The decision to implement a national breast cancer screening programme was made for four reasons:

- breast cancer is a significant health problem in New Zealand
- there is now very strong evidence of the efficacy of breast cancer screening in reducing breast cancer mortality among women aged 50-64 years
- studies show that breast cancer screening is value for money relative to other health interventions
- the early results of the pilot programmes demonstrate that screening can be done effectively and efficiently in New Zealand.

(Ministry of Health, 1995b)

The national programme is to be an organised, population-based programme which will offer free two-view mammographic screening every two years to all asymptomatic women aged from fifty to sixty-four in New Zealand. The national breast cancer screening programme will be modelled on the two pilot programmes (Ministry of Health 1995a).

The following recommendations are made for the New Zealand national breast cancer screening programme.

7.6.1 Organisation of the screening programme
The best way to minimise risks and ensure that screening reduces breast cancer mortality is to offer screening in a coordinated fashion, ensuring that each stage of the screening process is of high quality (Hakama et al 1985). Central coordination of the national breast cancer screening programme will be essential. It will not be adequate to contract out aspects of the programme piecemeal, because it is very important that all the stages of screening are coordinated to ensure timely and appropriate follow-up and adequate facilities for assessment and treatment (especially in the prevalence screen, when the flow-on from screening is highest).

Setting up a breast cancer screening service will require substantial managerial effort. Considerable co-operation between health authorities or boards, hospital specialists, Family Practitioner Committees and general practitioners will be needed. There will be training needs to be met and there may be difficulties in attracting suitable staff to a relatively narrow field of practice. The specialist assessment teams will need careful organisation and co-operation between different
authorities. Experience gained in the few existing breast cancer screening centres and the cervical cancer screening programme will be invaluable during the development of this new service.

(Breast Cancer Screening Working Group 1987)

The evaluation team was not charged with evaluating the way the Otago and Southland pilot programme was managed, but it became clear during the evaluation that the management of the programme was crucial to its success. Management of a breast cancer screening programme is very complex. It involves coordinating the efforts of people from many disciplines, and in different centres, often in a hospital setting, but in a programme dealing with well women.

Programme managers should have an overview of all screening, further investigation and treatment activities associated with the programme, coordinate all components and make decisions for monitoring and evaluation. They are responsible for identifying and calling all eligible women for mammographic examinations and ensuring that those with 'abnormal' findings are recalled for further investigation and referred to a specialist assessment team.

(Austoker 1990)

Each stage of screening must be carried out to a high standard if the programme as a whole is to be successful. The stages of screening include identifying and inviting eligible women, taking the screening mammograms, interpreting the films, informing women and their general practitioners of the results, referring some women to the assessment clinic, carrying out appropriate assessments, interpreting the results, informing women of these results, and referring women for appropriate treatment if required, and recalling women to the next screening round.

These stages require coordination between clerical staff, radiographers, radiologists, surgeons, pathologists, medical physicists, and breast care nurses or counsellors. For each stage to be carried out to a high standard experienced and dedicated staff are essential. Good management means that women who take part in screening experience a smooth and rapid transition from one stage to the next, and the staff working in the programme feel part of a team rather than isolated individuals offering part of a fragmented service (Breast Cancer Screening Working Group 1987, Austoker 1990). Fortunately the Otago and Southland pilot programme had an experienced manager with both clinical and management qualifications, and very dedicated staff. It will be essential that appropriate staff are recruited for the national breast cancer screening programme (section 7.6.2).
7.6.2 Programme staff
All the staff involved in the screening programme should be expert. A screening programme requires radiographers with expertise and experience in taking screening mammograms, a medical physicist with the expertise to calibrate and maintain the equipment so that the best images are produced at the lowest radiation exposure, radiologists with appropriate experience in reading screening films, pathologists with expertise in breast cancer histology, and surgeons with experience in treating women with screen-detected early breast cancer. The level of expertise of surgeons has been shown to affect the outcome for women with breast cancer (Gillis and Hole 1996). It would be sensible for the pilot programmes to become training centres for the national breast cancer screening programme. This approach was suggested and used successfully in Britain (Breast Cancer Screening Working Group 1987). The coordinator of the national breast cancer screening programme should be a person who has experience in managing a screening programme. Management of a screening programme is complex in a regional programme let alone a national programme, and requires a person with appropriate experience and expertise.

7.6.3 Invitation method
Personal invitations should be used, since these have been shown to increase participation in screening (Hobbs 1986, Williams and Vessey 1990), and contributed to the high coverage of the eligible population in the pilot programme (Chapter Five, section 5.1.3, Chapter Six, sections 6.2 and 6.6). The high level of self-referral that occurred in the Otago and Southland and Waikato pilot programmes may not occur in other parts of New Zealand, for instance major urban centres such as Auckland, or very isolated areas such as the West Coast of the South Island, so it will be important to have an efficient and effective system of inviting eligible women. The Otago and Southland pilot programme did not investigate alternative ways of inviting Maori women to be screened, since only about two percent of the eligible women in Otago and Southland were Maori. The Waikato pilot programme investigated Marae-based screening (Health Waikato Ltd. 1993), and is still developing appropriate ways to invite Maori women for screening. These results will be very important for the national breast cancer screening programme.

7.6.4 Referral rates
The pilot programme achieved excellent results by international standards, in most areas, including the sensitivity of the screening test. However, the specificity of the screening test was lower than
expected. Low specificity is directly related to a high referral rate, and this is a major challenge for screening in New Zealand. Women who were referred for assessment but did not have breast cancer did not benefit from taking part in the programme, but experienced unnecessary anxiety and investigations. Fortunately for most of these women, the investigations were not invasive, since the pilot programme achieved a low biopsy rate and a very good benign to malignant biopsy ratio.

Ideally a screening programme should combine a reduction in breast cancer mortality among women diagnosed with cancer, with the lowest possible risk for other women who take part in the programme. To achieve this, the referral rate in the national breast cancer screening programme must be lower than it was in the pilot programme. Methods to reduce the referral rate in the pilot programme should continue to be investigated so that appropriate referral criteria are adopted and can be taught to other radiologists who wish to read screening mammograms in the national breast cancer screening programme (Chapter Five, sections 5.3 and 5.4, Chapter Seven, section 7.2.7). Two-view mammography and double reading of films should be continued since these increase sensitivity and reduce referrals without markedly increasing costs (Wald et al 1995, Williams et al 1995).

7.6.5 Assessment of women with abnormal films

There should be specialist breast assessment clinics and clearly defined surgical and treatment protocols for the national screening programme (Forrest 1989, Aitken et al 1992). It has been shown that the ten-year survival of women with breast cancer treated by specialist breast surgeons is eight percent better than women treated by non-specialist surgeons (Gillis and Hole 1996). There was a sixteen percent (95% CI 6% to 25%) reduction in risk of dying over ten years, after adjustment for age, tumour size, socioeconomic status, and nodal involvement for women treated by specialist breast surgeons (Gillis and Hole 1996). This is probably related to specialist surgeons offering the optimum treatment, which has been shown to improve survival (Early Breast Cancer Trialists' Collaborative Group 1990) in women with early breast cancer (Chapter Two, section 2.2.3). The establishment of a nationwide network of specialist breast assessment centres is likely to improve treatment for symptomatic women as well as for women diagnosed through screening. In Britain, the National Health Service Breast Screening Programme's quality assurance guidelines for surgeons inspired the development of new guidelines for the treatment of symptomatic women (National Health Service Breast Screening Programme 1994).
7.6.6 Involvement of general practitioners

It has been shown that the response to an invitation is higher if the invitation is signed by the woman's general practitioner (Chapter Five, section 5.1.3). The other advantages of involving general practitioners in the screening programme are the avoidance of inappropriate invitations (for instance to women who are already being followed up after a diagnosis of breast cancer), and that women can discuss the invitation and the screening programme with their general practitioners, who will know when women in their practices are being invited and can provide support for women who require further assessment and investigation as part of the programme. A possible disadvantage to women of this involvement is the cost associated with consulting their general practitioners about breast cancer screening.

7.6.7 Screening programme computer systems

The screening programme should have a stand-alone computer system (preferably the same computer system should be used for all centres within the national programme). It is essential that all the screening centres can produce the information required for the management and evaluation of the programme quickly and easily. The computer system used for the Otago and Southland pilot programme was unsatisfactory (Chapter Four, section 4.4, Chapter Seven, section 7.1.1). In New Zealand we have the opportunity to learn from the experiences of other national screening programmes that are already in existence; in the United Kingdom limitations of the computer software used caused problems with the evaluation of the national breast cancer screening programme (Chamberlain et al 1993). Ideally the system chosen for the New Zealand national breast cancer screening programme should be one that has already been used successfully in breast cancer screening programmes. The system should be capable of managing appointments, storing results, flagging women who are due to be invited to each screening round, and generating invitation letters, results, and reports. A data manager and programming staff should be appointed to maintain the computer system and to ensure the efficient production of the data required for evaluating the national programme.

7.6.8 Evaluation and quality assurance

Evaluation of the pilot programmes should continue, so that information about the second and subsequent screening rounds can be collected, since this will be relevant to the national screening programme. The national programme should also be evaluated, to ensure that it offers screening in
a way that is acceptable, effective, and efficient. The information to be collected from each screening centre must be clearly specified, and it must be easy to obtain this information from each screening centre computer system. Appropriate standards for quality must be set and maintained throughout the national programme. The British national breast cancer screening programme has a quality assurance programme which is coordinated by national committees with representatives from each of the disciplines involved in the programme (Austoker 1990). A national screening evaluation unit specifies, checks, and analyses the data collected from each screening centre in the national breast cancer screening programme, and compares the results against an agreed standard (National Health Service Breast Screening Programme 1994). Equivalent mechanisms for quality assurance and evaluation should be incorporated in the New Zealand national breast cancer screening programme.

7.7 Conclusions

This evaluation of the Otago and Southland pilot programme was carried out to establish whether breast cancer screening could be carried out effectively, acceptably, and efficiently in New Zealand, and to provide information that was used to plan a national breast cancer screening programme. A goal-attainment evaluation method was used, where information was collected and compared against a standard derived from successful screening programmes in other countries.

The pilot programme met the targets that had been set for identification, coverage, sensitivity, biopsy rate, benign to malignant ratio, and screen prevalence (detection rate). These targets relate to the effectiveness of the pilot programme in detecting early breast cancer, and suggest that the Otago and Southland programme will reduce breast cancer mortality in the longer term, among screened women. The pilot programme failed to meet the targets set for specificity and the referral rate. Failure to meet these targets will not adversely affect breast cancer mortality but meant that too many women underwent unnecessary investigations and anxiety. The programme staff are aware of these results, and efforts are being made to reduce the referral rate and improve specificity, while maintaining high sensitivity.
Surveys of women who took part in the pilot programme showed that it provided an acceptable service and that most women plan to continue to participate in screening. This is encouraging as the effectiveness of the programme is enhanced if a high proportion of the eligible population is screened. The programme achieved these results at a cost that was comparable with similar programmes in other countries. Information from the pilot programme was used to estimate the cost of a national breast cancer screening programme in New Zealand.

The results of the evaluation of the Otago and Southland pilot programme influenced the decision to implement a national breast cancer screening programme in New Zealand (Ministry of Health 1995b), and the national breast cancer screening programme will be modelled on the Otago and Southland and Waikato pilot programmes (Ministry of Health 1995a). However there are still some issues which could affect the national screening programme, but could not be investigated fully in the pilot programmes, since the pilot areas could not be expected to be representative of the whole country.

An important issue will be finding the appropriate way to invite women and deliver screening to women in very isolated areas such as the West Coast of the South Island. Such women may have to travel considerable distances for screening, despite having mobile units available, and even further for assessment and treatment. Identifying and inviting highly mobile inner-city women, for instance in Auckland, will also be a challenge for the national programme. Another issue will be how to recruit and train adequate numbers of appropriate staff for a national screening programme. If the two pilot programmes become training centres for the rest of the country, it will be important to find a means to provide training while still maintaining the performance of the pilot programmes. A further issue will be establishing the best way to coordinate local and regional programmes into an acceptable, effective, and efficient national programme.

This method of evaluation has provided an early indication of the performance of the pilot programme compared with other breast cancer screening programmes. The method provided a rigorous evaluation of the acceptability, effectiveness, and economic efficiency of breast cancer screening in New Zealand. Breast cancer screening in New Zealand has been subjected to closer evaluation than most health interventions:
The evidence both for the overall benefit of the technique and for its value as applied in New Zealand is better than we have for the great majority of health interventions, and will justify a commitment to making breast cancer screening available to all women in the appropriate age range in New Zealand.

(Elwood 1994)

The evaluation method also provided rapid feedback for programme staff so that the successful aspects of the programme could be maintained, and changes made to any less successful aspects, as the pilot programme progressed.

The use of targets such as those used in this evaluation, is becoming more common in New Zealand and in other countries, in the planning and evaluation of health care. The World Health Organisation (WHO) global strategy for Health for All by the Year 2000 encourages member nations to set health goals and targets (World Health Organisation 1982). Australia has been developing and using health goals and targets since 1985 (Health Targets and Implementation Committee 1988, Australian Institute of Health 1990, Leeder 1995), and “Healthy People 2000” is a national strategy which uses targets to assess progress on improving health and preventing disease in the United States (Feinleib 1995). New Zealand introduced national health goals and targets in 1989 (Ministry of Health 1989).

There is also increasing emphasis on evidence-based medicine in the provision of health care. Evidence-based medicine promotes the use of interventions that are based on sound evidence from randomised controlled trials (Sackett and Rosenberg 1995, Glasziou and Irwig 1995). The system of targets used in this evaluation was independently developed after reviewing the relevant literature on breast cancer screening. Because the targets were derived from programmes which had proven efficacy demonstrated in randomised controlled trials, the targets provided a rigorous assessment of the performance of the breast cancer screening pilot programme. By linking the targets to evidence from randomised controlled trials, this method of evaluation ensures that evidence-based medicine is put into practice. Similar evaluation methods could be used for other medical interventions where there is good evidence of efficacy from randomised controlled trials.
REFERENCES


Bloom HJG, Richardson WW. Histological grading and prognosis in breast cancer. A study of 1409 cases of which 539 have been followed for 15 years. Br J Cancer 1957; 11: 359.


226


227


228


Irwig L, Turnbull D, McMurchie M. A randomised controlled trial of general practitioner-written invitations to encourage attendance at screening mammography. Com Hlth Studies 1990; 14: 357-64.


229


McKinnon H. Personal communication from manager of Waikato pilot breast cancer screening programme, 1996.


233


Seidman H, Gelb SK, Silverberg E, LaVerda N, Lubera JA. Survival experience in the breast cancer detection demonstration project. CA 1987; 37: 258-90.


Smart CR. Highlights of the evidence of benefit for women aged 40 to 49 years from the 14 year follow up of the Breast Cancer Detection Demonstration Project. Cancer 1994; 74 (1 suppl); 296-300.


Wilson JMG, Jungner G. Principles and practice of screening for disease. WHO Public Health Papers 1968; No. 34.


APPENDIX ONE
Impact of independent double reading of mammograms from the inception of a population-based breast cancer screening programme

S. M. Williams,* T. C. A. Doyle,* S. Chartres,* A. K. Richardson*† and J. M. Elwood*‡

*Department of Preventive and Social Medicine, †Department of Radiology, ‡Hugh Adam Cancer Epidemiology Unit, Otago Medical School, Dunedin, New Zealand

SUMMARY. The Otago-Southland Breast Screening programme was set up to find out how acceptable and effective breast screening would be for New Zealand women aged between 50 and 64 years. This report examines the performance of the radiologists in the first 18 months of the pilot programme. The majority, 5659 (80.0%), of the two view mammograms from 7074 women, were read independently by two radiologists. Women, who both radiologists believed had an abnormality, were automatically referred for further assessment. The mammograms of those thought to have an abnormality by only one were reconsidered before a referral by consensus was made. The majority of cancers (58) were detected in mammograms read by radiologists A and B. The positive predictive value for those read as abnormal by both was 12.4%. If only one had read the mammograms the number of cancers detected would have been 56 or 55 depending on the radiologist. Overall 73 women were found to have cancer, 67 of whom were identified by two radiologists. The positive predictive value for all the radiologists was 10.5% (95% CI 7.7–12.4).

The inter-observer agreement for radiologists A and B about whether or not a woman should be referred for further assessment measured by kappa was 0.65.Because of the debate about whether or not kappa is the most appropriate measure of agreement other measures are also presented. Alpha which is based upon the idea that mammograms fall into two classes, those that can be consistently classified by two observers according to some well defined rules and those, that because they are more difficult, are classified by chance. In this study alpha was 0.82 for radiologists A and B suggesting that they were in accord, taking account of chance, for 82% of the mammograms. The intraclass correlation, a measure of agreement among all radiologists was 0.64. These measures show higher levels of agreement than have been reported in other larger studies.

INTRODUCTION

Recent studies have shown that double reading of screening mammograms increases the number of cancers detected, but also increases the number of women who are unnecessarily recalled. Other studies have investigated the agreement between radiologists taking part in mammography screening programmes.

The Otago-Southland Breast Screening Programme is the first population-based programme in New Zealand: since 1991 all women aged 50-64 years have been offered 2-yearly two-view mammography, with an uptake rate of 65–75% and outcomes similar to those of European programmes. Mammograms were assessed with only two decision options: normal, that the woman should be offered routine screening again in 2-years time, or abnormal, that the woman should be referred for further assessment. Films were read independently by two radiologists, providing an opportunity to compare double reading with single reading, and to assess the agreement between readers.

PATIENTS AND METHODS

Data from the first 18 months of screening were available. Almost all mammograms were read by two radiologists. All women who reported breast symptoms were automatically referred for further assessment irrespective of results of their mammograms, and so were excluded from this analysis. The referral rate, detection rate of breast cancer, and positive predictive value (PPV) were calculated. Several measures of inter-reader agreement were used. Positive agreement is the ratio of the proportion described as abnormal by both observers to the mean proportion of those described as abnormal by each observer. Negative agreement is defined in a similar way. Kappa measures the level of agreement in
excess of chance. PABAK is a measure of agreement that takes account bias and prevalence. Another measure of agreement, alpha, is based upon postulating a mixed model for the observations so that a proportion are classified consistently and a more difficult sub group by chance. Alpha, a kappa-like statistic, provides an estimate of the proportion classified consistently by both observers. The bootstrap method was used to compute the 95% confidence intervals. A log linear model using a quasi independence model was used to see whether the level of agreement changed over time for the two radiologists who read the majority of the films.

RESULTS
In the 18-month period, 7182 women attended for screening. Of these, 6 had no films taken and 85 were excluded because they reported symptoms and were referred to the assessment clinic automatically. A further 16 were excluded because their films were read twice by the same reader and one because her films were the only set read by an 11th radiologist. This report was based upon the remaining 7074 women whose mammograms were read independently by 2 of 10 different radiologists.

Cancer detection, referral, and positive predictive value
Two radiologists, readers A and B, each independently read 5659 sets of mammograms, 80% of the total (Table 1). Of these, 7.5% were regarded as having an abnormality by both readers, an additional 1.9% were so assessed by reader A, and an additional 4.8% by reader B (Table 1); thus, reader A assessed 9.4% of women as having an abnormality, compared with 12.3% by reader B ($P < .001$).

Overall 14.3% of screenings were regarded as abnormal by one or both radiologists (Table 2). Assessed by 2-month time periods, this proportion was high in the first two time periods, then fell and rose again later, $\chi^2 (8 \text{ df}) = 28.1, P < 0.001$. Where only one of the radiologists initially found an abnormality, the films were reassessed by both radiologists and a consensus decision taken about referral for further assessment. The proportion referred to the assessment clinic declined, over time, from 84% in the first 2 months to 65% in the most recent period, a significant trend $\chi^2 (8 \text{ df}) = 49.6, P < 0.001, \chi^2$ (for trend) $= 34.3, P < 0.001$.

Of 582 women assessed 58 were found to have breast cancer, giving a PPV based on the assessment decision of 10.0%. The PPV was lower in the first two bimonthly periods than later (Table 2), but the trend over time was

<table>
<thead>
<tr>
<th>Classification of each set of mammograms, by reader</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time period</td>
</tr>
<tr>
<td>September–October 1991</td>
</tr>
<tr>
<td>November–December 1991</td>
</tr>
<tr>
<td>January–February 1992</td>
</tr>
<tr>
<td>March–April 1992</td>
</tr>
<tr>
<td>May–June 1992</td>
</tr>
<tr>
<td>July–August 1992</td>
</tr>
<tr>
<td>September–October 1992</td>
</tr>
<tr>
<td>November 1992–January 1993*</td>
</tr>
<tr>
<td>February–March 1993</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* No screening in December 1992.

<table>
<thead>
<tr>
<th>Time period</th>
<th>Total</th>
<th>Total abnormal by either reader</th>
<th>Total referred</th>
<th>Cancer</th>
<th>PPV</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>September–October 1991</td>
<td>407</td>
<td>57</td>
<td>48</td>
<td>84.2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>November–December 1991</td>
<td>167</td>
<td>28</td>
<td>20</td>
<td>71.4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>January–February 1992</td>
<td>417</td>
<td>44</td>
<td>35</td>
<td>79.5</td>
<td>3</td>
<td>9</td>
</tr>
<tr>
<td>March–April 1992</td>
<td>785</td>
<td>83</td>
<td>63</td>
<td>75.9</td>
<td>8</td>
<td>13</td>
</tr>
<tr>
<td>May–June 1992</td>
<td>709</td>
<td>85</td>
<td>65</td>
<td>76.5</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>July–August 1992</td>
<td>870</td>
<td>143</td>
<td>100</td>
<td>69.9</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>September–October 1992</td>
<td>667</td>
<td>116</td>
<td>83</td>
<td>71.6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>November 1992–January 1993*</td>
<td>678</td>
<td>109</td>
<td>75</td>
<td>68.8</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>February–March 1993</td>
<td>959</td>
<td>142</td>
<td>93</td>
<td>65.4</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>5659</td>
<td>807</td>
<td>582</td>
<td>72.1</td>
<td>58</td>
<td>10.0</td>
</tr>
</tbody>
</table>

* No screening in December 1992.
Table 3 Prevalence of an abnormality, and measures of agreement for two radiologists by time period

<table>
<thead>
<tr>
<th>Time Period</th>
<th>Prevalence</th>
<th>Positive agreement</th>
<th>Negative agreement</th>
<th>Kappa</th>
<th>PABAK</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. September–October 1991</td>
<td>0.12</td>
<td>0.79</td>
<td>0.97</td>
<td>0.76</td>
<td>0.90</td>
<td>0.87</td>
</tr>
<tr>
<td>2. November–December 1991</td>
<td>0.12</td>
<td>0.56</td>
<td>0.94</td>
<td>0.51</td>
<td>0.80</td>
<td>0.73</td>
</tr>
<tr>
<td>3. January–February 1992</td>
<td>0.07</td>
<td>0.63</td>
<td>0.97</td>
<td>0.90</td>
<td>0.88</td>
<td>0.84</td>
</tr>
<tr>
<td>4. March–April 1992</td>
<td>0.08</td>
<td>0.68</td>
<td>0.97</td>
<td>0.62</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td>5. May–June 1992</td>
<td>0.09</td>
<td>0.65</td>
<td>0.97</td>
<td>0.62</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td>6. July–August 1992</td>
<td>0.12</td>
<td>0.68</td>
<td>0.95</td>
<td>0.63</td>
<td>0.84</td>
<td>0.79</td>
</tr>
<tr>
<td>7. September–October 1992</td>
<td>0.14</td>
<td>0.73</td>
<td>0.96</td>
<td>0.68</td>
<td>0.85</td>
<td>0.82</td>
</tr>
<tr>
<td>8. November 1992–January 1993</td>
<td>0.13</td>
<td>0.72</td>
<td>0.96</td>
<td>0.68</td>
<td>0.86</td>
<td>0.82</td>
</tr>
<tr>
<td>9. February–March 1993</td>
<td>0.11</td>
<td>0.69</td>
<td>0.96</td>
<td>0.65</td>
<td>0.86</td>
<td>0.81</td>
</tr>
<tr>
<td>Total</td>
<td>0.11</td>
<td>0.69</td>
<td>0.96</td>
<td>0.65</td>
<td>0.86</td>
<td>0.82</td>
</tr>
</tbody>
</table>

95% confidence intervals (0.62–0.67) (0.80–0.84)

*Mean of proportion described as abnormal by radiologists; †number described as abnormal by both/mean number described as abnormal; ‡number described as normal by both/mean number described as abnormal.

not statistically significant. The PPV based on an initial recognition of abnormality by either radiologist was 7.2%. The PPV was higher for women in whom both radiologists found an abnormality, 53 cancers in 426 women (12.4%, 95% CI 9.3–16.3%); compared with five of 381 women (1.3%, 95 CI 0.4–3.1) for those in whom only one radiologist found an abnormality. Based on all women regarded as abnormal by either radiologist independently, the PPV for reader A was 10.3% and that for reader B 8.0%; this difference was not significant (P = 0.17).

Eight other radiologists were also involved in the programme, but none read more than 400 sets of screening films. Those described as abnormal by one of the radiologists were reviewed, usually by radiologists A and B, although occasionally a third radiologist was consulted, before a decision to refer the woman to the assessment clinic was made. Altogether 73 cancers were detected, 67 of which were found by two radiologists, making the positive predictive value for all the radiologists 10.5% (95% CI 7.7–12.4). Although the proportions regarded as abnormal ranged from 6% to 16%, and the PPV values ranged from 0% to 19%, neither of these variations is statistically significant. As the numbers are limited, the data are not presented in detail.

Inter-rater agreement

The level of agreement between the two radiologists A and B over the first 18 months of the programme was assessed by several measures, as shown in Table 3 and Figure 1. On all measures, the level of agreement was high in the first time period, much lower in the second time period, but the trend thereafter depended upon the measure of agreement used. Positive agreement, and kappa, follow very similar patterns and show a tendency to increase from time-period 3. The statistics PABAK and alpha show a very slight trend for reduction from time-period 3 through the rest of the study.

A further and rigorous method of assessment was made by fitting a log linear model to these data, using factors for each radiologist, for time and for agreement based on the diagonal cells. A model which excluded an interaction effect between time and the effect for agreement provided a good fit for the data, χ² (8 df) = 7.53, and addition of an interaction term did not improve...
this significantly, showing that the level of agreement between the radiologists did not change significantly in the course of the study. The model was also fitted, after excluding the data for the first time period. In this case, too, a model which excluded an interaction between time and agreement proved to be a good fit, $\chi^2 (7 \text{ df}) = 3.59$.

To illustrate further the use of these several measures of agreement, their values for each of the 10 radiologists are shown in Figure 2, although for some radiologists the numbers of sets of films read are fairly small. Again, positive agreement and kappa show very similar variations, while alpha shows some differences; for example radiologists E and H show similar levels of kappa but rather different values of alpha.

A random effects analysis of variance was used to compute the intraclass correlation which provides an overall measure of agreement among all the raters; it was 0.64.

**DISCUSSION**

This study is in the context of a population based screening programme for women aged 50–64 years, using invitations based on general practice lists and the electoral roll, with an uptake rate of 65–75%.\(^6\)\(^7\) It was the first large scale public funded programme in New Zealand. The radiologists concerned had considerable experience of diagnostic mammography at the start of the programme, but only limited experience of screening mammography on a large scale. As the programme was designed to be carefully assessed,\(^6\)\(^7\) independent double reading was introduced at the outset, and a system set up to ensure independence and systematic review of films where the initial opinions were inconsistent.

**Detection rates and PPV**

In a research setting, the performance of a screening test could be fully assessed if all subjects screened were also assessed by a definitive 'gold standard' diagnostic test. This is impossible in an operating screening programme, and the best assessment then requires the follow-up of all persons screened for a sufficient time so that missed cases will be detected by other means; the recording of these interval cancers allows the sensitivity and specificity of the test to be calculated.\(^9\) For breast cancer screening, this requires large numbers of subjects, as interval cancers are not common, and a 1 or 2 years follow-up period, and so is difficult to apply even to a whole programme. To compare the performance of different radiologist readers, or assess changes over time, is even more demanding. Thus, such assessments are usually restricted to those based on the initial mammogram result, and the follow-up of those women in whom the mammogram is normal. This allows calculation of the cancer detection rates, the referral rates, and the PPV, which can be used to compare readers. The PPV is the proportion of women with abnormal screening mammograms who are diagnosed with breast cancer after further assessment.

Most of the data relates to the two main radiologists (readers A and B). The proportion of women with an abnormal screening result who were subsequently diagnosed with breast cancer (the positive predictive value, PPV) was 12.4% where both radiologists found an abnormality; these women are likely to have had more clearly abnormal mammograms. Fifty-three cancers were diagnosed in this way. An additional 5 cases were initially regarded as abnormal by one reader; thus, had only one reader been used either 2 (3%) or 3 (5%) cases would have been missed, depending on which radiologist did
the reading. In the Finnish study 33% of those deemed abnormal by both radiologists were found to have cancer, and single reading would have led to either 4 (6%) or 7 (10%) cases being missed. In a Scottish study 10.9% of those with cancers would not have been detected had the mammograms been read by a single radiologist; thus, in general double reading seems to increase the detection rate by from 3% to 10%. A measure of consistency in terms of breast cancers detected would be the proportion of breast cancers detected in the opinion of both radiologists, compared with the number detected in the opinion of only one; which is 53/58, 91%, in this study.

In this study the positive predictive values for readers A and B, who read most of the mammograms, were 10.3% and 8.0%, similar to those that can be calculated from data reported for other studies. The PPV values for the two radiologists in a Finnish screening programme, using two-view mammography, were 11.0% and 12.4%; and those for three radiologists in a Scottish programme, using single oblique view mammography, who had much more experience with reading screening mammograms, were between 7.6% and 12.1%.

Effect of double reading on referral rate

The increase in the number of cancers detected was at the cost of an increased number of women being referred for further assessment. In our study, sets of films regarded as abnormal by only one reader were reviewed at a further session, and consensus was reached on whether referral was necessary. This greatly reduced the number of women referred; overall 72% of such women were referred after consensus, this proportion falling from 84% to 65% over the 18 months of the programme. As the proportion regarded as abnormal by at least one reader did not change with time, the review process led to fewer referrals to the assessment clinic. In the Finnish study the number of referrals was reduced by 45% by reviewing the mammograms found to be abnormal by either or both radiologists.

The PPV for women in whom only one radiologist categorized the films as abnormal, followed by a consensus decision to refer, was relatively low: 1.3%, as this group of women contribute only 8.6% to the numbers of cancers diagnosed, but 26.8% to the number of women referred. Further examination of the radiological features in these women might allow fewer referrals to be made without reducing the detection rate, and such studies are being done.

Single reading increased the number of referrals to the assessment clinic in both the Finnish and Scottish studies and in ours would depend upon the reader. Good agreement between the readers may make single reading a better option so whether one or two readers should be used will depend on the skill of the readers as well as a consideration of the costs and degree of negative effect of the extra recalls, and also on a consideration of the false negative rate, that is the cancers missed. These can only be ascertained after at least 1 year's follow-up of all women screened negative.

Measures of agreement

The positive predictive value for each radiologist gives an indication of the sensitivity of each radiologist, but is of limited value as it depends on the numbers of cancers detected, which will be small. Thus, in this study PPV can be assessed for the two radiologists who read several thousand films, but the results for other radiologists are too imprecise to be useful. The other way of comparing performance is by assessing the agreement, or consistency, between pairs of readers, using all the films they review. However, the appropriate choice of agreement measure is not obvious, and in this paper we have used several, and shown that one analysis: the trend in agreement with time, the result depends on which measure of agreement is used. One intuitive measure of agreement is the number of women for whom both radiologists recognized an abnormality, as a proportion of the number of women in whom either radiologist recognized an abnormality (the positive agreement). This is a measure of the consistency of making a referral decision, and in this study, comparing readers A and B, is 0.69. This takes no account of the likelihood of agreement by chance.

Kappa is a commonly used measure of the level of agreement in excess of that found by chance, a value of 1 signifying perfect agreement and 0 no agreement. It is possible to consider its statistical significance but it is generally argued that it is its magnitude that is important. Arbitrary bench marks for kappa applied to clinical; measurements in general have been provided; 0.41–0.60 is considered to be moderate agreement, 0.61–0.80 substantial agreement and 0.81–1.00 almost perfect agreement. Overall kappa was 0.65 for radiologists A and B and was between 0.48 and 0.74 when the other radiologists were compared with radiologist A. These results give a 'substantial' level of agreement. The intraclass correlation which provides an overall measure of agreement among all the raters was 0.64; it can be interpreted in the same way as kappa, suggesting that the agreement among all the raters was acceptable.

Results compared with other studies

The evaluation of agreement of independent assessment of mammography by two radiologists has been done in very few studies. One Canadian study, carried out in the planning stage of a multicentre randomized controlled trial of screening for breast cancer with mammography, compared the interpretation by 8 radiologists of a set of 100 selected xeromammograms, which comprised 10 histologically proven cancers, 40 benign abnormalities and 50 normal films. They were classified as a technically unsatisfactory film, normal breast, benign abnormality, suspicion of cancer and cancer present. Using weighted kappa as the measure of agreement between
pairs of raters, this was between 0.23 and 0.59. Agreement was highest for a diagnosis of cancer, with mean values of kappa of 0.50 or more for all but one reader. The radiologists also record their recommendations for management, as no recommendation. Follow-up after 12 months, follow-up between 6 and 12 months, follow-up within 6 months and aspiration or biopsy, giving values of weighted kappa for agreement were between 0.15 and 0.63. However, this study was based on small numbers, in a non-routine situation. A similar study in which 10 radiologists considered 150 mammograms found only moderate consistency between pairs of radiologists, weighted kappa was 0.47. The authors commented upon widely differing recommendations for management among the radiologists.

Studies within ongoing programmes are summarized in Table 4. In the Canadian National Breast Screening Study, a single reference radiologist blindly reviewed 5200 randomly selected two-view mammograms from women not found to have breast cancer, 575 screening detected breast cancers cases and 102 interval cancer cases. Agreement, measured by kappa, between the reference radiologist and the centre radiologists for those known to have breast cancer was 0.51; for those not known to have breast cancer it was 0.31. The more recent Finnish recent study provides figures from which it is possible to calculate kappa, which yields 0.31. Thus, the current results shown a significantly higher level of agreement than in these other two studies. Other series are available in which double reading has been used, but the data are not presented in a way in which a measure of agreement can be calculated.

Factors which would reduce the level of agreement include a low prevalence of abnormalities, poor technical quality or radiographic technique, particularly difficult to interpret films for biological reasons, lack of radiologists' experience, and chance.

Methodology

The value of the kappa statistic depends on the prevalence of the disease as well as the sensitivity and specificity of the test, approaching 0 as the prevalence of the disease approaches 0 or 1. A high level of observed agreement can be reduced when the prevalence of a condition is especially low, which is the situation in many screening procedures. This makes comparisons among studies or over different periods of time within a study difficult because it is impossible to distinguish the effects of changes in prevalence from those of changes in the pattern of agreement between the raters. When there is bias because the two observers differ in their assessment of the prevalence of a condition, the value of kappa is inflated. Because of these drawbacks and because most other measures of agreement can be shown to be mathematically equivalent to kappa, it has been recommended that for ongoing studies measures of both positive and negative agreement as well as kappa be reported. The results of this study suggest that reporting all these measures of agreement is not particularly useful as kappa corresponded closely with the measure of positive agreement and correcting the measure of positive agreement for chance would convert it to kappa. Elsewhere it has been suggested that kappa should be adjusted for bias and prevalence. The prevalence adjusted, bias adjusted kappa (PABAK) is linearly related to the observed overall agreement depending upon both the positive and negative results. Kappa is more strongly related to the agreement on positive results, that is, on who should be referred. It could be argued that this is the more important issue in a screening programme.

Kappa is adjusted for chance by assuming that every observation is classified independently. The statistic alpha is developed from a simple model which assumes that mammograms arise from two hypothetical groups: one, where abnormalities or the appearance of normality are clear, which (it is assumed) are classified totally and clearly by both observers, and a second group which are difficult to assess for which (it is assumed) the agreement between two readers is only due to chance. Alpha provides an estimate of the proportion of observations which are consistently classified. The overall value of alpha for radiologists A and B in this study was 0.82. Within increasing radiologists' proficiency alpha might be expected to increase. Although alpha depends upon the prevalence the condition of interest, it has been shown to be less affected by prevalence rates lower than 10% than is kappa. In this study the level of concordance was consistent with each radiologist finding about 10% of mammograms difficult to classify without much overlap for those they find difficult. The level of agreement in the current study as assessed by alpha was higher than the other available studies (Table 4).

<table>
<thead>
<tr>
<th></th>
<th>Kappa</th>
<th>95% CI</th>
<th>Alpha</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canadian Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No known cancer</td>
<td>5200</td>
<td>0.31</td>
<td>0.28</td>
<td>0.34</td>
</tr>
<tr>
<td>Screen detected cancers</td>
<td>575</td>
<td>0.51</td>
<td>0.42</td>
<td>0.60</td>
</tr>
<tr>
<td>Interval cancers</td>
<td>102</td>
<td>0.16</td>
<td>0.04</td>
<td>0.29</td>
</tr>
<tr>
<td>Finish Study</td>
<td>15457</td>
<td>0.31</td>
<td>0.27</td>
<td>0.35</td>
</tr>
<tr>
<td>Otago Southland</td>
<td>6800</td>
<td>0.66</td>
<td>0.63</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Table 4 | Kappa and alpha for the comparisons between the reference radiologist and the others for the data from the Canadian National Breast Screening Study and the Otago-Southland Pilot Programme.
CONCLUSION

The extent of agreement between the main radiologists was higher in this study than in other major studies which have been published. This result, along with the high cancer detection rate in the programme suggests that the standard of radiological assessment in the programme was high. We believe that the procedure of independent double reading, followed by a consensus decision for women whose films were initially regarded as abnormal by only one reader, is a useful approach. For the two main readers, the use of double reading increased the number of cancers detected by screening by either 2 or 3 (depending upon the radiologist) of the total of 58; that is by an average of 4%. This figure is somewhat lower than those shown in other studies, which is related to the higher degree of inter-observer consistency in our study. An interesting result is that the double reading process did not necessarily increase the number of women referred for assessment. Had the initial radiological decision from one radiologist been used, without the opportunity for discussion, the referral rate would have ranged from 9.4% to 12.3%, depending upon the radiologist. All these referral rates are relatively high, which is probably because this represents a first screening in a previously unscreened population, and the overall detection rate of cancer at this first screen has been high by international standards, around 10 per 100 women screened. The incidence of breast cancer is relatively high in New Zealand, and it is possible that the true incidence is higher than in many other areas where screening programmes have operated.

The most conservative approach would be to refer only women in whom both radiologists found an abnormality, and this would reduce the number of women referred by 26%. Correspondingly, the frequency of cancer in women referred after only one radiologist found an abnormality was quite low (1.3%), compared with the frequency where both radiologists found an abnormality (12.4%). However, we do not believe it is likely to be clinically acceptable only to refer women in whom both radiologists have found an abnormality, as there would be ethical and legal issues if no further investigation were undertaken on the others.

There are substantial benefits of double reading. It is an important educational experience, and encourages radiologists to take a realistic view of their abilities, and to continually improve. This is particularly helpful when a new radiologist joins the team or as in this case when the whole screening programme is new.

Acknowledgements

This report relates to the Otago-Southland Breast Screening Programme and acknowledges its dependence on all staff of the programme. In this case we are particularly grateful to Dr S. Crawford and G. Morris of Maribynong Clinic, Dunedin and Dr J. D. Billings, M. D. Brew, R. J. Chisholm, R. H. Coats, A. Long, T. M. M. Maiting, D. G. Ross, G. D. Watson and G. J. Welch of the Christchurch Radiology Group. We should also like to thank Mr R. Phipps and Mr M. Pfeifer (Surgery), Dr A. Miller, Dr R. Berkeley and Dr A. Pentegrew (Pathology), Mrs E. Bang (Manager), Mrs B. Morgan (Radiography) and Dr J. Nicol (Medical Physicist). In addition we thank the many other staff involved, the women and their general practitioners. The programme is financially supported by the New Zealand Department of Health, and the evaluation by the Cancer Society of New Zealand and the Health Research Council.

References


Date submitted 23 September 1994
Date accepted 7 February 1995
IN PRACTICE

The costs of mammography screening in New Zealand: evidence from the pilot programmes

Arun Menon, MB, BS, DPH, Department of Preventive and Social Medicine; Nancy J Devlin, BA, PhD, Economics Department; Ann K Richardson, MCCMNZ, Department of Preventive and Social Medicine, University of Otago, Dunedin.

Abstract

Aims. To measure the public health service costs associated with New Zealand's pilot mammography screening programmes. To compare the early evidence on cost per woman screened and per cancer detected in those programmes to that of overseas screening programmes. To estimate the cost of introducing a national screening programme in New Zealand.

Methods. Costs in each screening centre were obtained by a careful examination of screening budgets and public health service accounts; these were inflation adjusted using a consumers price index, and analysed in terms of equivalent annual operating costs.

Results. In the first year of screening the cost per woman screened (in $1991) was $182 in Waikato and $178 in Otago/Southland. The cost per woman screened in the third year of screening (with an assumed full screening throughput of 8000 women per annum) is estimated to fall to $106 and $113 for the Waikato and Otago/Southland programmes respectively. The cost per cancer detected in the first screening round differs between the two programmes. In the first year of screening the cost per cancer detected was $35 975 in Waikato and $21 908 in Otago/Southland. The difference was primarily attributable to a lower cancer detection rate in Waikato in that period (0.51% of women screened compared with 0.81% in Otago/Southland).

Conclusions. The initial performance of the New Zealand pilot programmes, both in terms of cost per woman screened and cost per cancer detected, falls within the range indicated from overseas experience. An established national screening programme is estimated to add between $9.3 and $9.9 million dollars (in 1991 dollar terms) to health service costs each year. These costs will be partly offset by savings resulting from the earlier detection of cancers.

Pilot mammography screening programmes were established in two regions of New Zealand, Otago/Southland (in November 1990) and Waikato (in February 1991), with the objective of determining the costs and benefits of screening for New Zealand. In each region, women between the ages of 50 and 65 years were invited to undergo screening at two-yearly intervals. Screening in the Otago/Southland programme was provided from two fixed units (one each in Invercargill and Dunedin) and one mobile unit, and in Waikato from one fixed and one mobile unit.

The aims of this paper are: (1) To measure the public health service costs associated with the recruitment, screening and subsequent clinical investigation of women participating in the pilot programmes. (2) To gauge the initial performance of those programmes, on indicators such as cost per woman screened and per cancer detected, relative to that of overseas screening programmes. (3) To estimate the cost of introducing a national screening programme.

**Methods**

Data was sought from each screening programme regarding the costs associated with the recruitment, screening, and follow-up investigation of participants, including diagnostic procedures such as clinical examination, further mammography, ultrasound, fine needle aspiration and surgical biopsy. Treatment costs relating to cancers detected by the programme have not been included. All costs exclude GST.

Cost data were primarily derived from the financial records kept by each programme for the years ending June 1991-3. No screening was undertaken in 1991 (year 0); 1992 and 1993 represent years 1 and 2 of screening activity. Costs relating to the use of wider hospital resources (such as those generated by follow-up assessments and overheads) were based mainly upon resource utilisation system (RUS) data, rather than on charges made against the screening programmes. In many cases such charges were deferred or negotiated at a rate below the actual cost. Some of the costs used by the screening programmes were funded by private donations. In order to reflect the costs of extending screening beyond the pilots, all direct capital and operating costs have been included, regardless of how these were funded.

Although significant overseas travel costs were incurred in establishing the pilot programmes, the results reported here include only the salary component of training costs on the basis that if screening were introduced on a national basis, clinicians and managers could be trained from the pilot centres rather than from overseas.

Actual costs in year 2 have been used to project costs for later periods. A screening throughput of 8 000 women per annum is assumed to be achievable by each programme by year 3. This is based upon an eligible population of approximately 20 000 women for each programme, screened every two years, with a coverage rate of 80%. Operating costs in year 3 incorporate the estimated additional inputs required by each programme to achieve this screening throughput.

Costs were calculated by a consumers price index (CPI) with a base in June 1991, and are reported in terms of annual operating costs to allow the performance of the programme in each period to be compared. Capital costs were annuitised to calculate the "equivalent annual cost", which incorporates both depreciation and opportunity costs, considerations regarding capital inputs. Any interest rate of 5% and a life of 10 years for all capital equipment is used. To check the sensitivity of results to this variable, and allowing for the possibility of technological change limiting the clinical life of equipment, results are also reported for capital having a life of 5 years.

Evidence on the costs of overseas screening programmes were, in each case, inflated to 1991 prices using the CPI for the relevant country, and converted to New Zealand dollars using 1991 mid-year exchange rates. Given that the prices of screening inputs may have increased more or less quickly than consumers prices in each country, and the variations in exchange rates which affect the conversion of other currencies, the resulting figures should be taken as an approximate guide only to the New Zealand dollar equivalents of the original results.

**Results**

Table 1 shows the costs (in 1991 dollar terms) of the Otago/Southland screening programme, expressed annually over a four year period.

<table>
<thead>
<tr>
<th>Year</th>
<th>Annual equivalent</th>
<th>Capital costs</th>
<th>Operating costs</th>
<th>Total cost</th>
<th>No. women screened</th>
<th>No. cancers detected</th>
<th>Cost per woman screened</th>
<th>Cost per cancer detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>55 953</td>
<td>40 676</td>
<td>174 264</td>
<td>230 249</td>
<td>2 701</td>
<td>22</td>
<td>$178</td>
<td>$113</td>
</tr>
<tr>
<td>1</td>
<td>55 953</td>
<td>40 676</td>
<td>174 264</td>
<td>230 249</td>
<td>2 701</td>
<td>22</td>
<td>$178</td>
<td>$113</td>
</tr>
<tr>
<td>2</td>
<td>55 953</td>
<td>40 676</td>
<td>174 264</td>
<td>230 249</td>
<td>2 701</td>
<td>22</td>
<td>$178</td>
<td>$113</td>
</tr>
<tr>
<td>3</td>
<td>55 953</td>
<td>40 676</td>
<td>174 264</td>
<td>230 249</td>
<td>2 701</td>
<td>22</td>
<td>$178</td>
<td>$113</td>
</tr>
</tbody>
</table>

**Table 1.** Otago/Southland: total annual costs of screening, cost per woman screened, and cost per cancer detected, in 1991 dollar terms.

The cost per woman screened in the first year of screening is $178, which decreases to $117 in year 2 and $113 in each subsequent year, assuming an annual screening throughput of 8 000 women from year 3 onwards. This result is sensitive to assumptions regarding the useful life of the programme's capital assets. Changing the assumed life of capital assets from 10 to 5 years increases the cost per programme screened to $206 in the first year of screening $130 in the second year and $123 in later periods. The cost per cancer detected is initially $21 908, which falls to $11 544 in year 2.

**Table 2.** Waikato: total annual costs of screening, cost per woman screened, and cost per cancer detected, in 1991 dollar terms.

The Waikato programme differs from the Otago/Southland programme in a number of respects. The Waikato programme was the first to be initiated, and hence initial performance should be more easily assessed than that of the Otago/Southland programme.

The Waikato programme was established in June 1991, with the objective of determining the costs and benefits of screening for New Zealand. In each region, women between the ages of 50 and 65 years were invited to undergo screening at two-yearly intervals. Screening in the Otago/Southland programme was provided from two fixed units (one each in Invercargill and Dunedin) and one mobile unit, and in Waikato from one fixed and one mobile unit.

The aims of this paper are: (1) To measure the public health service costs associated with the recruitment, screening and subsequent clinical investigation of women participating in the pilot programmes. (2) To gauge the initial performance of those programmes, on indicators such as cost per woman screened and per cancer detected, relative to that of overseas screening programmes. (3) To estimate the cost of introducing a national screening programme.

**Methods**

Data was sought from each screening programme regarding the costs associated with the recruitment, screening, and follow-up investigation of participants, including diagnostic procedures such as clinical examination, further mammography, ultrasound, fine needle aspiration and surgical biopsy. Treatment costs relating to cancers detected by the programme have not been included. All costs exclude GST.

Cost data were primarily derived from the financial records kept by each programme for the years ending June 1991-3. No screening was undertaken in 1991 (year 0); 1992 and 1993 represent years 1 and 2 of screening activity. Costs relating to the use of wider hospital resources (such as those generated by follow-up assessments and overheads) were based mainly upon resource utilisation system (RUS) data, rather than on charges made against the screening programmes. In many cases such charges were deferred or negotiated at a rate below the actual cost. Some of the costs used by the screening programmes were funded by private donations. In order to reflect the costs of extending screening beyond the pilots, all direct capital and operating costs have been included, regardless of how these were funded.

Although significant overseas travel costs were incurred in establishing the pilot programmes, the results reported here include only the salary component of training costs on the basis that if screening were introduced on a national basis, clinicians and managers could be trained from the pilot centres rather than from overseas.

Actual costs in year 2 have been used to project costs for later periods. A screening throughput of 8 000 women per annum is assumed to be achievable by each programme by year 3. This is based upon an eligible population of approximately 20 000 women for each programme, screened every two years, with a coverage rate of 80%. Operating costs in year 3 incorporate the estimated additional inputs required by each programme to achieve this screening throughput.

Costs were calculated by a consumers price index (CPI) with a base in June 1991, and are reported in terms of annual operating costs to allow the performance of the programme in each period to be compared. Capital costs were annuitised to calculate the "equivalent annual cost", which incorporates both depreciation and opportunity costs, considerations regarding capital inputs. Any interest rate of 5% and a life of 10 years for all capital equipment is used. To check the sensitivity of results to this variable, and allowing for the possibility of technological change limiting the clinical life of equipment, results are also reported for capital having a life of 5 years.

Evidence on the costs of overseas screening programmes were, in each case, inflated to 1991 prices using the CPI for the relevant country, and converted to New Zealand dollars using 1991 mid-year exchange rates. Given that the prices of screening inputs may have increased more or less quickly than consumers prices in each country, and the variations in exchange rates which affect the conversion of other currencies, the resulting figures should be taken as an approximate guide only to the New Zealand dollar equivalents of the original results.

**Results**

Table 1 shows the costs (in 1991 dollar terms) of the Otago/Southland screening programme, expressed annually over a four year period.

<table>
<thead>
<tr>
<th>Year</th>
<th>Annual equivalent</th>
<th>Capital costs</th>
<th>Operating costs</th>
<th>Total cost</th>
<th>No. women screened</th>
<th>No. cancers detected</th>
<th>Cost per woman screened</th>
<th>Cost per cancer detected</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>55 953</td>
<td>40 676</td>
<td>174 264</td>
<td>230 249</td>
<td>2 701</td>
<td>22</td>
<td>$178</td>
<td>$113</td>
</tr>
<tr>
<td>1</td>
<td>55 953</td>
<td>40 676</td>
<td>174 264</td>
<td>230 249</td>
<td>2 701</td>
<td>22</td>
<td>$178</td>
<td>$113</td>
</tr>
<tr>
<td>2</td>
<td>55 953</td>
<td>40 676</td>
<td>174 264</td>
<td>230 249</td>
<td>2 701</td>
<td>22</td>
<td>$178</td>
<td>$113</td>
</tr>
<tr>
<td>3</td>
<td>55 953</td>
<td>40 676</td>
<td>174 264</td>
<td>230 249</td>
<td>2 701</td>
<td>22</td>
<td>$178</td>
<td>$113</td>
</tr>
</tbody>
</table>

**Table 2.** Waikato: total annual costs of screening, cost per woman screened, and cost per cancer detected, in 1991 dollar terms.

The Waikato programme differs from the Otago/Southland programme with regards to its policy on women
who are invited, but not eligible, for a screen (due to, for
example, having "aged out" of the eligible age group). The
Otago-Southland programme screened only those eligible
at the date of screen, while the Waikato programme screened
all women who accepted invitations. The costs per woman
screened shown in Table 2 relate to all (eligible and ineligible)
women screened in the Waikato programme.

The results shown in Tables 1 and 2 can be used to
evaluate the costs of the national screening programme.
There are approximately 220,000 New Zealand women
between the ages of 50 and 65 years. With a
screening interval of two years, 110,000 women would be
eligible for screening each year. On the basis of the cost per
woman screened in the pilots in year 3, and assuming a
coverage rate of 80% of eligible women, a national
screening programme would add between $9.3 and $9.9
million to health service costs in each year (in 1991 dollars).
This estimate is based upon the annual cost per woman
screened in an established programme with a full screening
throughput, where capital costs have been spread evenly
over each year of operation. In year 2 the amortised capital
cost accounts for 11% of the cost per woman screened in
Waikato and 14% in Otago-Southland. The actual costs of
introducing a national screening programme would be
higher in early years when capital costs are incurred (and
operational costs per woman screened are higher) and lower
in later years (when capital costs have been sunk).

Discussion

The costs per woman screened which are evident from the
New Zealand pilot programmes are comparable to those
reported from overseas programmes.

For example, the cost per woman screened in the first
year of screening of $182 in Waikato and $178 in Otago
compare with approximately NZ$175 for the Sydney Breast
X-ray Programme. The Waikato and Otago-Southland cost
per woman screened in year 2 fell to $108 and $117
respectively. This compares favourably with the Sydney
programme in which costs for subsequent years remain
at around the year 1 level (personal communication, J Hall
29 September 1993).

These costs are considerably higher than the estimated
cost per woman screened of NZ$51 used by the Forrest
report, but fall within the range of screening costs evident
from other international programmes. The cost per woman
screened in the Netherlands varies from NZ$162 to NZ
$72. Similarly the New York Health Insurance plan
reports a cost of NZ$396 per screen. All costs are in
equivalent 1991 New Zealand dollars and are adjusted for
inflation in the countries concerned. However, while these
studies allow us to make some broad comparisons with the
results reported for New Zealand, comparisons need to be
approached with care. The methodology employed in each
case is different (with regards to what costs are considered)
as is the characteristics of the screening programmes being
evaluated (with respect to the eligible age group, the
screening interval and method of screening, as well as the
prevailing prices of screening inputs).

Diagnostic mammography screening costs (including the
fee charged plus government subsidies paid) currently
appear to fall in the range of $71.80 (Christchurch
Southern Cross) to $100 (Auckland) per woman screened.
These costs exclude two important elements of the activity
associated with a mass screening programme. First, the
recruitment of women to the programme, and second, the
costs associated with the follow on assessment of suspicious
diagnoses from screening. Although we were unable to
separately identify recruitment costs for the Otago-
Southland programme, Waikato data indicates that
approximately one tenth of the cost per woman screened
was attributable to recruitment and promotion activities.
A further one tenth of the cost per woman screened was
attributable to the follow on assessment of screened women.
Excluding these components, the cost of the basic screen
evident from Waikato data is comparable to private sector
costs at $88 in year 2. The option of contracting out the
provision of the basic screen under the auspices of an
extended screening programme is one which may merit
further consideration. Private sector provision of screening
may, however, generate other costs such as those arising
from the planning and coordination of recruitment and
screening activities which are not included in this
comparison.

While this paper presents some evidence on the costs and
intermediate outputs of the pilot programmes, these need
to be interpreted with caution.

First, it should be noted that the calculation of cost per
woman screened in years 1-3 to some extent disguises the
fact that significant set-up costs were incurred in year 0
before screening started. In both programmes there
were lengthy delays between the start of the programme and
the commencement of screening (9 months in Waikato and 10
months in Otago-Southland). The skills and experience
gained by managers and clinicians through the pilots could
hopefully be used to shorten the learning curve in
subsequent screening programmes.

Second, the number of cancers detected in the first
screening round is likely to be atypical of the longer term
detection rate. In the first screening round most of the
participating women will have been screened for the first time
and thus prevalent cancers will be detected. In later screening
rounds most women (with the exception of those aging into
the eligible age category) will have been screened earlier so
that only new (or incident) cancers will be detected. As
a result, the cost per cancer detected in subsequent periods.
all other things being equal can be expected to rise.

Finally, the costs reported here represent only the
additions to cost arising from the recruitment, screening
and assessment of screening participants. While the
introduction of screening adds to public health service
costs, it also has the potential to confer some resource
savings. These arise because screening effects the stage
(extent of advancement) at which cancers are detected,
relative to no screening conventional management. Earlier
detection of cancers means that the treatment of these
cancers may incur less costs (as well as producing more
years of life). For example, the resource savings generated
by breast cancer screening of the Dutch population, in
terms of lower assessment and treatment costs, may be
significant in the long term, offsetting up to one third of the
costs of screening itself. The decision as to whether a
national programme of mammography screening should be
implemented in New Zealand should be based upon a
careful examination of the net costs (including the
opportunity costs of New Zealand women foregone from other
health services), risks and benefits (both quantitative and
qualitative) evident from the pilot programmes, and the
acceptability to women of screening.

Acknowledgements. We are grateful to the managers of the pilot screening programmes.
Elizabeth Beale (Otago-Southland) and Helen McKinnon (Waikato) for their efforts in
providing information and their support. Thanks also to Sue Cashman, economist
Department of University of Otago, for assistance in collating price indices and exchange
rates.

Correspondence: Dr Anna Menon, Department of Preventive and Social Medicine, University
of Otago Medical School, PO Box 912, Dunedin

1. Richardson AW, Cafu A. Treatment of capital costs in evaluating health care

2. Drummond MF, Stoddart GL, Torrance GV. Methods for the economic evaluation of


4. Forrest P. Breast cancer screening: report to the Health Ministers of England, Wales,

5. Van der Maas PJ, De Koning NJ, Inselv M. The cost effectiveness of breast cancer

screening in Australia. Paper presented to the Second World Conference on Health

cost effectiveness: policy alternatives, quality of life considerations and the

1031-7.
EVALUATION DATA BASE:

(Information downloaded from pilot programmes’ computer systems. Missing data or information that does not apply to one of the programmes must be included in the form of blank fields so that files from the two programmes are compatible.)

Flag P
6 blanks (so we can add evaluation number)
Patient number (NMPI) 7 alphanumeric
Region: (1=Otago and Southland, 2=Waikato)
Area code (ward) 5 numeric
Date of birth (day, month, year) 6 numeric
Ethnic group
Otago: 1=Maori, 2=European, 3= Pacific Island Polynesian, 4=Asian, 5= Other
Waikato:
New Zealand European (1=Yes, 2= No)
New Zealand Maori (1=Yes, 2= No)
Samoan (1=Yes, 2= No)
Cook Island Maori (1=Yes, 2= No)
Tongan (1=Yes, 2= No)
Niuean (1=Yes, 2= No)
Other (1=Yes, 2= No)
Identification source:
General electoral roll (1=Yes, 2= No)
Maori electoral roll (1=Yes, 2= No)
GP list (1=Yes, 2= No)
self-referred (1=Yes, 2= No)
GP referral (1=Yes, 2= No)
BC investigation (1=Yes, 2= No)
BC treatment (1=Yes, 2= No)
Maori coordinator (1=Yes, 2= No)
other (1=Yes, 2= No)
Eligibility (1=eligible, 2=ineligible because outside eligible age-range, 3=ineligible because being followed up for breast cancer, 4=ineligible because deceased, 5=ineligible because of other reason)

Screen number 1 (Flag 1)
6 blanks (so we can add evaluation number)
Patient number (NMPI) 7 alphanumeric
Invitation sent? (1 numeric)
1= Yes - from programme
2= Yes (RCT) - programme Otago-Southland only
3= Yes (RCT) - GP
4= Yes - after specific request from woman
5= Yes - after specific request from GP
6= No - screened early (without invitation)
7= No - ineligible
8= No - other (does not apply to Waikato)
Reason for screening (1=response to invitation, 2=self-referred without invitation)
Date invitation produced (day, month, year) 6 numeric
Screening centre invited to (1=Dunedin, 2=Invercargill, 3=Otago/Southland mobile, 4=Hamilton, 5=Waikato mobile)
Appointment time (24h time, day, month, year) 10 numeric
(1st 4 fields blank for Waikato)
Response (1=accept, 2=accept for different appointment time, 3=decline, 4=no response)
Reason for declining (1=recent mammogram, 2=private screening, 3=breast disease)
Follow-up (1=not applicable, 2=second mailing, 3=telephone, 4=personal follow up)
Date of follow-up invitation (day, month, year) 6 numeric
Appointment time (24h time, day, month, year) 10 numeric
Response (1=accept, 2=accept for different appointment time, 3=decline, 4=no response)
Attendance:
centre (1=Dunedin, 2=Invercargill, 3=Otago/Southland mobile, 4=Hamilton, 5=Waikato mobile)
Location of mobile (15 alphabetic)
date and time (24h time, day, month, year) 10 numeric
mammogram taken (1=Yes, 2=No)
total number of films taken (2 numeric)
Extra films (1=Yes, 2=No)
Date of extra films (day, month, year) 6 numeric
Reason for extra films:
Technical (number of extra films taken for technical reasons) 2 numeric
Other (number of extra films taken for other reasons eg: large breasts, breast implants) 2 numeric
Result (2=normal, 3=abnormal)
Date result produced (day, month, year) 6 numeric
Result sent to (1=woman, 2=woman + GP, 3=woman + GP + assessment clinic) 3 does not apply to Waikato
Reason for referral (1=NA, 2=abnormal result, 3=symptom questionnaire, 4=abnormal result and symptom questionnaire, 5=technical, 6=abnormal and technical, 7=symptoms and technical, 8=abnormal and symptoms and technical) 3,4,6,7, and 8 Otago-Southland only
Assessment clinic attendance:
clinic (1=Dunedin, 2=Invercargill, 4=Hamilton)
date of appointment (day, month, year) 6 numeric
date of attendance (day, month, year) 6 numeric
Clinician number (3 numeric, 2 alpha)
Investigation:
Clinical Examination (1=NA, 2=normal, 3=abnormal)
assessment mammography (1=NA, 2=normal, 3=abnormal)
FNA (1=NA, 2=normal, 3=abnormal)
ultrasound (1=NA, 2=normal, 3=abnormal)
biopsy (1=NA, 2=normal, 3=abnormal)
Result from assessment clinic: (1=normal, 2=biopsy required)
Date of biopsy (day, month, year) 6 numeric
Biopsy result (1=breast cancer, 2=not breast cancer)
Final result (1=breast cancer, 2=no breast cancer)
Screen number 2

Flag 2
6 blanks (so we can add evaluation number)
Patient number (NMPI) 7 alphanumeric
Invitation sent? (1 numeric)
1= Yes - from programme
2= Yes (RCT) - programme Otago-Southland only
3= Yes (RCT) - GP
4= Yes - after specific request from woman
5= Yes - early invitation after specific request from woman
6= Yes - after specific request from GP
7= No - screened early (without invitation)
8= No - ineligible
9= No - other (does not apply to Waikato)
Date invitation produced (day, month, year) 6 numeric
Appointment time (24h time, day, month, year) 10 numeric
(1st 4 fields blank for Waikato)
Response (1=accept, 2=accept for different appointment time,
3=decline, 4=no response)
Follow-up (1=not applicable, 2=second mailing,
3=telephone, 4=personal follow up)
Date of follow-up invitation (day, month, year) 6 numeric
Appointment time (24h time, day, month, year) 10 numeric
Response (1=accept, 2=accept for different appointment time, 3=decline, 4=no response)
Attendance:
centre (1=Dunedin, 2=Invercargill, 3=Otago/Southland mobile,
4=Hamilton, 5=Waikato mobile)
Location of mobile (15 alphabetic)
date and time (24h time, day, month, year) 10 numeric
mammogram taken (1=Yes, 2=No)
total number of films taken (2 numeric)
Extra films (1=Yes, 2=No)
Date of extra films (day, month, year) 6 numeric
Reason for extra films:
Technical (number of extra films taken for technical reasons) 2 numeric
Other (number of extra films taken for other reasons eg: large breasts, breast implants) 2 numeric
Result (2=normal, 3=abnormal)
Date result sent (day, month, year) 6 numeric
Result sent to (1=woman, 2=woman + GP, 3=woman + GP +
assessment clinic) 3 does not apply to Waikato
Reason for referral (1=NA, 2=abnormal result, 3=symptom
questionnaire, 4=abnormal result and symptom questionnaire,
5=technical, 6=abnormal and technical, 7=symptoms and
technical, 8=abnormal and symptoms and technical)
3,4,6,7 and 8 Otago-Southland only
Assessment clinic attendance:
clinic (1=Dunedin, 2=Invercargill, 4=Hamilton)
date of appointment (day, month, year) 6 numeric
date of attendance (day, month, year) 6 numeric
Clinician number (3 numeric, 2 alpha)
Investigation:
Clinical Examination (1=NA, 2=normal, 3=abnormal)
assessment mammography (1=NA, 2=normal, 3=abnormal)
PNA (1=NA, 2=normal, 3=abnormal)
ultrasound (1=NA, 2=normal, 3=abnormal)
biopsy (1=NA, 2=normal, 3=abnormal)
Result from assessment clinic: (1=normal, 2=biopsy required)
Date of biopsy (day, month, year) 6 numeric
Biopsy result (1=breast cancer, 2=not breast cancer)
Final result (1=breast cancer, 2=no breast cancer)
APPENDIX FOUR
BREAST CANCER SCREENING QUESTIONNAIRE:

Name: ________________________________
Practice: ________________________________

To answer each question, please circle the appropriate answer:

1. Did you know before you received our letter that a pilot programme is being set up this year to offer free mammographic screening to eligible women in Otago and Southland?
   (1) Yes
   (2) No

2. What is your present policy for recommending mammographic screening to women in your practice? (Please indicate which age-groups your policy includes)

   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................

3. Would you encourage eligible women in your practice to take part in the pilot programme?
   (1) Yes
   (2) No
   If no, why not? .........................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
   ........................................................................................................
4. Is there any information that you would need before recommending screening to eligible women?

(1) Yes
(2) No

If yes, what information would you need? ....................
...........................................................................
...........................................................................
...........................................................................
...........................................................................

5. Does your practice have an age-sex register?

(1) Yes
(2) No

(If yes, please answer questions 6 to 8):

6. Can your register produce lists of patients with their addresses, by sex, for certain age-groups?

(1) Yes
(2) No

7. The screening programme has a register with the names and addresses of women aged between 50 and 64 in Otago and Southland but this does not include information about each woman's general practitioner. During the next two years these women will be sent invitations to be screened. We would like the invitations to be sent jointly from the screening programme and each woman's G.P. but we cannot do this unless we can match the women on our register with their G.P.s
Would you be willing to produce a list of the names and addresses of women in your practice who are 50-64, to be matched with the screening register?

(1) Yes

(2) Conditional yes
If conditional yes, under what conditions?.................
...........................................................................
...........................................................................
...........................................................................
...........................................................................
...........................................................................

(3) No
If no, why not?..................................................
...........................................................................
...........................................................................
...........................................................................
...........................................................................

8. Would you be willing to sign a letter inviting eligible women in your practice to be screened?

(1) Yes

(2) No
If no, why not?..................................................
...........................................................................
...........................................................................
...........................................................................
...........................................................................

Thank you for your help.
APPENDIX FIVE
OTAGO SOUTHLAND BREAST SCREENING PROGRAMME

CONSENT FORM

It would help both of us if you would complete the information required on this form.

Name: ...........................................................................

Surname Given Names

Previous Married Name (if remarried): ................................................. .

Maiden Name: .................................................................... .

Present Address: ..................................................................

Home Phone Number: ................... . Date of Birth: ........................... .

Work Phone Number: ................... .

Name of your family Doctor: ......................................................... .

Address of Surgery: ............................................................... .

The result of your breast X-Ray (mammogram) will be posted to you. We would like to send a copy of these to your family doctor. May we do this? (Please tick the appropriate box).

YES □ NO □

CONSENT TO PARTICIPATE IN THE PROGRAMME

1. I have read and understood the information contained in the pamphlet provided and expect to be given the opportunity to ask questions regarding the screening programme.

2. I understand that I am free to withdraw from the screening programme any time without disadvantage to my future relationship with either HealthCare Otago or Southern Health or my Doctor.

3. My records and information gained from me will be used to measure how effective this programme could be for other New Zealand women. Complete confidentiality will be kept in terms of my records, and I will not in any way be identified in any reports.

Signature: .......................................................... Date: ................... .
<table>
<thead>
<tr>
<th></th>
<th>R. Breast</th>
<th>L. Breast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Have you had a mammogram before?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>2.</td>
<td>Have you ever had a breast lump or cyst?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>3.</td>
<td>Have you ever had a biopsy, or has your doctor ever used a needle to remove fluid from your breast?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>4.</td>
<td>Have you had breast cancer treated?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>5.</td>
<td>Are you aware of a lump in your breast now?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>6.</td>
<td>Have you had a breast reconstruction or reduction?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>7.</td>
<td>Have you had any nipple discharge?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>8.</td>
<td>Are you taking hormone replacement therapy?</td>
<td>Yes / No</td>
</tr>
<tr>
<td>9.</td>
<td>Have you had painful breasts before menstruating?</td>
<td>Yes / No</td>
</tr>
</tbody>
</table>

![WART OR MOLE](image1)

![SCAR](image2)
**BREAST X-RAY PROGRAMME QUESTIONNAIRE**

This questionnaire is about your experience of having a breast x-ray. We would be very grateful if you would answer these questions honestly since it will help us to assess the Breast X-ray Programme. We assure you that the answers that you provide will remain strictly confidential.

Please tick in the box next to your answer. (You may need to tick more than one answer for some questions).

1. How did you find out about the Breast X-ray Programme? (Please tick as many as apply)
   
   (a) When I received my invitation to have a breast x-ray □
   (b) In the newspaper □
   (c) On the radio or TV □
   (d) From a friend or relative □
   (e) From a speaker at a group to which I belong □
   (f) From my family doctor □
   (g) From Iwi/Maori health co-ordinator □
   (h) Other (Please describe) □

2. Why did you decide to take part in the Breast X-ray programme? (Please tick as many as apply)
   
   (a) To detect breast cancer early □
   (b) Because it was recommended by my family doctor □
   (c) Because I have a breast symptom □
   (d) Because it is a pilot programme □
   (e) Because it is a free service □
   (f) Because I have a family history of breast cancer □
   (g) For reassurance □
   (h) Other (Please describe below) □

---

For Office Use Only

---

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23
3. If the programme did not exist, would you have arranged to have a breast x-ray anyway? (Please tick one box)
   (a) Yes [ ]
   (b) No [ ]
   (c) Yes, as long as it was a free service [ ]
   (d) Don't know [ ]
   (d) Depends (Please specify below) [ ]

4. Did you receive an invitation to have a breast x-ray?
   (a) Yes [ ]
   (b) No [ ]
   If yes, which of the following best describes your feelings when you received this invitation? (Please tick one box)
   (a) Not at all worried [ ]
   (b) A little bit worried [ ]
   (c) Quite worried [ ]
   (d) Very worried [ ]

5. What did you think about the green Breast X-ray programme information leaflet in terms of how easy it was to understand? This leaflet may have been posted to you with your invitation (Please tick one box).
   (a) Hard to understand [ ]
   (b) Just right [ ]
   (c) Too simple [ ]
   (d) Didn't see the leaflet/Don't know [ ]

6. Was there any information that you wanted to know that wasn't included in this leaflet? (Please tick one box)
   (a) Yes [ ]
   (b) No [ ]
   (c) Didn't see the leaflet [ ]
   If yes, what information should have been included?

   _______________________________________________________
   _______________________________________________________
   _______________________________________________________
   _______________________________________________________
The following questions relate to your experience of having a breast x-ray.

7. Which of the following best describes your feelings before you came in to have your breast x-ray? (Please tick one box)
   (a) Not at all worried
   (b) A little bit worried
   (c) Quite worried
   (d) Very worried

8. How long did it take you to travel from home or work, to the screening centre? (Please tick one box)
   (a) less than 15 minutes
   (b) between 15 minutes and half an hour
   (c) 30 minutes to an hour
   (d) more than an hour (please specify)

9. How did you travel to the screening centre? (Please tick one box)
   (a) walked
   (b) car
   (c) bus
   (d) taxi
   (e) other (please describe)

10. Was the waiting area large enough?
    (a) Yes
    (b) No - overcrowded

11. Was the changing room warm enough?
    (a) too hot
    (b) comfortable temperature
    (c) too cold

12. Was there enough privacy for changing?
    (a) Yes
    (b) No
13. Did you have any concerns about the way staff treated you, or anyone else who came with you, at the breast x-ray centre?
   (a) Yes □   (b) No □
   If yes, please comment
   ........................................................................................................

14. Did you feel that the way of taking the breast x-ray was explained to you so that you could understand?
   (a) Yes □   (b) No □
   If no, please comment
   ........................................................................................................

15. What time was your appointment? ...............................................  
   What time were you screened? ......................................................  

16. How would you describe your experience of having the breast x-ray?  
   (Please tick one box)
   (a) A little uncomfortable □
   (b) Very uncomfortable □
   (c) Painful □
   (d) Very painful □

17. Did you have any bruising after your x-ray? (Please tick one box)
   (a) None □
   (b) A little □
   (c) Moderate □
   (d) Severe □

18. How long did you spend at the breast x-ray centre from the time of your appointment to the end of having the breast x-ray? (Please tick one box).
   (a) Less than 15 minutes □
   (b) Between 15 minutes and half an hour □
   (c) 30 minutes to an hour □
   (d) More than an hour (Please specify below) □
19. What do you think is an acceptable time to wait for a breast x-ray? (Please tick one box).
   (a) Less than 15 minutes
   (b) Between 15 minutes and half an hour
   (c) 30 minutes to an hour

20. Did any other people come with you to the breast x-ray centre?
   (a) Yes
   (b) No

   If yes, why did they come? (Please tick any answers that apply).
   (a) For company/support
   (b) To transport me to the centre
   (c) To have a breast x-ray also
   (d) Other (Please describe below)

21. How many days after being x-rayed did you get your result? (Please fill in the space below).
   Number of days

22. Which of the following best describes how you felt while waiting for your results? (Please tick one box).
   (a) Not at all worried
   (b) A little bit worried
   (c) Quite worried
   (d) Very worried

23. Would you have another breast x-ray in two years time if you received an invitation?
   (a) Yes
   (b) No
If yes, would you prefer to be notified through the Breast X-ray Programme or make your own arrangements? (Please tick one box).

(a) Breast X-ray Programme □
(b) Arranged privately □
(c) Arranged through some other organization □
   (e.g. Maori based clinic)

24. How do you think the Breast X-ray Programme could be improved?
   (This information will help us to make the breast x-ray service as acceptable as possible).

We also need to find out about the hidden costs of this programme, so the following questions are about how much it cost you to be x-rayed.

25. Could you estimate how much it cost you to have a breast x-ray?

   Petrol, bus or taxi $.................. 58 59
   Parking fee $.................. 60 61
   Income lost because of x-ray appointment $.................. 62 63
   Any other cost (Please specify below) $.................. 64 65
26. If others came with you, could you please estimate how much it cost them?

Petrol, bus or taxi $.......................... 66 [ ] 67
Parking fee $.......................... 68 [ ] 69
Income lost because of x-ray appointment $.......................... 70 [ ] 71
Any other cost (Please specify below) $.......................... 72 [ ] 73

27. How easy or difficult was it for you to fit the breast x-ray appointment into your day? (Please tick one box).

(a) Very easy [ ]
(b) A little bit difficult [ ]
(c) Quite difficult [ ]
(d) Very difficult [ ]

28. What would you have been doing at the time if you had not come to have a breast x-ray? (Please tick one box).

(a) Working in paid employment [ ]
(b) Voluntary work [ ]
(c) Leisure activity [ ]
(d) Caring for a child or other relative [ ]
(e) Other (Please describe) [ ]

29. This screening programme is free. If the screening programme could not continue as a free service, how much would you be prepared to pay for a breast x-ray? (Please tick one box)

(a) I would not take part in the programme if there were a charge [ ]
(b) under $20 [ ]
(c) $20-$50 [ ]
(d) over $50 [ ]
Personal Details

The following questions will help us to see if there are any differences among women in terms of how they found the breast x-ray programme.

30. What is your age? (Please fill in the space below).
   Age in years ........................................

31. To which ethnic group or groups do you belong?
   (You may tick more than one answer)
   (a) European □
   (b) NZ Maori □
   (c) Pacific Island Polynesian □
   (d) Asian □
   (f) Other (Please specify below) □

32. Which of the following best describes you at the moment? (Please tick one answer)
   (a) Working for salary/wages □
   (b) Self-employed □
   (c) Not working in paid employment □
   (d) Retired □
   If you are working in paid employment, please describe your occupation, please be specific, e.g. "Secondary School Teacher" rather than "Teacher" (If retired, please describe your previous occupation).

33. What is your yearly household income before tax? (Please tick one answer)
   (a) Less than $15,000 □
   (b) $15,000 - $30,000 □
   (c) More than $30,000 □

34. What is the highest level of education you have reached? (Please tick one answer)
   (a) Primary School □
   (b) Secondary School □
   (c) University Education □
   (d) Trade or Vocational Qualification, e.g. Polytechnic □
35. At which town/city did you have your breast x-ray? (Please state)


36. Where do you live? (Please tick one box)

(a) In the Dunedin city area
(b) In the Invercargill city area
(c) In a rural town
(d) In the country


37. How would you rate your risk of getting breast cancer compared with other women your age? (Please tick one box)

(a) I think I would be more at risk
(b) I think I would be less at risk
(c) I think I would have the same risk
(d) Don't know


38. Has anyone in your family or close to you had breast cancer?

(a) Yes (b) No

If yes, please tick any of the following that apply:

(a) Close friend
(b) Mother
(c) Sister
(d) Don't know
(e) Other relative (Please specify below)


39. Did you require any help completing this questionnaire?

(a) Yes (b) No


40. Please fill in today's date


Thank you for taking time to fill out this questionnaire, which will help us to assess the Breast X-ray Programme.

Please post the completed questionnaire in the post-paid envelope provided.
ASSESSMENT CLINIC QUESTIONNAIRE

Thank you for completing this questionnaire about the assessment clinic.

1. Did you receive an invitation to have a breast x-ray?
   (a) Yes  
   (b) No  

   If yes, which of the following best describes your feelings when you received this invitation? (Please tick one of the following)
   (a) Not at all worried  
   (b) A little bit worried  
   (c) Quite worried  
   (d) Very worried  

2. How worried did you feel about being screened? (Please tick one of the following)
   (a) Not at all worried  
   (b) A little bit worried  
   (c) Quite worried  
   (d) Very worried  

3. How worried did getting the appointment for the assessment clinic make you feel? (Please tick one of the following)
   (a) Not at all worried  
   (b) A little bit worried  
   (c) Quite worried  
   (d) Very worried  

4. What did you think about the information that was given to you when you were phoned about your appointment? (Please tick one of the following)
   (a) too little information  
   (b) exactly the right amount of information  
   (c) too much information
5. How long did it take you to travel from home to the assessment clinic? (please tick one answer)
   (a) less than 15 minutes □
   (b) between 15 minutes and half an hour □
   (c) 30 minutes to an hour □
   (d) more than an hour □ (please specify) ............................................................

6. How did you travel to the assessment clinic? (please tick one answer)
   (a) walked □
   (b) car □
   (c) bus □
   (d) taxi □
   (e) other □ (please describe) .............................................................................

7. Did anyone come with you to the assessment clinic? (please tick one answer)
   (a) Yes □
   (b) No □

   If yes, why did they come?
   (i) for company □
   (ii) to transport me to the centre □
   (iii) to be screened □
   (iv) other □ (please describe) .............................................................................

8. Could you please estimate how much it cost you to go to the assessment clinic?
   Petrol, bus or taxi $.......................... □ □ 20
   Parking fee $.......................... □ □ 22
   Income lost because of x-ray appointment $.......................... □ □ 24
   Any other cost (Please specify below) $.......................... □ □ 26
9. If someone came with you could you please estimate how much it cost them?

   Petrol, bus or taxi $.........................
   Parking fee $.........................
   Income lost because of x-ray appointment $.........................
   Any other cost (Please specify below) $.........................

   ______________________________________________________

10. What would you have been doing if you hadn't had an appointment at the assessment clinic? (please tick one answer)

   (a) working in paid employment
   (b) voluntary work
   (c) leisure activity
   (d) caring for a child or other relative
   (e) other (please describe)............................

   ______________________________________________________

11. What time was your appointment? .........................................................

   What time were you seen? ..........................................................

   ______________________________________________________

12. How long did you spend at the assessment clinic altogether? (please tick one answer)

   (a) less than 15 minutes
   (b) between 15 minutes and half an hour
   (c) 30 minutes to an hour
   (d) more than an hour (please specify)

   ______________________________________________________

13. Was the waiting area large enough?

   Yes

   No - Overcrowded

   ______________________________________________________

14. Was the changing room warm enough? (please tick one answer)

   (a) too hot
   (b) comfortable temperature
   (c) too cold

   ______________________________________________________
15. Was there enough privacy for changing? (please tick one answer)
   (a) Yes □
   (b) No □

16. How worried did you feel during the assessment? (Please tick one answer)
   (a) Not at all worried □
   (b) A little bit worried □
   (c) Quite worried □
   (d) Very worried □

17. How did you find the staff at the assessment clinic? (Please tick one answer)
   (a) Unfriendly □
   (b) A little bit unfriendly □
   (c) Quite friendly □
   (d) Very friendly □

18. When did you get the result of the assessment? (Please tick one answer)
   (a) straight away (at the assessment clinic) □
   (b) within a week □
   (c) between 1 and 2 weeks after the assessment □
   (d) more than 2 weeks after the assessment □

19. How worried did you feel while waiting for the result? (Please tick one answer)
   (a) Not at all worried □
   (b) A little bit worried □
   (c) Quite worried □
   (d) Very worried □
20. How do you think the assessment clinic could be improved? (This information will help us make the clinic as acceptable as possible)

........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................
........................................................................................................................................

21. If you receive another invitation to have a breast x-ray (in 2 years time) do you think you will attend? (please tick one answer)

(a) Yes ☐
(b) No ☐

Why? ........................................................................................................................................
........................................................................................................................................
........................................................................................................................................

22. This screening programme is free. If the screening programme could not continue as a free service, how much would you be prepared to pay for a breast x-ray? (Please tick one answer)

(a) I would not take part in the programme if there were a charge ☐
(b) Under $20 ☐
(c) $20 - 50 ☐
(d) Over $50 ☐

23. What is your age? .................................................................
24. To which ethnic group or groups do you feel you belong? (please tick those which apply)

(a) European □
(b) NZ Maori □
(c) Pacific Island Polynesian □
(d) Asian □
(e) Other □ (please specify) ....................................................................................

25. Which of the following best describes your working situation? (Please tick one answer)

(a) Working for salary/wages □
(b) Self-employed □
(c) Not working in paid employment □
(d) Retired □

If you are working in paid employment please describe your occupation. (If retired please describe your previous occupation)

26. What is your yearly household income before tax? (Please tick one)

(a) Less than $15,000 □
(b) $15,000 - $30,000 □
(c) More than $30,000 □

27. What is the highest level of education you have reached? (please tick one answer)

(a) Primary school □
(b) Secondary school □
(c) University education □
(d) Trade or Vocational Qualification □

Please fill in today's date.................................................................

THANK YOU FOR YOUR HELP

PLEASE POST THIS IN THE POST-PAID ENVELOPE PROVIDED
APPENDIX EIGHT
TELEPHONE INTERVIEW FOR WOMEN WHO
DID NOT ATTEND SCREENING

ID Number ______________________

Date ______________________

Non-Respondent □

Non-Participant □

Introduction: (Information for interviewers is given in bold print)

Hello, my name is Margaret Bahr. I work in the Medical School at Otago University with the evaluation of the Otago-Southland Breast Screening Programme. It would be a great help to us if you could answer some confidential questions about the programme. Is it convenient for me to talk to you now?

1. Did you receive an invitation to have a breast x-ray?
   (i) Yes □
   (ii) No □

2. Could you please tell me how or where you heard about the screening programme? (Tick all replies) Prompt if necessary.
   (i) when I received an invitation to have a breast x-ray □
   (ii) newspaper or magazine □
   (iii) radio or TV □
   (iv) friend or relative □
   (v) speaker at a group to which you belong □
   (vi) family doctor □
   (vii) iwi authority □
   (viii) other (please describe) □
Ask questions 3-6 if the woman received an invitation. If not please go to question 7.

3. (For women who received the invitation)
   Could you please tell me what made you decide not to be screened? (Prompt)
   (i) recent mammogram
   (ii) already being followed up for a breast problem
   (iii) no transport
   (iv) too far to the screening centre
   (v) inconvenient appointment time
   (vi) negative reports from other women who had attended
   (vii) fear of the procedure
   (viii) fear of the possible outcome
   (ix) other (record below)

4. How worried did you feel when you got the appointment to be screened?
   (a) Not at all worried
   (b) A little bit worried
   (c) Quite worried
   (d) Very worried

5. Could you please tell me what you thought about the information that was posted to you with your invitation? (Prompt)
   (i) hard to understand
   (ii) just right
   (iii) too simple
   (iv) didn’t see the leaflet/don’t know
Was there any information that you wanted to know that wasn’t included in the leaflet? (please tick one answer)

(i) Yes  
(ii) No  

If yes, what information should have been included? ............................................................... ☐ 35

6. Could you please tell me which screening centre you were invited to?

(i) Dunedin hospital screening centre  
(ii) Mobile screening centre  
(iii) Kew hospital screening centre  

7. How would you have travelled to the screening centre?

(i) walked  
(ii) car  
(iii) bus  
(iv) taxi  
(v) other (please describe) ................................................................. ☐ 37

8. How long would it have taken you to travel to the screening centre?
(for women who did not receive an invitation, interviewer should specify nearest screening location)

(i) less than 15 minutes  
(ii) between 15 minutes and half an hour  
(iii) 30 minutes to an hour  
(iv) more than an hour (please specify)................................................................. ☐ 38
9. Can you please recall what you were doing at the time which was set for your appointment? (Prompt)
   (i) working in paid employment
   (ii) voluntary work
   (iii) leisure activity
   (iv) caring for a child or other relative
   (v) other (please describe) .................................................................

10. How easy would it have been for you to fit the screening appointment into your usual routine? (Prompt)
    (a) Very easy
    (b) Quite easy
    (c) A little bit difficult
    (d) Quite difficult
    (e) Very difficult

11. Could you please tell me how you would rate your risk of getting breast cancer compared with other women your age? (Prompt)
    (i) I think I would be more at risk
    (ii) I think I would be less at risk
    (iii) I think I would have the same risk
    (iv) Don’t know

12. Could you please tell me if anyone in your family or close to you has had breast cancer?
    (i) Yes
    (ii) No
    If yes, please specify:
    (i) close friend
    (ii) mother
    (iii) sister
    (iv) don’t know
    (v) other relative (please specify) ...................................................
13. If you receive an invitation to be screened (in 2 years time) will you attend?
   (i) Yes ☐
   (ii) No ☐

Why .................................................................
.................................................................
.................................................................
.................................................................
.................................................................

We now need some personal details so we can see if there is any difference between different groups of women and how they feel about the screening programme.

14. Could you please tell me your age? ........................................................

15. We need to know your ethnic group. Could you please tell me which ethnic group or groups you feel you belong to? (Prompt)
   (i) European ☐
   (ii) NZ Maori ☐
   (iii) Pacific Island Polynesian ☐
   (iv) Asian ☐
   (v) Other (please specify) .................................................................

16. Could you please tell me about your employment? (Prompt)
   Are you:
   (i) Working for salary/wages ☐
   (ii) Self-employed ☐
   (iii) Not working in paid employment ☐
   (iv) Retired ☐

If you are working in paid employment please describe your occupation. (If retired please describe your previous occupation)
.................................................................
.................................................................
.................................................................
.................................................................
17. Could you please tell me your yearly household income before tax? (Prompt)

(a) Less than $15,000 □
(b) $15,000 - $30,000 □
(c) More than $30,000 □
(d) Refused □

18. Could you please tell me the highest level of education you have reached? (Prompt)

(i) Primary school □
(ii) Secondary school □
(iii) University education □
(iv) Trade or Vocational Qualification (eg: Polytechnic) □
(v) Refused □

THANK YOU FOR YOUR HELP
TELEPHONE INTERVIEW FOR WOMEN WHO ATTENDED SCREENING

ID Number

Date

Participant

Introduction: (Information for interviewers is given in bold print)

Hello, my name is Margaret Bahr. I work in the Medical School at Otago University with the evaluation of the Otago-Southland Breast Screening Programme. It would be a great help to us if you could answer some confidential questions about the programme. Is it convenient for me to talk to you now?

1. Did you receive an invitation to have a breast x-ray?
   (i) Yes  
   (ii) No

2. Could you please tell me how or where you heard about the screening programme? (Tick all replies) Prompt if necessary
   (i) when I received an invitation to have a breast x-ray
   (ii) newspaper or magazine
   (iii) radio or TV
   (iv) friend or relative
   (v) speaker at a group to which you belong
Ask questions 3-6 if the woman received an invitation. If not please go to question 7.

3. Could you please tell me why you decided to take part in the Breast X-ray programme. (Prompt) (Please tick as many as apply)

(a) To detect breast cancer early □
(b) Because it was recommended by my family doctor □
(c) Because I have a breast symptom □
(d) Because it is a pilot programme □
(e) Because it is a free service □
(f) Because I have a family history of breast cancer □
(g) For reassurance □
(h) Positive reports from other women who had attended □
4. How worried did you feel when you got the appointment to be screened?
   (a) Not at all worried
   (b) A little bit worried
   (c) Quite worried
   (d) Very worried

5. Could you please tell me what you thought about the information that was posted to you with your invitation? (Prompt)
   (i) hard to understand
   (ii) just right
   (iii) too simple
   (iv) didn't see the leaflet/don't know

   Was there any information that you wanted to know that wasn’t included in the leaflet? (please tick one answer)
   (i) Yes
   (ii) No

   If yes, what information should have been included?

6. Could you please tell me which screening centre you were invited to?
   (i) Dunedin hospital screening centre
   (ii) Mobile screening centre
   (iii) Kew hospital screening centre
7. How did you travel to the screening centre?
   (i) walked 
   (ii) car 
   (iii) bus 
   (iv) taxi 
   (v) other (please describe) 

8. How long did it take you to travel to the screening centre?
   (i) less than 15 minutes 
   (ii) between 15 minutes and half an hour 
   (iii) 30 minutes to an hour 
   (iv) more than an hour (please specify) 

Your screening appointment was on (date) at (x)O'clock

9. Could you please tell me what you would have been doing at the time if you had not come to have a breast x-ray? (Prompt)
   (i) working in paid employment 
   (ii) voluntary work 
   (iii) leisure activity 
   (iv) caring for a child or other relative 
   (v) other (please describe) 


10. How easy was it for you to fit the screening appointment into your usual routine? (Prompt)
   (a) Very easy
   (b) Quite easy
   (c) A little bit difficult
   (d) Quite difficult
   (e) Very difficult

11. Could you please tell me how you would rate your risk of getting breast cancer compared with other women your age? (Prompt)
   (i) I think I would be more at risk
   (ii) I think I would be less at risk
   (iii) I think I would have the same risk
   (iv) Don’t know

12. Could you please tell me if anyone in your family or close to you has had breast cancer?
   (i) Yes
   (ii) No

   If yes, please specify:
   (i) close friend
   (ii) mother
   (iii) sister
   (iv) don’t know
   (v) other relative (please specify)
13. If you receive an invitation to be screened (in 2 years time) will you attend?
   (i) Yes 
   (ii) No

   Why ...........................................................................................................................................
   ...........................................................................................................................................
   ...........................................................................................................................................
   ...........................................................................................................................................
   ...............................................................................................................................................  

   'We now need some personal details so we can see if there is any difference between different groups of women and how they felt about the screening programme.

14. Could you please tell me your age? .................................................................  50

15. We need to know your ethnic group. Could you please tell me which ethnic group or groups you feel you belong to? (Prompt)
   (i) European
   (ii) NZ Maori
   (iii) Pacific Island Polynesian
   (iv) Asian
   (v) Other (please specify) ........................................................................................................  56
16. Could you please tell me about your employment? (Prompt)

Are you:

(i) Working for salary/wages

(ii) Self-employed

(iii) Not working in paid employment

(iv) Retired

If you are working in paid employment please describe your occupation. (If retired please describe your previous occupation)

........................................................................................................................................

........................................................................................................................................

17. Could you please tell me your yearly household income before tax? (Prompt)

(a) Less than $15,000

(b) $15,000 - $30,000

(c) More than $30,000

(d) Refused

18. Could you please tell me the highest level of education you have reached? (Prompt)

(i) Primary school

(ii) Secondary school

(iii) University education

(iv) Trade or Vocational Qualification
    (eg: Polytechnic)

(v) Refused

THANK YOU FOR YOUR HELP